Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Maaf Pak, Nanda lupa melampikan sumber data jadi saya lampirkan terpisah ya pak

Trims

Sumber :

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u act=8&ved=2ahUKEwj19a-c-

tnsAhX16nMBHYIuBa4QFjALegQICxAC&url=http%3A%2F%2Famutiara.staff.g unadarma.ac.id%2FDownloads%2Ffiles%2F71282%2FTutorial%2BRapidminer% 2B2.pdf&usg=AOvVaw32FHZ8LaMw7mvec9L-vrn2

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source= web&cd=&cad=rja&uact=8&ved=2ahUKEwi3sdXBtnsAhWE XMBHSZTCMoQFjAAegQIBBAC&url=https%3A %2F%2Fdinus.ac.id%2Frepository%2Fdocs%2Fajar%2FBela jar Data Mining dengan RapidMiner.pdf&usg=AOvVaw3G Ag5jJPm7n39g5DbAW2sT

Tugas 06

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban :

RapidMiner merupakan perangakat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan munggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan mengunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML)Extensible Markup Language) yang mendefenisikan proses analitis keingginan pengguna untuk diterpkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analis secara otomatis.

RapidMiner memiliki beberapa sifat sebagai berikut:

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukaran data.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari RapidMiner, antara lain:

• Banyaknya algoritma data mining, seperti decision treee dan self-organization map.

٥

- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation,

Tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer

Langkah – langkahnya :

1. Kita buka aplikasi RapidMiner 9.8

Interpretational State (1998) (2019) (201

2. Setel itu kita akan berada di halaman utama aplikasi seperti gambar dibawah ini :

<u>File Edit Process View Connections Since </u>	ettings E <u>x</u> tensions <u>H</u> elp											
	View	s: Design	Results	Turbo Prep	Auto Model	Deployments		Find	data, operatorsetc	🔎 Ali S	Studio	r
Tutorials ×	Repository ×	Process							Parameters	×		
Operators and Processes 🔺 View All	🔁 Import Data 🛛 = 💌	ProcessProc	ess			P P 🗎 🖥	👍 <table-cell> 🍯</table-cell>		Process			
AUE	• 🍽 Training Resources (cor	Process							logverbosity	init	•	(j)
Welcome to RapidMiner.	Community Samples (columnation)								logfile			(j)
	Samples) inp						res (
RapidMiner Studio has a lot of	Local Repository (Local)											
minutes, you will learn how you	DB (Legacy)											
can use RapidMiner Studio to perform basic data science												
techniques, such as: accessing												
building statistical models. We will	<											
use data describing the Titanic passengers.	Operators ×								Show advance	ed parameters		
	Search for Operators								✓ Change comp	atibility (9.8.000)	1	
EXPLANATION	h Data Access (EE)											
Each tutorial is broken into several steps to teach you the basics of	 Blending (82) 								Help \times			
RapidMiner Studio. Each step lists	Cleansing (28)								Proces	s		^
explanation of why those actions	Modeling (167)								RapidMine	r Studio Core		
matter. Notice the dotted-line preview in the Process Panel: this	Scoring (14)								Synopsis			
shows you the process you will be	 Utility (85) 								The root operato	r which is the o	outer	
building in that tutorial.		Leverage the Wisdor	n of Crowds to g	get operator recom	mendations base	d on your process de	sign!		most operator o	revery process.	•	
Head Rest	from the Marketplace			Activate V	Visdom of Crowds	3			Description			~

3. RapidMiner menyediakan contoh database yang dapat digunakan, berikut cara menggunakan Sample Data Repository.



4. Untuk tugas saya pilih database produk sebagai contoh tutorial

Pada bagian Repositori terdapat 3 buah lokasi repositori, yakni Samples, DB dan Local Repository.

Untuk mengambil Sample Data Repository, buka hirarki Samples, masuk ke folder Data. Sehingga seperti gambar berikut.



 Lakukan Drag dan Drop salah satu Example Repository. Kita ambil contoh Golf. Tarik dan lepaskan repository ke dalam Main Process, sehingga seperti gambar berikut dimana Repository berada dalam Main Process

<new process*=""> – RapidMiner Studio Educe Elle Edit Process View Connections</new>	cational 9.8.000 @ DESKTOP-A S Settings Extensions <u>H</u>	39118A elp							- 0	×
	•	Views: Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operatorsetc	🔎 All Studio	0 •
Repository ×	Process						Parameters	×		
🕒 Import Data 🛛 = 💌	Process			p p	la 🖪 🚦	🖸 🗣 🗿	Process			
Labor-Negotiations	Process						logverbosity	init		•
Market-Data	Retrieve Products						leafile			
Polynomial	Dinp out					res (logine			
Products										
Purchases										
Sopar										
Titanic										
Titanic Training										
Operators ×							Show advarted by the second	nced parameters mpatibility (9.8.000)		
The Data (array (55)										
Files (20)							Help ×			
 Database (3) 							Droce			^
a Read Database							RapidMi	:>> iner Studio Core		
📤 Write Database										
🍰 Update Database							Synopsis			
< Anniications (0) >	Leverage the Wisdom of C	Crowds to get operator recom	mendations based	l on your process d	esign!		process.	ator which is the outer most	operator of even	У
Get more operators from the Marketplace		🗸 Act	ivate Wisdom of Ci	owds			Description			~

 Hubungkan output pada Database ke Result seperti Gambar diatas. Lalu klik ikon Play . Gambar diatas adalah Sample data repository dari Produk. Coba lakukan untuk memasukkan Sample Repository yang lain.



	-	•		Views: Des	sign Results	Turbo Prep	Auto Model	Deploymer	nts	Find data, operatorsetc	<i>P</i> A	All Studio 🖣
sult History		ExampleSet (Ret	rieve Products)	×						Repository ×		
	Open in	🖣 Turbo Prep	🐴 Auto Model				Filter (178 / 178	examples):	all 🔻	Import I	Data	■
Data	Row No.	Broduct ID	Broduct Na	Broduct Cat	Brico					Samples		
	NOW NO.	FIGURCEID	Flouderina	Floudet cat	FILLE				1			
	1	1	Repressitol	9	35.540					Deals-Tes	stset	
Σ	2	2	Ritalout	10	35.480					Golf		
Statistics	3	3	Comanapracil	10	475.370					Golf-Tests	set	
	4	4	Serum 114	8	303.800					📕 Iris		
	5	5	Hypnocil	10	499.130					Labor-Ne	gotiations	
	6	6	Substance D	4	126.700					Market-Da	ıta	
a an a a a a a a a a a a a a a a a a a	7	7	Hydronium Br	4	89.580					Polynomia Des durts	al	
	8	8	Azoth	1	201.030					Products Purchase	e	
	0	0	Digitalia	5	5.610					Ripley-Se	t	
nnotations	9	5	Digitali	-	5.010					Sonar		
	10	10	Stim pack	/	268.900					Titanic		
	11	11	Damascus R	5	459.140					📒 Titanic Tra	aining	
	12	12	Nepenthe	6	365.700					Titanic Un	labeled	
	13	13	Blaccine	6	495.190					Transactio	ons	
	14	14	Teamocil	8	398.850					Weighting		
										processes		

7. Selanjutkan akan tampil data seperti dibawah ini :

Decision Tree pada RapidMiner

8. RapidMiner sebagai software pengolah data mining menyediakan tool untuk membuat decision tree. Hal ini tentu akan memudahkan kita membuat decision tree dengan menggunakan RapidMiner dibandingkan membuat decision tree secara manual yaitu dengan melakukan perhitungan menggunakan algoritma C4.5 yang telah dijelaskan sebelumnya. Pada contoh kali ini, kita akan membuat keputusan bermain tenis atau tidak. Untuk memudahkan dalam menggunakan RapidMiner untuk membuat decision tree, kita gunakan data sederhana yang ada pada sub bab decision tree. Pertama-tama data pada tabel 2 dibuat lagi dalam format excell dibawah ini :

А	В	С	D	E
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
Sunny	Hot	High	No	Don't Play
Sunny	Hot	High	Yes	Don't Play
Cloudy	Hot	High	No	Play
Rainy	Mild	High	No	Play
Rainy	Cool	Normal	No	Play
Rainy	Cool	Normal	Yes	Play
Cloudy	Cool	Normal	Yes	Play
Sunny	Mild	High	No	Don't Play
Sunny	Cool	Normal	No	Play
Rainy	Mild	Normal	No	Play
Sunny	Mild	Normal	Yes	Play
Cloudy	Mild	High	Yes	Play
Cloudy	Hot	Normal	No	Play
Rainy	Mild	High	Yes	Don't Play

9. Setelah data yang kita punya dibuat dalam bentuk tabel format xls, selanjutnya lakukan Importing Data kedalam Repositori, Lalu cari table Microsoft Excel yang telah dibuat dan masukan kedalam Local Repository seperti yang terlihat pada Gambar dibawah ini :



Lakukan Drag dan Drop Tabel PlayGolf kedalam Process view. Sehingga Operator Database muncul dalam View Proses seperti pada Gambar dibawah ini :



Pada view Process, tabel PlayGolf yang dimasukkan ke dalam proses akan dijadikan sebagai Operator Retrieve.

Untuk membuat decision tree dengan menggunakan RapidMiner, kita membutuhkan operator Decision tree, operator ini terdapat pada View Operators. Untuk menggunakannya pilih Modelling pada View Operator, lalu pilih Classification and Regression, lalu pilih Tree Induction dan pilih Decision Tree.





 Setelah menemukan operator Decision Tree, seret (*drag*) operator tersebut lalu letakkan (*drop*) ke dalam view Process. Kemudian susun posisinya disamping operator Retrieve, seperti yang tampak pada Gambar dibawah ini :



Selanjutnya, hubungkan operator Retrieve dengan operator Decision Tree dengan menarik garis dari tabel PlayGolf ke operator Decision Tree dan menarik garis lagi dari operator Decision Tree ke result di sisi kanan, seperti yang tampak pada Gambar 4.8. Operator Decision Tree berguna untuk memperdiksikan keputusan dari atribut-aribut yang dimasukkan ke dalam operator retrieve. Dengan mengubah tabel (atribut) yang dimasukkan menjadi sebuah pohon keputusan.

Nama : Nanda Tri Haryati NIM/Kelas : 202420016/MTI23-REG-A

6	Process 🕱 🛛 🖹 XML 🕱						
(- → - 1 🕅 Root >	đ	•	\$?	Ø	J	۹
inp	Main Process						res res

Pada operator Decision tree terdapat *input* **training set (tra)**, port ini merupakan output dari operator retrieve. Output dari operator lain juga dapat digunakan oleh port ini. Port ini menghasilkan ExampleSet yang dapat diperoses menjadi decision tree. Selain itu pada operator ini juga terdapat output model (mod) dan example set (exa). Mod akan mengonversi atribut yang dimasukkan menjadi mpdel keputusan dalam bentuk decision tree. **exa** merupakan port yang menghasilkan output tanpa mengubah inputan yang masuk melalui port ini. Port ini biasa digunakan untuk menggunakan kembali sama ExampleSet di operator lebih lanjut atau untuk melihat ExampleSet dalam Hasil Workspace.

11. Langkah selanjutnya ialah mengatur parameter sesuai dengan kebutuhan kita. Setelah menghubungkan operator retrieve dengan operator decision tree, atur parameter decision tree seperti pada gambar dibawah ini :

🛛 🛃 Parameters 🛛	🕒 🕒 Context 🔀
🚨 🗟 🗟 🔊 🛛	🕺 📑 🗝
💡 Dec	cision Tree
criterion	gain_ratio 🔹 🏫
minimal size for split	2
minimal leaf size	1
minimal gain	0.1
maximal depth	20
confidence	0.25
number of preprunin	.3
no pre pruning	
🗌 no pruning	

🦳 🛃 Parameters 🛛	Context 🔀
2 6 6 7 0	🕺 📑 🗝
💡 Dec	cision Tree
criterion	gain_ratio
minimal size for split minimal leaf size	gini_index accuracy
👔 <new process*=""> - RapidM</new>	iner 53.000 @ ambar-PC
Elle Edit Process Tools	Wew Help
📑 📦 🖩 🖬 🖉	Ø 🔊 🔺 🖉 🕒 II 🔳 🛐 🖉 👁

12. Setelah beberapa detik maka RapidMiner akan menampilkan hasil keputusan pada view Result. Jika kita pilih Graph view, maka akan ditampilkan hasilnya berbentuk pohon keputusan seperti pada gambar 4.12. Hasil pohon keputusan dapat disimpan dengan mengklik save image pada sisi kiri View Result.



NIM/Kelas : 202420016/MTI23-REG-A

13. Selain menampilkan hasil decision tree berupa graph atau tampilan pohon keputusan, RapidMiner juga menyediakan tool untuk menampilkan hasil berupa teks. view dengan mengklik button Text View seperti yang tampak pada Gambar dibawah ini :

```
🛒 Result Overview  🏹
                      💡 Tree (Decision Tree)  🛛
Graph View 🔘 Text View 🔘 Annotations
Tree
HUMIDITY = High
    OUTLOOK = Cloudy: Play {Don't Play=0, Play=2}
L
    OUTLOOK = Rainy
L
        WINDY = No: Play {Don't Play=0, Play=1}
L.
    1
        WINDY = Yes: Don't Play {Don't Play=1, Play=0}
    1
Е
    OUTLOOK = Sunny: Don't Play {Don't Play=3, Play=0}
L
HUMIDITY = Normal: Play {Don't Play=0, Play=7}
```

**Selesai **

NAMA	: OMAN ARROHMAN
NIM	: 202420042
MATA KULIAH	: ADVANCED DATABASE

TUGAS 6

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawab :

Data Berolahraga :

Hari	Cuaca	Temperatur	Kecepatan Angin	Berolahraga
Hari ke 1	cerah	normal	pelan	уа
Hari ke 2	hujan	tinggi	pelan	tidak
Hari ke 3	cerah	normal	kencang	уа
Hari ke 4	cerah	normal	pelan	уа
Hari ke 5	hujan	tinggi	kencang	tidak
Hari ke 6	hujan	tinggi	pelan	уа
Hari ke 7	cerah	normal	kencang	tidak

Langkah-langkah :

1. Klik menu File -> New Process -> Pilih Blank

Selanjutnya pada kota dialog Repository pilih **Import Data, -> My Computer,** cari lokasi file Data yang akan diimport.



	Import Data - Select the dat	a location.				×	peratorsetc 🎾 All Studio
Repository ×	-		Select the d	ata location.			×
G Import Data = •	Desumente						init 💌
Local Repository (Local)	Declimenta	File Neme	Pine	Tune	Loot Medified	<u> </u>	
B(Legent)	Bookmarks	Adobe	SIZE	File Folder	May 20, 2020	~	
E o (coloo))		Anowersoft		File Folder	Dec 4 2018		
		Assassin's Creed IV	Black FI	File Folder	Nov 7, 2017		
		Axure		File Folder	lan 22 2018	-	
		CAPCOM		File Folder	Nov 11, 2017		
		Corel		File Folder	Nov 7 2017		
		CPY SAVES		File Folder	Jan 13 2018		
		Criterion Games		File Folder	Apr 3, 2018		
		Custom Office Temp	ates	File Folder	Jan 8, 2018		
		data kamera		File Folder	Apr 21 2018		
		Downloads		File Folder	May 20, 2020		
		Electronic Arts		File Folder	Jan 13, 2018		
		GitHub		File Folder	Oct 6, 2019		
Operators ×		Hasil		File Folder	Aug 6, 2019		
	-	KONAMI		File Folder	Jan 13, 2018		d parameters
learch for Operators		MATERI PROPOSAL		File Folder	May 6, 2019		NEWE (0.0.000)
	-	maya		File Folder	May 10, 2018		1001110 (9.8.000)
Data Access (55)		My Games		File Folder	Nov 7, 2017		
Blending (82)		My Palettes		File Folder	Nov 7, 2017		1
Cleansing (28)		📴 oman		File Folder	Aug 3, 2019		
Modeling (167)		Presentation1		File Folder	Jan 8, 2018		
Scoring (14)		Drozi		Eilo Coldor	0433 3030		Studio Core
Validation (30)							
	All Elles						
Ounty (85)	MITIES						which is the outer most
Extensions (2)					Please select a	file to import.	process.
					← Previous → Mext	Cancel	
Get more operators from the			. A set and the set of a		Louin	processina	t contain exactly one operator of

2. Setelah menemukan file datanya, klik Next, kemudian setelah sampai pada kotak dialog berjudul Format Your Columns dan pada masing-masing atribut pilih logo setelan.



Kriteria :

			Format you	r columns.				
	Quite format MMM d. y	wheness as	•	Replace errors with	missingvalues (2)			
	Jumiah MK o +	Absensi (%) & + /ea/	Jurstah SK., e +	IPS St e +	IPS S2 0 mar	• 8°5 53 /ea/	a -	N P
1	0	0.000	0	0.000	2	2		1
2	0	0.000	a	0.000	7	7		Ĩ
3	23	38.002	39	0.290	0.900	0.000		5
4	1	53.750	50	2.350	2.490	0.000		0
5	37	60.952	60	2.250	2,820	1.800		
4	20	47.509	60	1.580	2.050	0.100		
7	-36	71.845	59	1.979	2.820	2,778		
	-43	54.771	18	2.268	1.810	1.372		
9	8	6.253	10	0.363	0.000	9		
10	9	6.253	60	0.360	0.000	9		
11	35	75.921	29	3.030	2.710	2.799		
12	17	28.365	0	2.530	0.000	0.000		
	<		1000 R.				no people	ì
					1		-	

Pilih Change Type. Kriterianya sebagai berikut ;

- Polynominal = untuk atribut yang memiliki lebih dari 2 kategori
- Binominal = untuk atribut yang memiliki 2 kategori saja
- Real = untuk tipe data yang memiliki nilai koma, atau decimal
- Integer = untuk file bernilai integer atau bilangan buat tanpa koma.

3. Kemudian pada atribut kelasnya atur Change Role dan isi valuenya sebagai label, selanjutnya **klik Nex**t sampai dengan finish.

	-	•		Views:	Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operate	orsetc	P	All Studio 🔻
Result History	E	xampleSet (//Loc	al Repository/d	ataset berolahra	nga) ×						Repository	×		
Data	Open in	Turbo Prep Berolahraga	Auto Model	Cuaca	Temperatur	Kecepatan		Filter (7	7 / 7 examples):	all 🔻	Samples) Import Dat	a cal)	≣▼
	1	ya	Hari ke 1	cerah	normal	pelan					🕨 🗲 Conn	ections		
Σ	2	tidak	Hari ke 2	hujan	tinggi	pelan					data			
Statistics	3	ya	Hari ke 3	cerah	normal	kencang					datas	et berolahra	ga (11/5/	20 3:23 PM – 2 kE
	4	уа	Hari ke 4	cerah	normal	pelan					DB (Lega	cy)		
(5	tidak	Hari ke 5	hujan	tinggi	kencang								
Visualizations	6	уа	Hari ke 6	hujan	tinggi	pelan								
	7	tidak	Hari ke 7	cerah	normal	kencang								
Annotations														
	ExampleSet (7	examples, 1 special	attribute, 4 regula	r attributes)							<			>

4. Setelah finish pada halaman Repository akan muncul data teman-teman yang sudah Terimpor oleh sistem. Kemudian **Drag atau seret** ke halaman kerja RapidMiner kemudian klik Run (F11).



5. Pada kotak dialog **Operators**, pilih menu-menu seperti **Random Tree, Split Data, Apply Model, dan Permormance.** Cara mencarinya tulisan di kotak pencarian pada kotak dialog tersebut.

🧶 <new process*=""> - RapidMiner Studio Educ</new>	ational 9.8.000 @ Aaroh								- 0 ×
Lie Lait Process View Connections	Settings Extensions Help	Views: Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operatorsetc	🔎 All Studio 🔻
Repository ×	Process							Parameters ×	
🕒 Import Data 🛛 = 👻	Process >				o o	S 🖬 🧃 💰	ø 🛛	% Performance (Performance (Classification))
Samples) inp						re ^	main criterion first	• •
Local Repository (Local)	Ret	rieve dataset ber					re		Ð
Gonnections	[out)						accuracy	W
🕨 📴 data								classification error	(I)
processes									
dataset berolahraga (11/5/20 3:		Random Tree						kappa	Ð
DB (Legacy)	•	tra mod exa						weighted mean recall	٢
		0.510.1						weighted mean precision	Ð
<	•	exa par						spearman rho	Ð
Operators ×	l							kendall tau	I
		Apply Model						abaaluta arrar	(D)
performace ×	9	mod 👷 lab						Show advanced parameters	
Multi Horizon Performance	٩	uni mod							
Sliding Window Validation	l							Help ×	
ation (20) 🔤		Performance							/
erformance (18)	•	lab 🥱 per						% Performance (Cla	ssification)
Predictive (7)	6	per exa						RapidMiner Studio Core	
% Performance (Classification)	l							Tags: <u>Accuracy</u> , <u>Errors</u> , <u>Precision</u> , <u>F</u>	Recall, Kappa, uations Metrics
% Performance (Binominal Classific								Confusion Matrix, Predictive	and the second second
% Performance (Regression)							~	Synopsis	
% Performance (Costs)	<						>	This operator is used for statistic	al performance
% Performance (Ranking) V	Leverage the Wisdom of Crowd	Is to get operator recommendation	ons based on your	process design!				evaluation of classification tasks	This operator
< > No results were found.		~	Activate Wisdom	of Crowds				delivers a list of performance crit classification task.	eria values of the

Setelah menemukan semua **Opertors** tersebut, jangan lupa Drag atau seret ke halaman kerja satu persatu.

6. Selanjutnya pada **Operator Split Data klik 2x**, kemudian atur Sampling Type pada kota Parameters , ini bertujuan untuk memisah atau membagi menjadi data latih dan data uji guna untuk mengetahui permorma dari model yang dibentuk oleh k-NN

100% Q Q 4001	Parameters	×
	partitions	Edit Enumer D
	sampling type	linear sampli • D
Spit Data as par D 1	use local ra	ndom seed 🛛 🛈
	🕈 Hide advan	ced parameters.
	Help ×	

7. Setelah itu atur porposisi atau kapasitas antara data latih dan data uji. Klik pada menu edit *Enumeration*. kemudian akan muncul kotak dialog sepeti dibawah ini :



8. Add Entry kemudain Ok.

Sumber :

https://www.youtube.com/watch?v=h82NuHDNhKI

https://www.youtube.com/watch?v=jXRXI66inDM

https://www.youtube.com/watch?v=jB_hKsi408

https://www.youtube.com/watch?v=idHF3t4xMak&t=286s

Nama : Puspita Dewi Setyadi

Nim : 202420011

Rapidminer adalah salah satu tools yang digunakan pada data mining

Langkah-langkah menggunakan tools rapidminer sebagai berikut :

- 1. Mendownload tools rapidminer terlebih dahulu, Disisni saya menggunakan rapidminer versi 5.3
- 2. Setelah terinstal pada laptop atau pe lalu kita klik pada tools rapidminer setelah terbuka klik new
- 3. Kemudian klik rea pada kolom, lalu muncul beberapa read. Sesuaikan tipenya
- 4. Setelah itu klik pada layar main process
- 5. Selanjutnya jika ingin memasukkan dataset yang digunakan klik next
- 6. Pada tahap selanjutnya klik finish, setelah itu ketikkan validation, validation ini berguna untuk testingnya
- 7. Kemudian hubungkan dataset dan validationnya, setelah itu double klik pada validationnya
- 8. Dan muncul ada 2 box yaitu box training dan box testing
- 9. Pada box training ini (algoritma yang akan kita gunakan)
- 10. Dapat melihat prediksi dan faktanya yang menghasilkan.

Nama : Robby Prabowo

NIM : 202420001

Kelas : MTIA1

- 1. Donload aplikasi Rapid Miner di google dan install aplikasi
- 2. Buatlah datasheet di excel
- 3. Import datasheet ke Rapid Miner
 - a. Cari di Operators Read Excel
 - b. Double klik Read Excel atau drag ke area Process
 - c. Double klik Read Excel di area Process dan cari datasheet yang sudah kita buat



4. Cari Multiply di Operators lalu double klik



5. Cari Splid Data di Operators lalu double klik

In the process -> - RapidMiner Studio Educational 9.8.000 @ DESKTOP-QT46LVL Elle Edit Process View Qonnections Settings Extensions Help

Turbo Prep 🔎 All Studio 🔻 •
 • Design Results Auto Model Deployments Find data, operators...et Tutorials Parameters Repository Process T Split Data Process P P 🐚 🖡 📮 🏹 🔍 Operators and Processes 🔺 View All 🕒 Import Data 🛛 = 💌 Edit Enumera... (1) partitions Samples Local Repository (Read Excel Solit Data Multiply Welcome to RapidMiner. automatic • (sampling type DB (Legacy) + RapidMiner Studio has a lot of features to offer. In the next few minutes, you will learn how you can use RapidMiner Studio to Welcome to RapidMiner! perform basic data science techniques, such as: accessing We have a super brief introduction to using data, transforming data, and RapidMiner on the left of your screen. Check it ou building statistical models. We will use data describing the Titanic Operators × passengers. Show advanced parameters split data × EXPLANATION T Blending (1) Each tutorial is broken into several Help × ▼ 📇 Examples (1) steps to teach you the basics of RapidMiner Studio. Each step lists 🕶 🛅 Sampling (1) 🍸 Split Data specific actions to perform, plus an explanation of why those actions P Split Data Split Data dMiner Studio Core Partitions an example set into subsets according to the specified relative sizes. Tags: <u>Divide, Separate, Part, Training,</u> Testing, <u>Samples, Subsets, Partitions,</u> matter. Notice the dotted-line preview in the **Process** Panel; this shows you the process you will be Sampling building in that tutorial. d on your process design! Synopsis Press "F3" for focus Activate Wisdom of Crow This operator produces the desired number of subsets of the given No results were found. Type here to search хI へ い (20:5 へ い (29/10/ w

o ×

6. Cari id3 di Operators lalu double klik



7. Cari Apply Model di Operators lalu double klik



- 8. Double klik pada Split Data, tentukkan Rationya
 - a. Tentukan Ratio pertama : Bagi sisa jumlah sisa data yang akan diuji dengan jumlah

total data, misalkan jumlah data 6 sisa data 5 berarti 5:6=0.833 untuk ratio pertama

n ×

- b. Tentukkan Ratio kedua : Bagi jumlah data yang ingin diiuji dengan jumlah total datasheet, misalkan 1:6=0.167
- 9. Sambungkan antara
 - a. Read Excel dengan Splid Data,
 - b. Splid Data dengan ID3,
 - c. ID3 dengan Apply Model,
 - d. Apply Model dengan Result,
 - e. ID3 dengan Result,
 - f. Parameter yang ada di Split data dengan unl pada Apply Model
 - g. Parameter yang ada di Split Data dengan Result
 - h. Out Multiply dengan Result

🧶 <new process*> – RapidMiner Studio Educational 9.8.000 @ DESKTOP-QT46LVL





10. Klik tombol Play diatas, secara otomatis Rapid Miner akan membuat pohon keputusan

Nama : Shabila Fitri Aulia Nim : 202420024 Kelas : MTI A23

TUGAS 06

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumberrujukan tutorial yang anda gunakan.

Jawaban :

Pada kasus ini saya mengambil pengolahan data terkait dengan Market Basket yang akan dicoba pada tool rapidminer. Adapun tahapannya sebagai berikut :

NO	PULPEN	PENGHAPUS	PENSIL	PENGARIS	STABILO	TIPEK	
1	1	1	1	0	0	0	
2	0	1	0	0	1	1	
3	1	1	0	1	0	0	
4	0	0	0	1	1	1	

1. Create tambel master datanya di Excel kemudian export ke Rapidminer.

Ext	oort	ke	Ra	pidm	niner	dan	sesuaika	n type	e datany	a	•
LA	5011	кc	ILU	prun	mor	uun	Sesuarka	n type	unitity	u	٠

istory	Ex:	ampleSet (//Loca	al Repository/ex	celtugas06)	×				
	Open in 🔢	Turbo Prep	Auto Model			Filter (5	/ 5 examples):	all	•
а	Row No.	A	В	с	D	E	F	G	
	1	NO	PULPEN	PENGHAPUS	PENSIL	PENGARIS	STABILO	TIPEK	-
	2	1	1	1	1	0	0	0	
tics	3	2	0	1	0	0	1	1	
	4	3	1	1	0	1	0	0	
Ī	5	4	0	0	0	1	1	1	

2. Create Modul Association Rule, FP-Growth, dan Numerical to Binomial.

Numerical to Binomial.

Sesuaikan dengan parameter – parameter yang ada.

Interpretation of the state	ducational 9.8.000 @ DESKTOP-S ns <u>S</u> ettings Egtensions <u>E</u>	ISRFV2F Help				8 X
	-	Views: Design Results	Turbo Prep Auto Model	Deployments	Find data, opera	atorsetc 🔎 All Studio 🔻
Repository ×	Process				Parameters >	
🕒 Import Data 🛛 🖃 💌	Process >		, ₽ , ₽ 🐚	🚦 📮 🍒 🥔 🖸	Numerical to Bino	minal
F Training Resources (connected	Process				attribute filter type 💙	all 🔻 🛈
 Community Samples (connect) Samples 	inp fil to out	Numerical to Binominal	FP-Growth	res	invert selection	(D)
Teolo	1	ori D	fre d	res	include special attr	ibutes 💙 🕕
Deals Deals-Testset					min	0.0
Golf-Testset					max	0.0
<pre>Operators ×</pre>			Create Association			
create ass ×			ite rul			
 Modeling (1) Associations (1) Create Association Rules 						
We found "MeaningCloud Text	Recommended Operators	1		~	• Activate V	Vindows
Analytics" in the Marketplace. <u>Show me!</u>	C Retrieve	2 70% Select Attributes	2 ² 39% 👿 Set Role	2² 37%	 Change compatibility 	lity (9.8.000) S to activate Windows.
₽ Type here to search		H 🖪 😂 🗾 👂	R 🧕 🖬 💶 🛛	I 🚾 📎 🖸	27 🗐 ^	¢») 🖮 ENG 14.50 28/10/2020 ₹1

FP-Growth

Sesuaikan dengan parameter yang ada :

In the second	ducational 9.8.000 @ DESKTOP-95R ns <u>S</u> ettings Extensions <u>H</u> el	FV2F				- 0	×
	•	Views: Design Results	Turbo Prep Auto Model Deploym	nents	Find data, operatorsetc	🔎 All St	udio 🔻
Repository ×	Process				Parameters ×		
🕂 Import Data 🛛 🗉 💌	Process >		🔎 🔎 🐚 📑 📑	🤹 🖝 🖸	🛒 FP-Growth		
Training Resources (connected	Process			n	nin requirement	support	• @
Samples	inp fil out	Numerical to Binomi	FP-Growth exa exa	res m	in support	0.95	•
• 😋 data		v		m	nin items per itemset	1	Ð
Deals-Testset				m	nax items per itemset	0	
Golf-Testset				m	nax number of itemsets	1000000	D
			Create Association		🗸 find min number of itemsets		٢
create ass ×			fite mul	m	in number of itemsets	100	•
🕶 🛅 Modeling (1)				n	nax number of retries	15	Ð
 Associations (1) Create Association Rules 				n	equirement decrease factor	0.9	Ð
				m	nust contain list	🗒 Edit Enum.	
	Province de la constante de la	τ.		n	nust contain regexp		
We found "MeaningCloud Text Analytics" in the Marketplace.	Recommended Operators		at nov		Activate Window	S	~
Show me!	C. Kettleve	Select Attributes	2º 39%	2- 37%	Go to settings to activ		
₽ Type here to search	Ĕ	# 📃 😂 💶 🕫	R 👩 🗄 💶 💷	S	へ (3) 物 E	NG 14.50 28/10/2020	5

Association Rule

Sesuaikan dengan parameter yang ada :

🌖 <new process*=""> - RapidMiner Studio Ec</new>	ducational 9.8.000 @ DESKTOP-95	RFV2F					 ×
	v m	Views: Design Results	Turbo Prep Auto Model	Deployments	Find data, ope	eratorsetc 🔎	All Studio 💌
Repository ×	Process				Parameters	×	
🕒 Import Data 🛛 = 💌	Process >		,e ,e 🗈 👔	🗵 🖷 🗸 🖬	🛒 Create Associat	tion Rules	
Training Resources (connected	Process				criterion	confidence	•
Community Samples (connect Samples	Read Excel	Numerical to Binomi	FP-Growth	res	min confidence	0.8	Ð
▼ <mark></mark> data	V	ori	fre fre	res	gain theta	2.0	Ð
Deals-Testset					laplace k	1.0	٩
Golf-Testset							
Operators ×			Create Association Rules)			
create ass 🗙			a de la companya de l				
The Modeling (1)							
Associations (1) Create Association Rules							
We found "MeaningCloud Text	Recommended Operators	1		~	Antinata	147 - I	
Analytics" in the Marketplace. Show mel	<u> </u> <i>d</i> Retrieve	2 70% Select Attributes	🝰 39% 🚺 Set Role	2² 37%	Activate	windows arameters ligs to activate Wind	
🕂 🔎 Type here to search		H 🛤 😂 💶 🤌	🙊 🧔 🔒 💶 💵	🐖 📎 🚾	-	へ ¢») ¶æ ENG <u>28/10</u> /	51 /2020 🖣

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 355 rules yang ditemukan.

<new process*=""> File Edit Proce</new>	– RapidMiner Studio Educational 9.8.000 @ ss. View. Connections. Settings. Ext	DESKTOP	95RFV2F Heln			– 0 ×
			Views: Design	Results Turbo Prep Auto Model	Deployments	Find data, operatorsetc 🔎 All Studio 🔻
Result History	🛒 AssociationRules (Cre	ate Asso	ciation Rules) 🗙 📘 Ex:	ampleSet (//Local Repository/exceltugas06)	×	Repository ×
	Show rules matching	No.	Premises	Conclusion	Support	🕒 Import Data 🛛 🖃 💌
Dett:	all of these conclusions:	340	D, B, E, F, C	G	0.250	Training Resources (connected)
Data	D	341	G, C	D, B, E, F	0.250	Community Samples (connected)
	B	342	D, O, C	8, E, F	0.250	▼ <mark>></mark> Samples
Z	F	343	B, G, C	D, E, F	0.250	Toolo
Graph	G	344	DBGC	FF	0.250	Deals-Testset
		245	5,5,0,0	 D.B.5	0.250	Golf
=		343	2,0,0	5,5,1	0.250	Golf-Testset
_		346	D, E, G, C	B, F	0.250	Iris
Description		347	B, E, G, C	D, F	0.250	Labor-Negotiations
		348	D, B, E, G, C	F	0.250	Market-Data
		349	F, G, C	D, B, E	0.250	Polynomial Products
Annotations		350	D. F. G. C	B. E	0.250	Purchases
		361	BEOC	D.E	0.250	Ripley-Set
		331	5,1,0,0	0,2	0.230	Sonar 📒
		352	D, B, F, G, C	E	0.250	Titanic
	Min. Criterion:	353	E, F, G, C	D, B	0.250	Titanic Training
	confidence 🔻	354	D, E, F, G, C	в	0.250	Titanic Unlabeled
	Min. Criterion Value:	355	BEEGO	D	0.250	Transactions
		<			>	Go to Settings to activate Windows.

Graph



Nama : Siti Ratu Delima Nim : 202420025 Kelas : MTI24

TUGAS 06

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamubuatsendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban :

Pada kasus ini saya mengambil pengolahan data terkait dengan Market Basket yang akan dicoba pada tool rapidminer. Adapun tahapannya sebagai berikut :

1.	Create	tambel m	aster dat	anya di Excel	kemudiar	n export ke	Rapidmin	ler.

 							_
NO	PULPEN	PENGHAPUS	PENSIL	PENGARIS	STABILO	TIPEK	
1	1	1	1	0	0	0	
2	0	1	0	0	1	1	
3	1	1	0	1	0	0	
4	0	0	0	1	1	1	

Export ke Rapidminer dan sesuaikan type datanya :

listory	I E	xampleSet (//Lo	cal Repository/e	cceltugas06)	×				
	Open in	🖣 Turbo Prep	👫 Auto Model			Filter	(5/5 examples):	all	•
a	Row No.	A	В	с	D	E	F	G	
	1	NO	PULPEN	PENGHAPUS	PENSIL	PENGARIS	STABILO	TIPEK	
	2	1	1	1	1	0	0	0	
ics	3	2	0	1	0	0	1	1	
	4	3	1	1	0	1	0	0	
	5	4	0	0	0	1	1	1	

2. Create Modul Association Rule, FP-Growth, dan Numerical to Binomial.

Numerical to Binomial.

Sesuaikan dengan parameter - parameter yang ada.

Interprocess*> – RapidMiner Studio E Elle Edit Process Ylew Connection	ducational 9.8.000 @ DESKTOP- ons Settings Extensions !	95RFV2F ∃elp			Г		- 8 ×
	•	Views: Design	Results Turbo Prep	Auto Model Deploy	rments	Find data, opera	torsetc 🔎 All Studio 👻
Repository ×	Process					Parameters ×	
🕒 Import Data 🛛 🖃 💌	Process >			PP 🗅 🗈 🗖	🖸 🗣 🗿	Numerical to Binor	minal
F Training Resources (connected	Process					attribute filter type 💙	all 🔻 🛈
 Community Samples (connect- Samples 	Read Excel	Numerical to Binomin	al 🛛 🚽	FP-Growth	res	invert selection	0
💌 📴 data	√	ori	~	o of fre	res	include special attr	ibutes 🂙 🕕
Deals-Testset						min	0.0
Golf-Testset						max	0.0
Operators ×				·	Ļ		
create ass X			a contraction in the second se	P			
 Modeling (1) Associations (1) 				-			
Create Association Rules							
We found "MeaningCloud Text	Recommended Operators	0			*	Activata	
Analytics" in the Marketplace. <u>Show me!</u>	L Retrieve	🗳 70% 🚺 Select Att	tributes 🙎 39%	🚺 Set Role	2² 37%	 <u>Change compatibil</u> 	ity (9.8,000) S to activate Windows.
							14 50 —
P Type here to search			🧶 🖉 😐	N 💁 📲 🖉	N 🛛	· ·	(14.30 (1

FP-Growth

Sesuaikan dengan parameter yang ada :

Interpret Studio Education Studio Education Education Education Education Studio Educati	lucational 9.8.000 @ DESKTOP-95F ns <u>S</u> ettings Extensions <u>H</u> e	RFV2F Ip				- 0	×
	•	Views: Design Results	Turbo Prep Auto Model Deplo	yments	Find data, operatorsetc	🔎 Ali s	Jtudio 🔻
Repository ×	Process				Parameters ×		
🕂 Import Data 🛛 🗉 💌	Process >		🔎 🔎 🗎 📘	🛛 🗣 🍒	₩ FP-Growth		
Training Resources (connected	Process	Numerical to Rinomi	EP Growth		min requirement	support	•
 Samples 		exa 🌮 exa	exa	res	min support	0.95	
Teals	✓	✓			min items per itemset	1	D
Deals-Testset					max items per itemset	0	Ð
Golf-Testset					max number of itemsets	1000000	D
Operatore			Create Association		✓ find min number of itemsets		١
create ass ×			te rul		min number of itemsets	100	
 Modeling (1) Accessizations (1) 			v		max number of retries	15	D
Create Association Rules					requirement decrease factor	0.9	D
					must contain list	🖐 Edit Enum	1 ①
	Recommended Operators	0		~	must contain regexp		
 We found "MeaningCloud Text Analytics" in the Marketplace. <u>Show mel</u> 	<u>/</u> Retrieve	2º 70% Select Attributes	💒 39% 🚺 Set Role	2² 37%	The advanced parameters	rs ate Windows	
₽ Type here to search		H 📻 🚖 💶 🏚	🞗 🌖 🛱 🚳 💵 🖷	S) 🔽	🌖 🔷 🗘	NG 14.50	- R 1

Association Rule

Sesuaikan dengan parameter yang ada :

《new process*> - RapidMiner Studio Er File Edit Process View Connectio	ducational 9.8.000 @ DESKTOP-95F Ins Settings Extensions He	RFV2F					
	•	Views: Design Results	Turbo Prep Auto Model [Deployments	Find data, op	eratorsetc 🔎 All	Studio 🔻
Repository ×	Process				Parameters	×	
🔂 Import Data 🛛 = 💌	Process >		🔎 🔎 🗎 📑	🖂 🗣 🖻	🛒 Create Associa	tion Rules	
F Training Resources (connected	Process				criterion	confidence	•
Community Samples (connect)	Read Excel	Numerical to Binomi	FP-Growth		min confidence	0.8	Ð
 Samples 	Dinp 🖌 fil 📩 out 🕅	exa 🔔 exa	exa exa	res			
Teals	✓	v			gain theta	2.0	•
Deals-Testset					laplace k	1.0	١
Golf							
Golf-Testset							
< III >							
Operators ×			te unit rul	/			
create ass 🗙			a de la de l				
🕶 📇 Modeling (1)			\checkmark				
 Associations (1) 							
🛒 Create Association Rules							
We found "MeaningCloud Text	Recommended Operators	0		~	A	MC - I	
Analytics" in the Marketplace. <u>Show me!</u>	C Retrieve	2 70% Select Attributes	📽 39% 🚺 Set Role	2² 37%	E Hide advanced	Parameters Higs to activate Window	
		_					
🛨 🔎 Type here to search		# 📻 📚 💶 🥵	😟 📀 🔒 💶	🛛 📎 🔽		へ (14.51 へ (19) 第回 ENG 28/10/202	20 🗟

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 355 rules yang ditemukan.

Interprocess Interprocess Elle Edit Proce	– RapidMiner Studio Educational 9.8.000 @ ss _Ylew _Connections _Settings Ext	DESKTOP	-95RFV2F Help			– 0 ×		
	• •		Views: Design Results	Turbo Prep Auto Model Deployments	F	Find data, operatorsetc 🔎 All Studio 🔻		
Result History	🛒 AssociationRules (Cre	ate Asso	ciation Rules) × 📕 ExampleSet (//L	.ocal Repository/exceltugas06) 🛛 🛛 👋		Repository ×		
	Show rules matching	No.	Premises	Conclusion	Support	🗘 Import Data 🛛 📼 🔻		
Data	all of these conclusions: 🔹 🔻	340	D, B, E, F, C	G	0.250	Training Resources (connected)		
	D	341	G, C	D, B, E, F	0.250	Community Samples (connected)		
	8	342	D, G, C	8, E, F	0.250	▼ <mark>™</mark> Samples		
Z	F	343	B, G, C	D, E, F	0.250	Teals		
Graph	G	344	D, B, G, C	E, F	0.250	Deals-Testset		
		345	EGC	DBF	0.250	Golf		
=		040	2,0,0	5,5,7	0.230	Golf-Testset		
_		346	D, E, G, C	B, F	0.250	Iris		
Description		347	B, E, G, C	D, F	0.250	Labor-Negotiations		
		348	D, B, E, G, C	F	0.250	Market-Data		
1		349	F, G, C	D, B, E	0.250	Polynomial Products		
Annotations		350	D. F. G. C	B. E	0.250	Purchases		
		261	BEOC	DE	0.250	Ripley-Set		
		331	8, 1, 0, 0	0,2	0.250	Sonar 🗧		
		352	D, B, F, G, C	E	0.250	Titanic		
	Min. Criterion:	353	E, F, G, C	D, B	0.250	Titanic Training		
	confidence 💌	354	D, E, F, G, C	в	0.250	Titanic Unlabeled		
	Min. Criterion Value:	355	B, E, F, G, C	D	0.250	Iransactions		
	<u> </u>	<			>	Go to Settings to activate Windows.		

Graph



SUMBER : BelajarDataMiningdenganRapidMiner Penyusun:

- 1. Dennis Aprilla C
- 2. Donny Aji Baskoro
- 3. Lia Ambarwati
- 4. I Wayan

Simri Wicaksana:

Remi Sanjaya

Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C

BELAJAR DATA MINING DENGAN RAPID MINER

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati Wayan Simri Wicaksana

BELAJAR DATA MINING DENGAN RAPID MINER

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana

Identitas

Belajar Data Mining dengan RapidMiner

Penyusun:

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor: Remi Sanjaya

Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C
Kata Pengantar

Dengan mengucapkan puji syukur kepada Tuhan YME atas Berkah Rahmat dan Hidayah-Nya, penulis dapat menyelesaikan buku yang berjudul Belajar Data Mining dengan RapidMiner.

Produk-produk perangkat lunak gratis (freeware) dan bersifat open source yang demikian banyak jumlahnya, telah memudahkan kita dalam melakukan proses pengolahan dan analisis data. Dalam melakukan analis terhadap data mining, RapidMiner merupakan salah satu solusi yang dapat kita gunakan. Keberadaan RapidMiner yang berupa freeware dan dapat dijalankan pada berbagai sistem operasi tidak hanya menguntungkan penyedia aplikasi karena tidak perlu mengeluarkan biaya untuk lisensi perangkat lunak, tetapi juga memudahkan pengembang maupun calon pengembang dalam mempelajari dan mencoba sendiri fitur-fitur yang ada.

Buku ini diharapkan dapat membantu pembaca mempelajari RapidMiner, melalui rangkaian tutorial bertahap mulai dari proses instalasi hingga pemrograman. Pada buku ini juga dibahas beberapa teori penunjang mengenai data mining seperti, decision tree, neural network dan market basket analysis untuk membuka wawasan pembaca mengenai data mining sebelum melakukan analisis data mining.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu penyelesaian buku ini.

Akhir kata, penulis menyadari masih terdapat kekurangan dalam penyusunan buku ini baik pada teknis penulisan maupun materi, mengingat akan kemampuan yang dimiliki penulis. Untuk itu kritik dan saran dari semua pihak penulis harapkan demi penyempurnaan pembuatan buku ini. Semoga buku ini dapat bermanfaat bagi para pembaca.

Jakarta, April 2013

Penulis

Daftar Isi

Kata Pengantari
Daftar Isi iii
Daftar Gambar v
Daftar Tabelvii
Kecerdasan Buatan 2
Definisi Kecerdasan Buatan2
Ruang Lingkup Kecerdasan Buatan5
Perbedaan Komputasi Kecerdasan Buatan dan Komputasi Konvensional6
RapidMiner Error! Bookmark not defined.8
Apa itu RapidMiner?8
Instalasi Software11
Pengenalan Interface16
Cara Menggunakan Repositori28

Data Mining	39
Mengenal Data Mining	39
Pengelompokan Teknik Data Mining	43
Decision Tree	45
Mengenal Decision Tree	45
Algoritma c4.5	48
Kelebihan Pohon Keputusan	55
Kekurangan Pohon Keputusan	56
Decision Tree pada RapidMiner	56
Neural Network	84
Market Basket Analysis	96
Memahami Market Basket Analysis	96
Metodologi Association Rules	100
Contoh Association Rules	102
Frequent Itemset Generation dan Rule Generati	on
	105
Market Basket Analysis pada RapidMiner	107
Glossarium	122
Daftar Pustaka	125

Daftar Gambar

Gambar 1.1 Proses Kecerdasan Buatan	4
Gambar 2.1 Form Awal Instalasi	14
Gambar 2.2 Form Persetujuan Lisensi	14
Gambar 2.3 Form Pemilihan Lokasi Instalasi	15
Gambar 2.4 Form Proses Instalasi	15
Gambar 2.5 Form Instalasi selesai	16
Gambar 2.6 Tampilan Welcome Perspective	17
Gambar 2.7 Welcome Perspective	19
Gambar 2.8 Header Tab	20
Gambar 2.9 Tampilan Design Perspective	21
Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki	23
Gambar 2.11 Tampilan Parameter View	25
Gambar 2.12 Problem & Log View	27
Gambar 2.13 Kumpulan Sample Data Repository	28
Gambar 2.14 Tampilan Design Perspective Awal	29
Gambar 2.15 Repository berada dalam Main Process	29
Gambar 2.16 Menghubungkan Output Repositori ke Result	30
Gambar 2.17 Isi Sample Golf Data Repository	30
Gambar 2.18 Repository	32
Gambar 2.19 Step 1 of 5 Import Wizard	32
Gambar 2.20 Step 2 of 5 Import Wizard	33
Gambar 2.21 Step 3 of 5 Import Wizard	34
Gambar 2.22 Step 4 of 5 Import Wizard	34
Gambar 2.23 Tipe Data	35

Gambar 2.24 Step 5 of 5 Import Wizard	35
Gambar 2.25 Repository yang sudah diimport	36
Gambar 2.26 Menghubungkan Output Repositori pada Result	36
Gambar 2.27 Tabel Repository	37
Gambar 4.1 Bentuk Decision Tree Secara Umum	48
Gambar 4.2 Grafik Entropi	50
Gambar 4.3 Tabel Keputusan dalam Format xls	57
Gambar 4.4 Lokasi Tabel pada Repository	58
Gambar 4.5 Repository PlayGolf pada Main Process	59
Gambar 4.6 Daftar Operator pada View Operators	59
Gambar 4.7 Posisi Operator Decision Tree	60
Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator Decision	
Tree	61
Gambar 4.9 Parameter Decision Tree	62
Gambar 4.10 Tipe Criterion	62
Gambar 4.11 Ikon Run	66
Gambar 4.12 Hasil Berupa Graph Pohon Keputusan	66
Gambar 4.13 Hasil Berupa Penjelasan Teks	67
Gambar 4.14 Tabel SakitHipertensi dalam format xls	69
Gambar 4.15 Lokasi Tabel pada Repository	69
Gambar 4.16 Tabel SakitHipertensi pada Main Process	70
Gambar 4.17 Hirarki Operator X-Validation	72
Gambar 4.18 Operator Validation	72
Gambar 4.19 Parameter X-Validation	74
Gambar 4.20 Hirarki Operator Apply	77
Gambar 4.21 Operator Apply Model	78
Gambar 4.22 Parameter Apply Model	79
Gambar 4.23 Hirarki Operator Performance	80
Gambar 4.24 Operator Performance	81
Gambar 4.25 Parameter Performance	82
Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Performar	nce
	82
Gambar 4.27 Susunan Operator Retrieve dengan Operator Validation	83
Gambar 4.28 Tampilan Decision Tree	83
Gambar 6.1 Frequent Item Set tanpa Apriori	106
Gambar 6.2 Frequent Item Set dengan Apriori	106

Gambar 6.3 Tabel Penjualan Sederhana	108
Gambar 6.4 Repositori	108
Gambar 6.5 Database dalam Main Process	109
Gambar 6.6 Operator Create Association Rules	109
Gambar 6.7 Operator FP-Growth	110
Gambar 6.8 Operator Numerical to Binominal	110
Gambar 6.9 Pencarian Operator Numerical to Binominal	111
Gambar 6.10 Pencarian Association Rules	112
Gambar 6.11 Menghubungan Database TransaksiMakanan pada C	perator
Numerical to Binomial	112
Gambar 6.12 Parameter Numerical to Binomial	113
Gambar 6.13 Menghubungkan Operator Numerical to Binomial de	ngan
Operator FP-Growth	114
Gambar 6.14 Parameter FP-Growth	115
Gambar 6.15 Menghubungkan Operator FP-Growth dengan Opera	itor
Create Association Rules	115
Gambar 6.16 Parameter Association Rules	116
Gambar 6.17 Susunan Operator Association Rules	117
Gambar 6.18 Hasil Association Rules Pertama	117
Gambar 6.19 Operator FP-Growth	118
Gambar 6.20 Mengubah Parameter FP-Growth	119
Gambar 6.21 Operator Create Association Rules	119
Gambar 6.22 Mengubah Parameter Association Rules	120
Gambar 6.23 Hasil Association Rules Kedua	120
Gambar 6.24 Hasil dalam bentuk Graph View	121

Daftar Tabel

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konve	ensional7
Tabel 4.1 Keputusan Bermain Tenis	52
Tabel 4.2 Perhitungan Simpul 1	53
Tabel 6.1 Tabel Transaksi	102
Tabel 6.2 Kombinasi Produk dan Nilai Support	103
Tabel 6.3 Association Rules dan Nilai Confidence	104

Bagian Satu Pendahuluan

Pengenalan Kecerdasan Buatan

Pengenalan RapidMiner

Kecerdasan Buatan

Definisi Kecerdasan Buatan

Manusia memiliki kecerdasan, manusia memiliki kemampuan untuk menganalisa suatu masalah dengan menggunakan pengetahuan dalam otaknya dan

pengalaman yang pernah dilaluinya. Pengetahuan ketika manusia belajar, maka dari itu datang pembelajaran merupakan faktor penting bagi manusia untuk mencapai sebuah kecerdasan. Namun pengetahuan tidak akan cukup untuk menyelesaikan masalah jika tidak memiliki pengalaman, karena pengalaman akan selalu membawa pengetahuan baru. Tetapi akan sia sia, jika seseorang yang memiliki banyak pengalaman tetapi tidak memiliki akal untuk menalar

semua pengetahuan dan pengalaman yang ia miliki. Kombinasi dari pengetahuan, pengalaman, dan kemampuan menalar inilah yang membuat manusia menjadi cerdas dan dapat menyelesaikan permasalahan yang ia hadapi.

Berdasarkan konsep diataslah kecerdasan buatan dibuat. Agar mesin dapat bertindak seperti seorang manusia, maka mesin tersebut harus memiliki pengetahuan dan sejumlah pengalaman serta kemampuan menalar yang dapat mengubah pengetahuan dan pengalaman tersebut menjadi sebuah keputusan dalam menyelesaikan sebuah permasalahan.

Komputer awalnya diciptakan hanya untuk melakukan sebuah perhitungan saja. Jaman terus akhirnya berkembang hingga komputer kini diberdayakan manusia untuk membantu pekerjaannya kesehariannya. Maka dari itu dalam komputer diharapkan memiliki kemampuan yang hampir sama dengan manusia agar dapat mengerjakan segala sesuatu vang bisa dikerjakan oleh manusia Kecerdasan Buatan.

The art of creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)

The study of how to make computers do things at which, at the moment, people are better (Rich dan Knight, 1991)

A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes (Schalkoff, 1990)

The branch of computer science that is concerned with the automation of intelligent behavior (Luger dan Stubblefield, 1993)

Jadi apakah kecerdasan buatan itu? Kecerdasan buatan adalah salah satu bagian dari ilmu komputer yang membuat agar mesin dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia. Dengan demikian, untuk menciptakan sebuah aplikasi kecerdasan buatan terdapat dua bagian utama yang sangat dibutuhkan.



Gambar 1.1 Proses Kecerdasan Buatan

Ruang Lingkup Kecerdasan Buatan

Kecerdasan buatan merupakan teknologi yang fleksibel, dan dapat diterapkan di berbagai macam bidang ilmu. Kemampuan kecerdasan buatan menjadi sangat dibutuhkan di bidang ilmu lain, karena konsepnya tak lagi procedural melainkan meniru cara berpikir manusia. Tak heran kecerdasan buatan bisa di gunakan untuk bidang psikologi yang dikenal dengan cognition dan psycolinguistic. Namun yang paling sering dekat dengan kita ialah robotika, yakni kecerdasan buatan di dalam ilmu elektornika.

Semakin banyaknya ilmu yang menggunakan kecerdasan buatan, semakin sulit juga bagi manusia untuk mengkategorikannya, maka dari itu dibentuklah ruang lingkup kecerdasan buatan yang dapat mewakilinya (Turban dan Frenzel, 1992, pp21-26):

 Sistem Pakar. komputer digunakan untuk menyimpan pengetahuan para pakar. Dengan demikian komputer akan memiliki keahlian untuk menyelesaikan permasalahan dengan meniru keahlian yang dimiliki oleh pakar.

- Pengolahan Basa Alami. dengan pengolahan bahasa alami ini diharapkan user dapat berkomunikasi dengan komputer dengan menggunakan bahasa sehari-hari.
- Pengenalan Ucapan. Melalui pengenalan ucapan diharapkan manusia dapat berkomunikasi dengan komputer dengan menggunakan suara.
- 4. Robotika dan Sistem Sensor
- Computer Vision. Mencoba untuk dapat menginterpretasikan gambar atau objek-objek tampak melalui komputer.
- Intelligent Computer-aided Instruction.
 Komputer dapat digunakan sebagai tutor yang dapat melatih dan mengajar.
- 7. Game Playing.

Perbedaan Komputasi Kecerdasan Buatan dan Komputasi Konvensional

Komputasi Konvensional merupakan Komputer yang hanya digunakan untuk alat hitung. Sangatlah berbeda, kerja dan konsep dari kedua komputasi ini. Agar dapat memberikan gambaran, table berikut adalah

detail dari perbedaan komputasi kecerdasan buatan dan komputasi konvensional.

Dimensi	Komputasi Kecerdasan	Komputasi
	Buatan	Konvensional
Pemrosesan	Mengandung konsep-	Algoritmik
	konsep simbolik	
Sifat Input	Bisa tidak lengkap	Harus lengkap
Pencarian	Kebanyakan bersifat	Biasanya didasarkan
	heuristic	pada algoritma
Keterangan	Disediakan	Biasanya tidak
		disediakan
Fokus	Pengetahuan	Data dan Informasi
Struktur	Kontrol dipisahkan dari	Kontrol terintegrasi
	pengetahuan	dengan informasi
Kemampuan	Ya	Tidak
menalar		

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konvensional

RapidMiner

Apa itu RapidMiner?

RapidMiner merupakan perangakat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan munggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya dikembangkan pada tahun mulai 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan mengunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML)Extensible Markup Language) yang mendefenisikan proses analitis keingginan pengguna untuk diterpkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analis secara otomatis. RapidMiner memiliki beberapa sifat sebagai berikut:

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukaran data.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari RapidMiner, antara lain:

- Banyaknya algoritma data mining, seperti decision treee dan self-organization map.
- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation,

loading), data preprocessing, visualisasi, modelling dan evaluasi

- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI
- Mengintegrasikan proyek data mining Weka dan statistika R

Instalasi Software

System Requirement

Sebelum melakukan instalasi software RapidMiner, terdapat beberapa spesifikasi minimal dimiliki komputer harus vang vang pengguna. Spesifikasi minimal bergantung pada komputer dan sistem operasi yang akan diinstal. Berikut ini beberapa spesifikasi minimal dibutuhkan software yang RapidMiner:

1. Sistem Operasi

RapidMiner merupakan software yang multiplatform, sehingga software ini dapat dijalankan pada berbagai sistem operasi. Berikut ini beberapa jenis sistem operasi yang dapat diinstal RapidMiner:

- ✓ Microsoft Windows (x86-32) → Windows XP, Windows Server 2003, Windows Vista, Windows Server 2008, Windows 7
- ✓ Microsoft Windows (x64) → Windows XP untuk x64, Windows Server 2003 untuk x64, Windows Vista untuk x64, Windows Server 2008 untuk x64, Windows 7 untuk x64
- ✓ Unix sistem 32 atau 64 bit
- ✓ Linux sistem 32 atau 64 bit
- Apple Macintosh sistem 32 atau 64 bit
 Sebagai bahan pertimbangan, kami merekomendasikan untuk penggunaan sistem
 64 bit. Hal ini dikarenakan jumlah maksimum yang dapat digunakan oleh RapidMiner terbatas pada sistem operasi dengan sistem
 32, yaitu hanya sebasar 2GB.
- 2. Java Runtime Environment versi 6

Selain itu, penggunaan server RapidAnalytics dalam kombinasi dengan RapidMiner dapat memaksimalkan proses analisis pada RapidMiner, meskipun tugas analisis sudah banyak dapat dijalankan dengan RapidMiner desktop client. Dalam hal ini proses analisa dirancang dengan RapidMiner, kemudian dieksekusi oleh server RapidAnalytics.

Instalasi RapidMiner

Seperti yang yang telah dikemukakan sebelumnya bahwa RapidMiner merupakan software gratis yang bersifat terbuka (open source). Software ini dapat dijalankan pada sistem operai Windows, Linux, maupun Mac. RapidMiner dapat diunduh pada situs resminya, yaitu <u>www.rapid-i.com</u>. Pada bagian ini, akan dijelaskan bagaimana cara melakukan instalasi software RapidMiner versi 5.3 pada sistem operasi Microsoft Windows.

Untuk memulai instalasi software RapidMiner pada sistem operasi Microsoft Windows, jalankan file installer RapidMiner-5.3.000x32-install.exe, sehingga akan muncul tampilan wizard seperti pada Gambar 2.



Gambar 2.1 Form Awal Instalasi

Klik **Next >** untuk melanjutkan pada form persetujuan dan lisensi seperti pada Gambar 2.3



Gambar 2.2 Form Persetujuan Lisensi

Pilih **I Agree** untuk melanjutkan. Kemudian, wizard akan menampilkan form seperti pada gambar 2.4.

Choose Install Location		R
Choose the folder in which to install RapidMir	ner 5.	U.
Please select the folder where RapidMiner 5 administrator rights you should select a direc your home directory. Installing into a writabi use the automatic update service of RapidMi	should be installed. If yo tory into which you can e directory is especially iner 5 without having ad	ou are a user without write, e.g. a directory in important if you want to ministrator rights.
Destination Folder		Browse
Space required: 187.7MB Space available: 63.8GB		

Gambar 2.3 Form Pemilihan Lokasi Instalasi

Pilih **Install** untuk melakukan proses instalasi. Kemudian wizard akan menampilkan progress dari proses tersebut, seperti yang ditunjukkan pada Gambar 2.5.

③ RapidMiner 5 Setup	
Installation Complete Setup was completed successfully.	
Completed	
Show details	
	Carrel
	Carca

Gambar 2.4 Form Proses Instalasi

Setelah proses selesai, pilih **Next** > untuk melanjutkan, maka wizard akan menampilkan informasi bahwa proses instalasi telah selesai dilakukan, seperti yang ditunjukkan pada Gambar 2.6.



Gambar 2.5 Form Instalasi selesai

Pilih Finish untuk mengakhiri proses instalasi.

Pengenalan Interface

RapidMiner menyediakan tampilan yang *user friendly* untuk memudahkan penggunanya ketika menjalankan aplikasi. Tampilan pada RapidMiner dikenal dengan istilah Perspective. Pada RapidMiner terdapat 3 Perspective, yaitu; Welcome Perspective, Design Perspective dan Result Perspective.

Welcome Perspective

Ketika membuka aplikasi Anda akan disambut dengan tampilan yang disebut dengan Welcome Perspective, seperti yang ditunjukkan pada Gambar 6. Pada bagian toolbar, terdapat toolbar **Perspectives** yang terdiri dari ikon-ikon untuk menampilkan perspective dari RapidMiner. Toolbar ini dapat dikonfigurasikan sesuai dengan kebutuhan Anda. Sedangkan **Views** menunjukkan pandangan (view) yang sedang Anda tampilkan.



Gambar 2.6 Tampilan Welcome Perspective

Jika komputer Anda terhubung dengan internet, maka pada bagian bawah Welcome Perspective akan menampilkan berita terbaru mengenai RapidMiner. Bagian ini dinamakan **News**. Pada bagian tengah halaman terlihat daftar **Last Processes** (Recent Processes), bagian ini menampilkan daftar proses analisis yang baru saja dilakukan. Hal ini akan memudahkan Anda jika ingin melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut. Bagian **Actions** menunjukkan daftar aksi yang dapat Anda lakukan setelah membuka RapidMine. Berikut ini rincian lengkap daftar aksi tersebut:

- New : Aksi ini berguna ntuk memulai proses analis baru. Untuk memulai proses analisis, pertama-tama Anda harus menentukan nama dan lokasi proses dan Data Repository. Setelah itu, Anda bisa mulai merancang sebuah analisis baru.
- 2. Open Recent Process : Aksi ini berguna untuk membuka proses yang baru saja ditutup. Selain aksi ini, Anda juga bisa membuka proses yang baru ditutup dengan mengklik dua kali salah satu daftar yang ada pada Recent Process. Kemudian tampilan Welcome Perspective akan otomotasi beralih ke Design Perspective.
- Open Process : Aksi ini untuk membuka Repository Browser yang berisi daftar proses. Anda juga bisa memilih proses untuk dibuka pada Design Perspective.
- 4. **Open Template** : Aksi ini menunjukkan pilihan lain yang sudah ditentukan oleh proses analisis.

5. Online Tutorial : Aksi digunakan untuk memulai tutorial secara online (terhubung internet). Tutorial yang dapat secara langsung digunakan dengan RapidMiner ini, memberikan perkanalan dan beberapa data mining. konsep Hal ini direkomendasikan untuk Anda yang sudah memiliki pengetahuan dasar mengenai data mining dan sudah akrab dengan operasi dasar RapidMiner.

RapidMiner dapat menampilkan beberapa view pada saat bersamaan. Seperti yang ditunjukkan pada Gambar 7, pada tampilan Welcome Perspective tedapat **Welcome view** dan **Log View**. Ukuran dari setiap view tersebut dapat diubah sesuai dengan kebutuhan Anda dengan Mengklik dan menarik garis batas diantara keduanya ke atas atau ke bawah.



Gambar 2.7 Welcome Perspective



Gambar 2.8 Header Tab

Anda bisa melakukan beberapa aksi terhadap view, dengan mengklik salah satu ikon yang tampak pada bagian view, seperti yang ditunjukkan pada gambar 2.8. Berikut ini beberapa aksi yang dapat Anda lakukan:

- Close : Aksi ini untuk menutup view yang ditampilkan pada perspective. Anda bisa menampilkan view kembali dengan mengklik menu view dan memilih view yang ingin ditampilkan.
- 2. *Maximize* : Aksi ini untuk memperbesar ukuran view pada perspective.
- 3. *Minimize* : Aksi ini untuk memperkecil ukuran view pada perspective.
- Detach : Aksi ini untuk melepaskan view dari perspective menjadi jendela terpisah, kemudian Anda juga dapat memindahkannya sesuai dengan keinginan Anda.

Design Perspective

Design Perspective merupakan lingkungan kerja RapidMiner. Dimana Design Perspective ini merupakan perspective utama dari RapidMiner yang digunakan sebagai area kerja untuk membuat dan mengelola proses analisis. Seperti yang ditunjukkan pada Gambar 2.10, perspective ini memiliki beberapa view dengan fungsinya masing-masing yang dapat mendukung Anda dalam melakukan proses analisis data mining. Anda bisa mengganti perspective dengan mengklik salah satu ikon dari tollbar perspective yang sebelumnya telah dijelaskan. Selain dengan cara tersebut, Anda juga bisa mengganti perspective dengan mengklik menu view, kemudian pilih perspective, lalu pilih perspective yang ingin Anda tampilkan.



Gambar 2.9 Tampilan Design Perspective

Sebagai lingkungan lingkungan kerja, Design Perspective memiliki beberapa view. Berikut ini beberapa view yang ditampilkan pada Design Perspective:

1. Operator View

Operator View merupakan view yang paling penting pada perspective ini. Semua operator atau langkah kerja dari RapidMiner disajikan dalam bentuk kelompok hierarki di Operator View ini sehingga operator-operator tersebut dapat digunakan pada proses analisis, seperti yang ditunjukkan pada Gambar 2.10. Hal ini akan memudahkan Anda dalam mencari dan menggunakan operator yang sesuai dengan kebutuhan Anda. Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut:

- Process Control : Operator ini terdiri dari operator perulangan dan percabangan yang dapat mengatur aliran proses.
- Utility : Operator bantuan, seperti operator macros, loggin, subproses, dan lain-lain.
- Repository Access : Kelompok ini terdiri dari operator-operator yang dapat digunakan untuk membaca atau menulis akses pada repository.
- Import : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk membaca data dan objek dari format tertentu seperti file, database, dan lain-lain.
- Export : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk menulis data dan objek menjadi format tertentu.

- Data Transformation : kelompok ini terdiri dari semua operator yang berguna untuk transformasi data dan meta data.
- Modeling : kolompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru.
- Evaluation : kelompok ini berisi operator yang dapat digunakan untuk menghitung kualitas pemodelan dan untuk data baru.



Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki

2. Repository View

Repository View merupakan komponen utama dalam Design Perspective selain Operator View. View ini dapat Anda gunakan untuk mengelola dan menata proses Analisis Anda menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.

3. Process View

menunjukkan langkah-langkah View Process tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Anda dapat menambahkan langkah baru dengan beberapa cara. hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali. bekerja dengan RapidMiner ialah Pada dasarnya mendefinisikan proses analisis. yaitu dengan menunjukkan serangkaian langkah kerja tertentu. Dalam RapidMiner, komponen proses ini dinamakan RapidMiner sebagai operator. Operator pada didefinisikan sebagai beikut:

- * Deskripsi dari input yang diharapkan.
- * Deskripsi dari output yang disediakan.
- Tindakan yang dilakukan oleh operator pada input, yang akhirnya mengarah dengan penyediaan output.
- Sejumlah parameter yang dapat mengontrol action performed.

4. Parameter View

Beberapa operator dalam RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar. Namun terkadang parameter tidak mutlak dibutuhkan, meskipun eksekusi operator dapat dikendalikan dengan menunjukkan nilai parameter tertentu. Parameter view memiliki toolbar sendiri sama seperti view-view yang lain. Pada Gambar 2.12, Anda dapat melihat bahwa pada Parameter View ini terdapat beberapa ikon dan nama-nama operator terkini yang dikuti dengan aktual parameter.



Gambar 2.11 Tampilan Parameter View

Huruf tebal berarti bahwa parameter mutlak harus didefinisikan oleh analis dan tidak memiliki nilai default. Sedangkan huruf miring berarti bahwa parameter diklasifikasikan sebagai parameter ahli dan seharusnya tidak harus diubah oleh pemula untuk analisis data. Poin pentingnya ialah beberapa parameter hanya ditunjukkan ketika parameter lain memiliki nilai tertentu.

5. Help & Comment View

Setiap kali Anda memilih operator pada Operator View atau Process View, maka jendela bantuan dalam Help View akan menunjukkan penjelasan mengenai operator ini. Penjelasn yang ditampilkan dalam Help View meliputi:

- Sebuah penjelasan singkat mengenai fungsi operator dalam satu atau beberapa kalimat.
- Sebuah penjelasan rinci mengenai fungsi operator.
- Daftar semua parameter termasuk deskripsi singkat dari parameter, nilai default (jika tersedia), petunjuk apakah parameter ini adalah parameter ahli serta indikasi parameter dependensi.

Sedangkan Comment View merupakan area bagi Anda untuk menuliskan komentar pada langkahlangkah proses tertentu. Untuk membuat komentar, Anda hanya perlu memilih operator dan menulis teks di atasnya dalam bidang komentar. Kemudian komentar tersebut disimpan bersama-sama dengan definisi proses Anda. Komentar ini dapat berguna untuk melacak langkah-langkah tertentu dalam rancangan nantinya.

6. Problem & Log View

Problem View merupakan komponen yang sangat berharga dan merupkan sumber bantuan bagi Anda selama merancang proses analisis. Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam Problem View, seperti yang ditunjukkan pada Gambar 2.13.



Gambar 2.12 Problem & Log View

Pada kolom Message, Anda akan menemukan ringkasan pendek dari masalah. Kolom Location berisi tempat di mana masalah muncul dalam bentuk nama Operator dan nama port input yang bersangkutan. Kolom Fixes memberikan gambaran dari kemungkinan solusi tersebut, baik secara langsung sebagai teks (jika hanya ada satu kemungkinan Solusi) atau sebagai indikasi dari berapa banyak kemungkinan yang berbeda untuk memecahkan masalah.

Cara Menggunakan Repositori

Repositori merupakan Tabel, database, koleksi teks, yang kita miliki untuk dapat digali datanya untuk mendapatkan informasi yang kita inginkan. Ini merupakan awal dari seluruh proses Data Mining. Maka dari itu adalah penting bagi kita untuk mengetahui cara menggunakan repository.

Sample Data Repository

RapidMiner menyediakan contoh database yang dapat digunakan, berikut cara menggunakan Sample Data Repository.



Gambar 2.13 Kumpulan Sample Data Repository

Pada bagian Repositori terdapat 3 buah lokasi repositori, yakni Samples, DB dan Local Repository.
Untuk mengambil Sample Data Repository, buka hirarki Samples, masuk ke folder Data. Sehingga seperti gambar berikut.



Gambar 2.14 Tampilan Design Perspective Awal

Lakukan Drag dan Drop salah satu Example Repository. Kita ambil contoh Golf. Tarik dan lepaskan repository ke dalam Main Process, sehingga seperti gambar berikut.



Gambar 2.15 Repository berada dalam Main Process



Gambar 2.16 Menghubungkan Output Repositori ke Result

Hubungkan output pada Database ke Result seperti Gambar diatas. Lalu klik ikon Play ▶. Gambar 2.17 adalah Sample data repository dari Golf. Coba lakukan untuk memasukkan Sample Repository yang lain.

🛛 🐺 Resi	ult Overview	🗙 🖊 📳 Exa	mpleSet (Retri	eve Golf) 🏾	
Data Vie	w 🔘 Meta D	ata View 🔘 F	Plot View 🔘 A	dvanced Cha	rts 🔘 Annotatior
ExampleSe	t (14 example	es, 1 special at	tribute, 4 regul	ar attributes)	
Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

Gambar 2.17 Isi Sample Golf Data Repository

Import Repository

Dibanyak kesempatan lain, kita akan selalu menggunakan database yang kita miliki. RapidMiner menyediakan layanan pengguna agar dapat mengimport database miliknya. Namun, tidak seperti kebanvakan tools Data Mining Lain. RapidMiner memiliki kelebihan tersendiri yakni dapat langsung melakukan import file dengan ekstensi .xls atau .xlsx, yakni file dari Microsoft Excel, Program yang relatif sering digunakan oleh pengguna. Berikut adalah cara untuk melakukan import file Microsoft Excel.

Lihat pada bagian Repository. Klik pada ikon import seperti gambar 2.18. Seperti yang dapat kita lihat, ada beberapa ekstensi file yang dapat kita masukkan kedalam repository kita. CSV File, Excel Sheen File, Access Database Table File, Database Table, Binary File. Namun pada Dasarnya cara melakukan import pada semua file ini sama. Sebagai contoh, pilih Import Excel Sheet.



Gambar 2.18 Repository

Setelah itu, akan muncul window baru yakni Step 1 dari 5 Step Data import Wizard. Disini akan diarahkan oleh RapidMiner bagaimana langkah untuk melakukan import data.



Gambar 2.19 Step 1 of 5 Import Wizard

Cari file excel kalian dengan klik pada bagian Look in

Look In: Documents

yang dibutuhkan lalu Klik tombol Next 🗩 🕬.

Berikutnya pada Step 2 ialah, pilih Sheet yang akan dimasukkan. Pada dasarnya, Repository RapidMiner hanya menyediakan 1 repositori untuk 1 buah table.

-		Data import wizard - Step 2 of 5						
"At	This wizard guides you to import your data. Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into RapidMiner. Furthermore, you can mark a range of cells to be loaded.							
Sheet1								
A	В	С	D	E				
OUTLOOK	TEMPERATU							
SUNNY								
SUNNY								
OVERCAST								
RAIN			false					
RAIN			false					
RAIN								
OVERCAST								
SUNNY								
SUNNY								
RAIN								
	Previous Next Prish X Cancel							

Gambar 2.20 Step 2 of 5 Import Wizard

Klik tombol Next 🖃 🕬. Berikutnya ialah memberikan anotasi. Jika data kita tidak memiliki nama attribute, tidak usah melakukan apa-apa pada step 3 ini.

Data import wizard - Step 3 of 5 This wizard guides you to import your data. Step 3: In RapidNiner, each attribute can be annotated. The most important annotation data attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.						
A	В	С	D	E		
OUTLOOK	TEMPERATU	HUMIDITY	WINDY	PLAY		
SUNNY	85	85	false	Don't Play		
SUNNY	80	90	true	Don't Play PLAY		
OVERCAST	83	78	false	Play		
RAIN	70	96	false	Play		
RAIN	68	80	false	Don't Play		
RAIN	65	70	true	Play		
OVERCAST	64	65	true	Don't Play	1	
SUNNY	72	95	false	Play		
SUNNY	69	70	false	Play	L	
ΡΔΙΝΙ	75	80	folco	Play	$\mathbf{\nabla}$	
	annotation of names of the his property. can assign th A DUTLOOK SUNNY SUNNY OVERCAST RAIN RAIN OVERCAST SUNNY SUNNY SUNNY RAIN	anotation of an attribute is attributes it y his property. If further anno can assign them here.	annotation of an attribute is its name - an attributes. If your data does this property if further annotations are color an assign them here. A B C JUTLOOK TEMPERATI. HUMIDITY SUNNY 85 85 SUNNY 80 90 OVERCAST 83 78 RAIN 70 96 RAIN 65 70 OVERCAST 64 65 SUNNY 72 95 SUNNY 76 90	anotation of an attribute is its name - a row with this 2 manes of the attributes. If your data does not contain a his property. If further anotations are contained in the can assign them here. A B C D OUTLOOK TEMPERATL HUMIDITY WINDY SUNNY 85 85 false SUNNY 85 90 true OVERCAST 83 78 false RAIN 70 96 false RAIN 65 70 true OVERCAST 64 65 true SUNNY 72 95 false SUNNY 72 95 false SUNNY 72 97 false SUNNY 75 90 false	anotation of an attribute is its name - a row with this annotation defines the tames of the attributes. If your data does not contain attribute marks, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.	

Gambar 2.21 Step 3 of 5 Import Wizard

Klik tombol Next 🗩 . Step ke 4 adalah memberikan tipe data pada tabel kita. Sebenarnya Rapid Miner akan memberikan tipe data yang tepat secara otomatis.

Beload data	Duess value types	Preview uses only	trist 100 rows. 🖸 ble for	mat
2	ø	Ø	Ø	Ø
OUTLOOK	TENPERATURE	HUVIDITY	WINDY	PLAY
polynominal +	integer +	Integer +	binominal +	binominzi =
atribult +	attribute +	attribute +	attribute +	attribute +
SUNNY	85	80	talse	Con 1 Play
SUNW	80	90	tue	Dom't Play
OVERCAST	83	70	Salare	Pter .
RAN	70	96	talse	Play
RAN	60	80.	taise	Don't Play
0				D
😳 0 enors.			1	🛛 Jancre errors 🔲 Show only gro
Row Cours	1	Error	Original value	Massage

Gambar 2.22 Step 4 of 5 Import Wizard

Namun, jika kita merasa tipe data yang diberikan RapidMiner tidak cocok, kita bisa mengubahnya.



Gambar 2.23 Tipe Data

Klik tombol Next 🕬 Step ke 5 adalah memasukkan database kita kedalam repository. Disarankan untuk memasukkannya kedalam Local Repository untuk memudahkan kita mencarinya. Jangan lupa untuk memberikan nama repository kita.



Gambar 2.24 Step 5 of 5 Import Wizard

Kemudian klik tombol finish Prinish

Hasil Import Repository akan terlihat pada bagian Repository seperti dalam gambar 2.25.



Gambar 2.25 Repository yang sudah diimport

Untuk melihat isi dari repository kita, hubungkan output pada repository kearah result seperti gambar 2.26.



Gambar 2.26 Menghubungkan Output Repositori pada Result

klik ikon Play 🖻. Dan berikutnya akan muncul isi dari tabel yang kalian miliki.

🛛 🐺 Resu	lt Overview 🖇	👔 🗐 Exan	npleSet (//Loc	al Repository/	data/golfdatası	et) 🕺 🔪
Data Vier	w 🔘 Meta Da	ita View 🔘 P	lot View 🔘 A	dvanced Chai	ts 🔘 Annotat	ions
ExampleSet	(14 examples	, 0 special att	ributes, 5 regu	ular attributes)		
Row No.	OUTLOOK	TEMPERAT	HUMIDITY	WINDY	PLAY	
1	SUNNY	85	85	false	Don't Play	
2	SUNNY	80	90	true	Don't Play	
3	OVERCAST	83	78	false	Play	
4	RAIN	70	96	false	Play	
5	RAIN	68	80	false	Don't Play	
6	RAIN	65	70	true	Play	
7	OVERCAST	64	65	true	Don't Play	
8	SUNNY	72	95	false	Play	
9	SUNNY	69	70	false	Play	
10	RAIN	75	80	false	Play	
11	SUNNY	75	70	true	Play	
12	OVERCAST	72	90	true	Play	
13	OVERCAST	81	75	false	Play	
14	RAIN	71	80	true	Don't Play	
14	RAIN	/1	00	uue	DUILTIAY	

Gambar 2.27 Tabel Repository

Bagian Dua Data Mining

Pengenalan Data Mining

Pengenalan Decision Tree

Pengenalan Neural Network

Pengenalan Market Basket Analysis

Chapter 3

Data Mining

Mengenal Data Mining

Pengertian Data Mining

Sebelum kita mulai, ayo kita coba beberapa eksperimen sebagai berikut.

- Pilih angka antara 1 sampai 10
- Kalikan dengan angka 9
- Hasil dari perkalian tersebut jumlahkan masingmasing angkanya
- Kalikan hasil dengan 4
- Bagi dengan 3
- Kurangi dengan 2

Jawabannya adalah 2. Kebetulan? Sebagai seorang analis, pasti jawabannya adalah tidak.

Bagaimana dengan kejadian acak lainnya, seperti "lempar koin." Tentu jika temanmu menebak secara langsung dan hasil dari kejadian tersebut ternyata tepat seperti yang temanmu tebak, kau pasti akan mengatakan bahwa itu merupakan kebetulan.

Kita ambil satu contoh sederhana lagi. Terdapat kejadian seperti: Seseorang menjatuhkan sebuah gelas dari ketinggian tertentu. Detik pertama orang tersebut menjatuhkan gelasnya, kau pasti akan mengatakan dengan pasti bahwa gelas tersebut akan pecah, padahal hukum fisika belum menunjukkan proses penghancuran gelas tersebut ketika bersentuhan dengan tanah. Dan lagi, tebakanmu itu dikatakan bukanlah kebetulan. Jadi secara logika, bagaimana kau tahu dengan sangat tepat hasil dari kejadian tersebut? Bukankah kondisinya sama seperti kejadian "lempar koin" sebelumnya?

Jadi apakah yang kita lakukan dalam otak kita? Kita mempertimbangkan karakteristik-karakteristik dari kejadian ini. Pada kasus gelas yang jatuh, kita dengan cepat mengetahui karakteristik penting dari serangkaian kejadian tersebut, bahan gelas, ketinggian, tipe pijakan, dan lain-lain. Kemudian kita menjawab dengan cepat berdasarkan analogi, contohnya kita kita

membuat perbandingan dengan kejadian gelas atau cangkir atau piring yang jatuh sebelumnya. Berarti dua diperlukan hal yang adalah: pertama, kita membutuhkan data dari kejadian-kejadian sebelumnya, dan kedua, seberapa mirip kejadian yang di tempat dengan kejadian sebelumnya. Kita bisa membuat estimasi atau prediksi dengan mencari kejadian yang paling mirip dengan kejadian di tempat. Karena kita lebih sering melihat bahwa benda berbahan kaca dijatuhkan akan pecah, maka secara otomatis inilah yang menjadi prediksi kita.

Bagaimanapun, prosedur diatas tidak cocok untuk kejadian "lempar koin." Ini disebabkan terdapat lebih banyak faktor yang harus dipertimbangkan, ada yang sulit dan ada yang tidak bisa diukur. Belum lagi kita harus dapat memikirkan proses kejadian menuju hasil dengan baik, memikirkan analogi yang paling cocok dengan kejadian untuk melakukan prediksi. Ditambah "lempar koin" memiliki kondisi yang dapat berubah-ubah tiap kejadiannya dan berlangsung cepat, ini berarti perhitungan juga harus dilakukan secara cepat. Mustahil untuk seorang manusia? Benar. Tetapi tidak mustahil untuk metode data mining.

Data Mining adalah serangkaian proses untuk menggali nlai tambah dari suatu kumpulan data

berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006)

Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya. (Pramudiono, 2006)

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005)

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. (Larose, 2005)

Kata *Mining* merupakan kiasan dari bahasa inggris, mine. Jika mine berarti menambang sumber daya yang tersembunyi di dalam tanah, maka Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), machine learning, statistik dan basis Data.

Pengelompokan Teknik Data Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan popular adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur herhirarki

Association

Digunakan untuk mengenali kelakuan dari kejadiankejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya adalah Market Basket Analysis, yaitu salah sati metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan.

Clustering

Digunakan menganalisis pengelompokkan untuk berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokkan belum didefinisikan sehelum dijalankannya tool data mining. Biasanya menggunkan metode *neural network* atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool data mining.

Decision Tree

Mengenal Decision Tree

diketahui bahwa selalu Seperti manusia menghadapi berbagai macam masalah dalam di sehari-hari. kehidupannya Masalah-masalah yang timbul dari berbagai macam bidang ini memiliki tingkat kesulitan dan kompleksitas yang sangat bervariasi, mulai dari masalah yang sangat sederhana dengan sedikit faktor-faktor terkait hingga masalah yang sangat rumit dengan banyak sekali faktor-faktor yang terkait, sehingga factor-faktor yang berkaitan dengan masalah tersebut perlu untuk diperhitungkan.

Seiring dengan perkembangan kemajuan pola pikir manusia, manusia mulai mengembangkan sebuah sistem yang dapat membantu manusia dalam menghadapi masalah-masalah yang timbul sehingga dapat menyelesaikannya dengan mudah.

keputusan atau yang lebih dikenal Pohon dengan istilah Decision Tree ini merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebutdengan memperhitungkan berbagai macam faktor yang berkaitan di dalam lingkup masalah tersebut.

Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah sehingga dengan memperhitungkan faktor-faktor tersebut dapat dihasilkan penyelesaian terbaik untuk masalah tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah.

Pohon keputusan dalam analisis pemecahan masalah pengambilan keputusan merupakan pemetaan alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon keputusan juga memperlihatkan faktor-faktor kemungkinan yang dapat mempengaruhi alternative-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut.

Secara umum, pohon keputusan adalah suatu gambaran permodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah kepada solusi yang dihasilkan. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Seiring dengan perkembangannya, pohon keputusan kini telah banyak dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang prediksi. (Han, J., & Kamber, M. (2006). Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.)



Gambar 4.1 Bentuk Decision Tree Secara Umum

Algoritma c4.5

Pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Seperti yang telah dijelaskan sebelumnya, klasifikasi merupakan Suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Metode ini popular karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Banyak algoritma yang dapat digunakan untuk membangun suatu decision tree, salah satunya ialah algoritma C45.

Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C.45 menggunakan rasio perolehan (gain ratio). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropi.

Konsep Entropy

Entropy(S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropi digunakan untuk mengukur ketidakaslian S.



Gambar 4.2 Grafik Entropi

Besarnya Entropy pada ruang sampel S didefinisikan dengan:

 $Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$

Dimana:

- S : ruang (data) sampel yang digunakan untuk pelatihan
- p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu
- p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

- Entropi(S) = 0, jika semua contoh pada S berada dalam kelas yang sama.
- Entropi(S) = 1, jika jumlah contoh positif dan negative dalam S adalah sama.
- 0 > Entropi(S) > 1, jika jumlah contoh positif dan negative dalam S tidak sama.

Konsep Gain

Gain (S,A) merupakan Perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

$$Gain(S, A) \equiv Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$$

Dimana:

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- |S_i| : Jumlah sampel pada pertisi ke --i
- |S| : Jumlah sampel dalam S

Untuk memudahkan penjelasan mengenai algoritma C4.5berikut ini disertakan contoh kasus yang dituangkan dalam Tabel 4.1:

No	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Tabel 4.1 Keputusan Bermain Tenis

Tabel 1 merupakan kasus yang akan dibuat pohon keputusan untuk menentukan main tenis atau tida. Data ini memiliki atribut-atribut yaitu, keadaan cuaca (outlook), temperatur, kelembaban (humidity) dan keadaan angin (windy).

Berikut merupakan cara membangun pohon keputusan dengan menggunakan algoritma:

- 1. Pilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut-atribut yang ada.
- 2. Buat cabang untuk masing-masing nilai
- 3. Bagi kasus dalam cabang
- 4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

NODE			JUMLAH KASUS	NO (S ₁)	YES (S ₂)	ENTROPY	GAIN
1	TOTAL		14	4	10	0.863120569	
	OUTLOOK						0.258521037
		CLOUDY	4	0	4	0	
		RAINY	5	1	4	0.721928095	
		SUNNY	5	3	2	0.970950594	
	TEMPERATURE						0.183850925
		COOL	4	0	4	0	
		HOT	4	2	2	1	
		MILD	6	2	4	0.918295834	
	HUMIDITY						0.370506501
		HIGH	7	4	3	0.985228136	
		NORMAL	7	0	7	0	
	WINDY						0.005977711
		FALSE	8	2	6	0.811278124	
		TRUE	6	4	2	0.918295834	

Tabel 4.2 Perhitungan Simpul 1



Dari hasil pada Tabel 4.2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah HUMIDITY yaitu sebesar 0.37. Dengan demikian HUMIDITY dapat menjadi node akar.

Ada 2 nilai atribut dari HUMIDITY yaitu HIGH dan NORMAL. Dari kedua nilai atribut tersebut, nilai atribut NORMAL sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya Yes, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut HIGH masih perlu dilakukan perhitungan lagi hingga semua kasus masuk dalam kelas seperti yang terlihat pada Gambar di sebelah kanan.

Kelebihan Pohon Keputusan

Dalam membuat keputusan dengan menggunankan pohon keputusan, metode ini memiliki kelebihan sebagai berikut:

- Daerah pengambilan keputusan lebih simpel dan spesifik.
- Eliminasi perhitungan-perhitungan tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
- Fleksibel untuk memilih fitur dari internal node yang berbeda. Sehingga dapat meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
- Dengan menggunakan pohon keputusan, penguji tidak perlu melakukan estimasi pada distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Karena metode ini menggunakan kcriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Kekurangan Pohon Keputusan

Pohon keputusan sangat membantu dalam pengambilan keputusan, namun pohon keputusan juga memiliki beberapa kekurangan, diantaranya:

- Kesulitan dalam mendesain pohon keputusan yang optimal.
- Hasil kualitas keputusan yang didapat sangat tergantung pada bagaimana pohon tersebut didesain. Sehingga jika pohon keputusan yang dibuat kurang optimal, maka akan berpengaruh pada kualitas dari keputusan yang didapat.
- Terjadi overlap terutama ketika kelas-kelas dan criteria yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
- Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.

Decision Tree pada RapidMiner

RapidMiner sebagai software pengolah data mining menyediakan tool untuk membuat decision tree. Hal ini tentu akan memudahkan kita membuat decision tree dengan menggunakan RapidMiner dibandingkan membuat decision tree secara manual yaitu dengan melakukan perhitungan menggunakan algoritma C4.5 yang telah dijelaskan sebelumnya.

Contoh Kasus:

Keputusan Bermain Tenis

Pada contoh kali ini, kita akan membuat keputusan bermain tenis atau tidak. Untuk memudahkan dalam menggunakan RapidMiner untuk membuat decision tree, kita gunakan data sederhana yang ada pada sub bab decision tree. Pertama-tama data pada tabel 2 dibuat lagi dalam format excel seperti yang terlihat pada Gambar 4.3.

	А	В	С	D	E	F
1	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY	
2	Sunny	Hot	High	No	Don't Play	
3	Sunny	Hot	High	Yes	Don't Play	
4	Cloudy	Hot	High	No	Play	
5	Rainy	Mild	High	No	Play	
6	Rainy	Cool	Normal	No	Play	
7	Rainy	Cool	Normal	Yes	Play	
8	Cloudy	Cool	Normal	Yes	Play	
9	Sunny	Mild	High	No	Don't Play	
10	Sunny	Cool	Normal	No	Play	
11	Rainy	Mild	Normal	No	Play	
12	Sunny	Mild	Normal	Yes	Play	
13	Cloudy	Mild	High	Yes	Play	
14	Cloudy	Hot	Normal	No	Play	
15	Rainy	Mild	High	Yes	Don't Play	
16						

Gambar 4.3 Tabel Keputusan dalam Format xls

Setelah data yang kita punya dibuat dalam bentuk tabel format xls, selanjutnya lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada Bab 2. Lalu cari table Microsoft Excel yang telah dibuat dan masukan kedalam Local Repository seperti yang terlihat pada Gambar 4.4.



Gambar 4.4 Lokasi Tabel pada Repository

Lakukan Drag dan Drop Tabel PlayGolf kedalam Process view. Sehingga Operator Database muncul dalam View Proses seperti pada Gambar 4.5. Pada view Process, tabel PlayGolf yang dimasukkan ke dalam proses akan dijadikan sebagai Operator Retrieve.



Gambar 4.5 Repository PlayGolf pada Main Process

Untuk membuat decision tree dengan menggunakan RapidMiner, kita membutuhkan operator Decision tree, operator ini terdapat pada View Operators. Untuk menggunakannya pilih Modelling pada View Operator, lalu pilih Classification and Regression, lalu pilih Tree Induction dan pilih Decision Tree.



Gambar 4.6 Daftar Operator pada View Operators

Setelah menemukan operator Decision Tree, seret (*drag*) operator tersebut lalu letakkan (*drop*) ke dalam view Process. Kemudian susun posisinya disamping operator Retrieve, seperti yang tampak pada Gambar 4.7.



Gambar 4.7 Posisi Operator Decision Tree

Selanjutnya, hubungkan operator Retrieve dengan operator Decision Tree dengan menarik garis dari tabel PlayGolf ke operator Decision Tree dan menarik garis lagi dari operator Decision Tree ke result di sisi kanan, seperti yang tampak pada Gambar 4.8. Operator Decision Tree berguna untuk memperdiksikan keputusan dari atribut-aribut yang dimasukkan ke dalam operator retrieve. Dengan mengubah tabel (atribut) yang dimasukkan menjadi sebuah pohon keputusan.

6° Process X ▼ XML X 4= + ⇒ + ↑ ■ Root >	<i>d</i> • 🗉 🛊 🛛 🕹
inp) Main Process Retrieve Play out DecisionTree tra mod era	() res (res

Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator Decision Tree

Pada operator Decision tree terdapat input training set (tra), port ini merupakan output dari operator retrieve. Output dari operator lain juga dapat digunakan oleh port ini. Port ini menghasilkan ExampleSet yang dapat diperoses menjadi decision tree. Selain itu pada operator ini juga terdapat output model (mod) dan example set (exa). Mod akan mengonversi atribut yang dimasukkan menjadi mpdel keputusan dalam bentuk decision tree. exa merupakan port yang menghasilkan output tanpa mengubah inputan yang masuk melalui port ini. Port ini biasa digunakan untuk menggunakan kembali sama ExampleSet di operator lebih lanjut atau untuk melihat ExampleSet dalam Hasil Workspace.

Langkah selanjutnya ialah mengatur parameter sesuai dengan kebutuhan kita. Setelah menghubungkan operator retrieve dengan operator decision tree, atur parameter decision tree seperti pada gambar 4.9.

🛛 🛃 Parameters 💈	Context 🗶					
🚨 🖷 🖷 🔛 🛛	🕺 📑 👻					
Pecision Tree						
criterion	gain_ratio 💌	Â				
minimal size for split	2					
minimal leaf size	1					
minimal gain	0.1					
maximal depth	20					
confidence	0.25					
number of preprunin	.3					
no pre pruning						
no pruning		~				

Gambar 4.9 Parameter Decision Tree





- Criterion, berguna memilih kriteria untuk menetapkan atribut sebagai akar dari decision tree. kriteria yang dapat dipilih, antara lain
 - Gain ratio merupakan varian dari information_gain. Metode ini menghasilkan information gain untuk

setiap atribut yang memberikan nilai atribut yang seragam

- 2. Information_gain, dengan metode ini, semua entropi dihitung. Kemudian atribut dengan entropi minimum yang dipilih untuk dilakukan perpecahan pohon (split). Metode ini memiliki bias dalammemilih atribut dengan sejumlah besar nilai.
- Gini_index merupakan ukuran ketidakaslian dari suatu ExampleSet. Metode ini memisahkan pada atribut yang dipilih memberikan penurunan indeks gini rata-rata yang dihasilkan subset.
- Accuracy, metode ini memimilih beberapa atribut untuk memecah pohon (split) yang memaksimalkan akurasi dari keseluruhan pohon.
- Minimal size of split, Ukuran untuk membuat simpul-simpul pada decision tree. simpul dibagi berdasarkan ukuran yang lebih besar dari atau sama dengan parameter Minimal size of split. Ukuran simpul adalah jumlah contoh dalam subset nya

- Minimal leaf size, Pohon yang dihasilkan sedemikian rupa memiliki himpunan bagian simpul daun setidaknya sebanyak jumlah minimal leaf size.
- Minimal gain merupakan nilai gain minimal yang ditentukan untuk menghasilkan simpul pohon keputusan. Gain dari sebuah node dihitung sebelum dilakukan pemecahan. Node dipecah jika gain bernilai lebih besar dari Minimal Gain yang ditentukan. Nilai minimal gain yang terlalu tinggi akan mengurangi perpaecahan pohon dan menghasilkan pohon yang kecil. Sebuah nilai yang terlalu tinggi dapat mencegah pemecahan dan menghasilkan pohon dengan simpul tunggal.
- Maximal depth, Parameter ini digunakan untuk • membatasi ukuran Putusan Pohon. Proses ketika generasi pohon tidak berlanjut kedalaman pohon adalah sama dengan kedalaman maksimal. Jika nilainya diatur ke '-1', parameter kedalaman maksimal menempatkan tidak terikat pada kedalaman pohon, pohon kedalaman maksimum dihasilkan. Jika nilainya diatur ke '1 ' maka akan dihasilkan pohon dengan simpul tunggal.
- Confidence, Parameter ini menentukan tingkat kepercayaan yang digunakan untuk pesimis kesalahan perhitungan pemangkasan.
- number of prepruning alternatives. Parameter ini menyesuaikan jumlah node alternatif mencoba untuk membelah ketika split dicegah dengan prepruning pada simpul tertentu.
 - no prepruning, Secara default Pohon Keputusan yang dihasilkan dengan prepruning. Menetapkan parameter ini untuk menonaktifkan benar prepruning dan memberikan pohon tanpa prepruning apapun.
 - no pruning Secara default Pohon Keputusan yang dihasilkan dengan pemangkasan. Menetapkan parameter ini untuk menonaktifkan benar pemangkasan dan memberikan sebuah unpruned

Setelah parameter diatur, klik ikon Run pada toolbar, seperti pada gambar 40 untuk menampilkan hasilnya. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 4.11 Ikon Run

Setelah beberapa detik maka RapidMiner akan menampilkan hasil keputusan pada view Result. Jika kita pilih Graph view, maka akan ditampilkan hasilnya berbentuk pohon keputusan seperti pada gambar 4.12. Hasil pohon keputusan dapat disimpan dengan mengklik save image pada sisi kiri View Result.



Gambar 4.12 Hasil Berupa Graph Pohon Keputusan

Selain menampilkan hasil decision tree berupa graph atau tampilan pohon keputusan, RapidMiner juga menyediakan tool untuk menampilkan hasil berupa teks view dengan mengklik button Text View seperti yang tampak pada Gambar 4.13.

🛛 🛒 Result Overview 🕺 🎧 Tree (Decision Tree) 🕺						
Graph View . Text View: Annotations						
Tree						
HUMIDITY = High						
OUTLOOK = Cloudy: Play {Don't Play=0, Play=2}						
OUTLOOK = Rainy						
WINDY = No: Play {Don't Play=0, Play=1}						
WINDY = Yes: Don't Play {Don't Play=1, Play=0}						
OUTLOOK = Sunny: Don't Play {Don't Play=3, Play=0}						
HUMIDITY = Normal: Play {Don't Play=0, Play=7}						

Gambar 4.13 Hasil Berupa Penjelasan Teks

Contoh Kasus : Keputusan seseorang mempunyai potensi menderita hipertensi

Sebelumnya kita telah mengetahui bagaimana pohon membuat keputusan untuk menentukan bermain tenis dengan menggunakan operator decision tree. Pada pembahasan kali ini kita akan membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi atau tidak. Untuk menambah pengeatahuan kita mengenai kegunaan operator yang ada pada RapidMiner, oleh karena itu untuk membuat pohon keputusan kali ini kita menggunakan operator X-Validation, Apply Model dan Performance. Selain itu, kita juga tetap menggunakan operator decision tree dalam pembuatan pohon keputusan kali ini.

Cara yang digunakan dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi, tidak jauh berbeda dengan cara membuat pohon keputusan yang sebelumnya, yaitu pertama-tama import data ke dalam repository RapidMiner, lalu lakukan drag dan drop data tersebut pada view process untuk mengubah data yang berisi atribut pohon keputusan menjadi operator retrieve. setelah itu, lakukan drag dan drop operator decision tree ke dalam view process dengan cara yang sama seperti penjelasan sebelumnya.

	А	В	С	D	E
1	Usia	Berat	Kelamin	Hipertensi	
2	Muda	Overweight	Pria	Ya	
3	Muda	Underweight	Pria	Tidak	
4	Muda	Average	Wanita	Tidak	
5	Tua	Overweight	Pria	Tidak	
6	Tua	Overweight	Pria	Ya	
7	Muda	Underweight	Pria	Tidak	
8	Tua	Overweight	Wanita	Ya	
9	Tua	Average	Pria	Tidak	
10					

Gambar 4.14 Tabel SakitHipertensi dalam format xls



Gambar 4.15 Lokasi Tabel pada Repository



Gambar 4.16 Tabel SakitHipertensi pada Main Process

Untuk membuat pohon keputsan kali ini kita menggunakan operator X-Validation. Operator ini melakukan validasi silang untuk memperkirakan kinerja statistik operator pembelajaran (biasanya pada set data yang tak terlihat). Operator ini juga digunakan untuk memperkirakan seberapa akurat suatu model yang akan tampil dalam praktek. Operator X-Validasi merupakan operator bersarang yang memiliki dua subproses: training subprocess (subproses percobaan) testing subprocess (subproses dan pengujian). Subproses percobaan digunakan untuk melatih sebuah model. Model yang terlatih kemudian diterapkan dalam subproses pengujian.

Biasanya proses belajar mengoptimalkan parameter model untuk membuat model sesuai dengan data percobaan. Jika kita kemudian mengambil sampel independen dari data pengujian, umumnya model tersebut tidak cocok dengan data percobaan maupun data pengujian. Hal ini disebut dengan istilah 'over-pas', dan sangat mungkin terjadi ketika ukuran set data training kecil, atau ketika jumlah parameter dalam model besar. Sehingga validasi silang merupakan cara untuk memprediksi kesesuaian model untuk satu set pengujian hipotesis ketika set pengujian eksplisit tidak tersedia.

Untuk menemukan operator X-Validation, pilih Evaluation pada View Operator, lalu pilih Validation, lalu pilih X-Validation .Setelah menemukan operator X-Validation, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.17 Hirarki Operator X-Validation



Gambar 4.18 Operator Validation

Operator X-Validation memiliki port input yaitu, training example set (tra) sebagai port input memperkirakan ExampleSet untuk melatih sebuah model (training data set). ExampleSet yang sama akan digunakan selama subproses pengujian untuk menguji model.

Selain itu, operator ini juga memiliki port output sebagai berikut:

- model (mod), Pelatihan subprocess harus mengembalikan sebuah model yang dilatih pada input ExampleSet. Harap dicatat bahwa model yang dibangun ExampleSet disampaikan melalui port ini.
- training example set (tra), The ExampleSet yang sebagai masukan pada port input diberikan pelatihan dilewatkan tanpa mengubah ke output melalui port ini. Port ini biasa digunakan untuk menggunakan kembali ExampleSet sama di operator lebih lanjut atau untuk melihat ExampleSet dalam Workspace Result.
- averagable (ave), subproses pengujian harus mengembalikan Vector Kinerja. Hal ini biasanya dengan menerapkan model dihasilkan dan mengukur kinerjanya. Dua port tersebut diberikan tetapi hanya dapat digunakan jika diperlukan. Harap dicatat bahwa kinerja statistik dihitung dengan skema estimasi hanya perkiraan (bukan perhitungan yang tepat) dari kinerja yang akan dicapai dengan model yang dibangun pada set data vang disampaikan secara lengkap.

🖉 📴 Parameters 🕅	Context 🚿			
2 5 5 7 5	t 📭 🕶			
% Validation	n (X-Validation)			
🖌 average performar	nces only			
leave one out				
number of validations	10			
sampling type	stratified sampling 🔻			
use local random s	seed			
55				
A A A A A A A A A A	^			
Compatiblity level	5.3.000			

Gambar 4.19 Parameter X-Validation

Operator X-Validation juga memiliki parameter yang perlu diatur, diantaranya:

- average performances only (boolean), ini merupakan parameter ahli yang menunjukkan jika vector kinerja harus dirata-ratakan atau semua jenis dari hasil rata-rata.
- leave one out (boolean) Seperti namanya, leave one out validasi silang melibatkan penggunaan satu contoh dari ExampleSet asli sebagai data pengujian (dalam pengujian subproses), dan contoh-contoh yang tersisa sebagai data pelatihan (dalam pelatihan subprocess). Namun hal ini biasanya sangat mahal untuk ExampleSets besar dari sudut

pandang komputasi karena proses pelatihan diulang sejumlah besar kali (jumlah waktu contoh). Jika diatur dengan benar, parameter number of validations dapat diabaikan.

- number of validations (integer), parameter ini menentukan jumlah subset ExampleSet yang harus dibagi (setiap subset memiliki jumlah yang sama dari contoh). Juga jumlah yang sama dari iterasi yang akan berlangsung. Setiap iterasi melibatkan pelatihan model dan pengujian model. Jika ini ditetapkan sama dengan jumlah contoh dalam ExampleSet, Hal ini akan setara dengan operator X-Validasi dengan parameter leave one out set true.
- sampling type (selection), Operator X-Validasi dapat menggunakan beberapa jenis sampling untuk membangun subset. Sampel yang tersedia, diantaranya:
 - linear_sampling, Linear sampling hanya membagi ExampleSet ke partisi tanpa mengubah urutan contoh yaitu subset dengan contoh-contoh berturut-turut diciptakan.
 - shuffled_sampling, Shuffled Sampling membangun subset acak ExampleSet. Contoh dipilih secara acak untuk membuat subset.
 - 3. stratified_sampling, Stratified Sampling membangun subset acak dan memastikan

bahwa distribusi kelas dalam himpunan adalah sama seperti dalam ExampleSet seluruh.

- use local random seed (boolean), Parameter ini menunjukkan jika local random seed harus digunakan untuk mengacak contoh subset. Dengan menggunakan nilai yang sama dengan local random seed maka akan menghasilkan subset yang sama. Mengubah nilai parameter ini mengubah cara contoh menjadi acak, sehingga subset akan memiliki satu set yang berbeda dari contoh. Parameter ini hanya tersedia jika Shuffled atau Stratified sampling dipilih. Hal ini tidak tersedia untuk pengambilan sampel Linear karena tidak membutuhkan pengacakan, contoh yang dipilih secara berurutan
 - Iocal random seed (integer), Parameter ini hanya tersedia jika parameteruse local random seed dipilih. parameter ini menentukan local random seed

Seperti yang telah disebutkan sebelumnya bahwa dalam membuat pohon keputusan pada contoh ini, kita menggunakan operator Apply Model. Operator ini menerapkan suatu model terlatih pada sebuah ExampleSet. Sebuah model pertama kali dilatih di sebuah ExampleSet, informasi yang berkaitan dengan ExampleSet dipelajari oleh model. Maka model tersebut dapat diterapkan pada ExampleSet yang lain dan biasanya untuk prediksi. Semua parameter yang diperlukan disimpan dalam objek model. Ini adalah wajib bahwa kedua ExampleSets harus persis nomor yang sama, order, jenis dan peran atribut. Jika sifat meta data dari ExampleSets tidak konsisten, hal itu dapat menyebabkan kesalahan serius.

Untuk menemukan operator Apply Model, pilih Modeling pada View Operator, lalu pilih Model Application, lalu pilih Confidence dan pilih Apply Model .Setelah menemukan operator Apply Model, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.20 Hirarki Operator Apply

Operator ini memiliki port input yaitu, **model** (**mod**) port ini mengharapkan model. Port ini harus memastikan bahwa nomor, order, jenis dan peran atribut dari ExampleSet pada model yang dilatih konsisten dengan ExampleSet pada port input data unlabeled. **unlabelled data (unl)** port ini mengharapkan suatu ExampleSet. Ini harus memastikan bahwa nomor, order, jenis dan peran atribut ExampleSet ini konsisten dengan ExampleSet pada model yang dikirim ke port input model dilatih.

Operator ini juga memiliki port output, diantaranya, **labeled Data (lab)**, Model yang diberikan dalam input diterapkan pada ExampleSet yang diberikan dan ExampleSet terbaru disampaikan dari port ini. Beberapa informasi akan ditambahkan ke input ExampleSet sebelum dikirimkan melalui port output. Dan **model (mod)**, Model yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui port ini.



Gambar 4.21 Operator Apply Model

Seperti yang terlihat pada gambar 4.22, Operator Apply Model hanya memiliki dua parameter yaitu, application parameters (menu) parameter ini merupakan parameter ahli yang berguna memodelkan parameter untuk aplikasi (biasanya tidak diperlukan). Dan create view (boolean) Jika model diterapkan pada port input mendukung Views, Hal ini mungkin untuk bukannya mengubah data yang membuat View mendasarinya. Transformasi yang akan biasanva dilakukan langsung di data kemudian akan dihitung setiap kali nilai diminta dan hasilnya dikembalikan mengubah Beberapa tanpa data. model tidak mendukung Views.



Gambar 4.22 Parameter Apply Model

Dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit menggunakan Hipertensi, kita juga operator Performance. Operator ini digunakan untuk evaluasi kinerja. Operator ini memberikan daftar nilai kriteria kinerja. Kriteria kinerja secara otomatis ditentukan agar sesuai dengan jenis tugas belajar. Berbeda dengan operator lain, operator ini dapat digunakan untuk semua jenis tugas belajar. Secara otomatis menentukan jenis tugas belajar dan menghitung kriteria yang paling umum untuk jenis tersebut.

Untuk menemukan operator Performance, pilih Evaluation pada View Operator, lalu pilih Performance and Measurement, lalu pilih Performance. Setelah menemukan operator Performance, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.23 Hirarki Operator Performance

Operator Performance memiliki port input vaitu, labelled data (lab), Port ini mengharapkan mengharapkan ExampleSet berlabel. Apply Model merupakan contoh yang baik dari operator yang menyediakan berlabel. data Pastikan bahwa ExampleSet memiliki atribut label dan atribut prediksi. performance (per) Ini adalah parameter opsional yang membutuhkan Performance Vector.

Selain itu, Operator ini juga memiliki port output yaitu, **performance (per)**, port ini memberikan Peformance Vector (kita menyebutnya outputperformance-vektor untuk saat ini). Peformance Vector adalah daftar nilai kinerja kriteria. **example set (exa)**, ExampleSet yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui port ini.



Gambar 4.24 Operator Performance

Operator ini hanya memiliki satu parameter yaitu, **use example weights** (boolean) Parameter ini memungkinkan contoh bobot contoh yang akan digunakan untuk perhitungan kinerja jika memungkinkan. Parameter ini memiliki tidak memiliki efek jika atribut tidak memiliki peran bobot.



Gambar 4.25 Parameter Performance

Selanjutnya, susun dan hubungkan port-port dari operator decision tree, operator Apply Model dan operator Performance seperti yang terlihat pada Gambar 55.



Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Performance

Kemudian hubungkan operator retrieve (tabel SakitHipertensi) dengan operator validation dengan menarik garis pada port input dan output yang terdapat pada operator tersebut, seperti yang tampak pada Gambar 56.



Gambar 4.27 Susunan Operator Retrieve dengan Operator Validation

Setelah parameter dari masing-masing operator diatur, dan posisi operator disusun dengan benar, klik Run, lalu tunggu beberapa detik hingga RapidMiner akan menampilkan hasil Keputusan decision tree berupa graph pohon. seperti yang tampak pada Gambar 4.28.



Gambar 4.28 Tampilan Decision Tree

Chapter 5

Neural Network

Apa itu Neural Network?

Dapat dikatakan bahwa neural netwok dapat mempelajari pemetaan input data ke output data. Neural network merupakan model komputasi yang terinspirasi oleh prinsip-prinsip mengenai bagaimana cara otak manusia bekerja. Mereka dapat mempelajarinya dari data, mereka mampu mengeneralisasi dengan baik, dan mereka tahan dengan kebisingan.

Biasanya jaringan saraf digunakan untuk masalah-masalah seperti klasifikasi (classification), prediksi (prediction), pengenalan pola (pattern recognition), pendekatan (approximation), dan asosiasi (association). Mereka hanya perlu belajar dari beberapa data sampel, dan setelah mereka telah mempelajarinya, mereka dapat bekerja dengan input data yang tidak diketahui, atau bahkan input data yang bising maupun tidak lengkap.

Secara umum Neural Network (NN) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia. NN ini merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut.

Secara sederhana NN adalah sebuah alat pemodelan data statistik non-linear. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pada data. Secara mendasar. pola-pola sistem pembelajaran merupakan proses penambahan pengetahuan pada NN yang sifatnya kontinuitas sehingga pada saat digunakan pengetahuan tersebut maksimal akan dieksploitasikan dalam secara mengenali suatu objek. Neuron adalah bagian dasar dari pemrosesan suatu Neural Network. Dibawah ini merupakan bentuk dasar dari suatu neuron.

Bentuk Neural Network

Setiap neural network terdiri dari unit pengolahan dasar yang saling berhubungan, yang disebut Neuron. Network belajar dengan memodifikasi bobot hubungan antara neuron selama proses pelatihan. Bentuk dasar arsitektur suatu Neural Network adalah sebagai berikut:



Gambar 5.1 Arsitektur Dasar Neural Network

Secara umum, terdapat tiga jenis Neural Network yang sering digunakanberdasarkan jenis network-nya, yaitu:

- 1. Single-Layer Neural Network
- 2. Multilayer Perceptron Neural Network
- 3. Recurrent Neural Networks

Single-Layer Neural Network

Neural Network jenis ini memiliki koneksi pada inputnya secara langsung ke jaringan output.



Gambar 5.2 Single-layer Neural Network

Jenis Neural Network ini sangatlah terbatas, hanya digunakan pada kasus-kasus yang sederhana.

Multilayer Perceptron Neural Network

Jenis Neural Network ini memiliki layer yang dinamakan "hidden", ditengah layer input dan output. Hidden ini bersifat variable, dapat digunakan lebih dari satu hidden layer.



Gambar 5.3 Multilayer Perceptron Neural Network

Gambar di atas menunjukkan sebuah jaringan saraf sederhana yang dibuat dengan easyNeurons. Jenis jaringan ini disebut Multi Layer Perception dan itu merupakan salah satu jaringan yang paling umum digunakan.

Recurrent Neural Network

Neural network jenis ini memiliki ciri, yaitu adanya koneksi umpan balik dari output ke input.



Gambar 5.4 Recurrent Network

Kelemahan dari jenis ini adalah Time Delay akibat proses umpan balik dari output ke titik input.

Proses Pembelajaran pada Neural Network

Proses pembelajaran merupakan suatu metoda untuk proses pengenalan suatu objek yang sifatnya kontinuitas yang selalu direspon secara berbeda dari setiap proses pembelajaran tersebut. Tujuan dari pembelajaran ini sebenarnya untuk memperkecil tingkat suatu error dalam pengenalan suatu objek.

Secara mendasar, neural network memiliki sistem pembelajaran yang terdiri atas beberapa jenis berikut:

- 1. Supervised Learning
- 2. Unsupervised Learning

Supervised Learning

Sistem pembelajaran pada metoda Supervised learning adalah system pembelajaran yang mana, setiap pengetahuan yang akan diberikan kepada sistem, pada awalnya diberikan suatu acuan untuk memetakan suatu masukan menjadi suatu keluaran yang diinginkan. Proses pembelajaran ini akan terus dilakukan selama kondisi error atau kondisi yang diinginkan belum tercapai. Adapun setiap perolehan error akan dikalkulasikan untuk setiap pemrosesan hingga data atau nilai yang diinginkan telah tercapai.

Unsupervised Learning

Sistem pembelajaran pada neural network, yang mana sistem ini memberikan sepenuhnya pada hasil komputasi dari setiap pemrosesan, sehingga pada sistem ini tidak membutuhkan adanya acuan awal agar perolehan nilai dapat dicapai. Meskipun secara mendasar, proses ini tetap mengkalkulasikan setiap langkah pada setiap kesalahannya dengan mengkalkulasikan setiap nilai weight yang didapat.

Siapa yang menggunakan Neural Network?

Beberapa aplikasi yang khas adalah gambar (image), sidik jari dan pengenalan wajah (fingerprint face recognition), prediksi and saham (stock prediction), prediksi untuk taruhan (sport bets prediction), klasifikasi pola dan pengakuan (pattern classification and recognition), pengawasan dan (monitoring pengendalian and control). Mereka digunakan dalam industri, kedokteran (diagnosa), aplikasi militer (seperti radar pada pengenalan citra),

keuangan dan robotika. Akhir-akhir ini mereka sangat popular di industri game karena berkat mekanisme belajar yang dilakukan, mereka dapat memberikan kontrol adaptif dan pembelajaran untuk karakter yang dikendalikan computer.

Kegunaan Neural Networks

- 1. Pengenalan karakter optikal (Optical character recognition)
- 2. Pengenalan citra (Image recognition)
- 3. Pengenalan sidik jari (Fingerprint recognition)
- 4. Prediksi saham (Stock prediction)
- 5. Prediksi taruhan (Sport bets prediction)
- 6. Kontrol computer untuk karakter game (Computer controlled game characters)
- 7. Model statistical (Statistical modeling)
- 8. Data mining

Neural Network pada RapidMiner

Kita mulai dengan menggunakan data sederhana dalam tabel GE.xls. Data tersebut juga bisa kita dapatkan dengan melakukan pengunduhan melalui salah satu add-ins Microsoft Excel yang bernama DownloaderXL, dimana data mengenai harga saham yang terjadi dalam rentang waktu tertentu telah dicatat pada sebuah web hosting.

Contoh Kasus:

Perkiraan harga saham dengan menggunakan metoda Neural Network.

M	licrosoft Exc	el - GE						
	File Edit	View Inser	t Format	Tools Dat	a Window	Monte Carlo	xl bit Help Adobe PDF	
2111	200	alar	A AND AN			n nu 14		0004
: 14			£ ✓ ⊪2	6 40 4	🛛 • 🔺 🕛	1 + (* +)		.00% + (
A	rial	- 10	- B 1	<u>u</u> ∣≣⊧	콜 콜 🔜	\$ % ,	號 🔐 🏥 🗰 🖽 • 🔇	- <u>A</u> -
: 00	aluzerYI 🔻 🗍 Br	dk∩untee¥L ▼	Downloade		ntecYl • D	edictorYI + F	Racktesting 1 - Classifier VI -	PortfolioYI
: 01	O10	ancorossia a	EF 47544	00440404	iocoste i D	concorne 11	gachicostingne i <u>C</u> iassino ne i	Cordione
-	610	• p	< 55.17541	00140191	F	-	0	
-	A	8	6	0	E	F	6	н
1	Date	Open	High	Low	Close	Volume	StochasticOscillator(8,5)	
2	05/08/2009	14.18	14.35	13.95	14.33	129813025	ን 90.00	
3	05/11/2009	14.18	14.22	13.90	13.99	94827338	87.69	
4	05/12/2009	14.04	14.11	13.08	13.49	119778085	83.76	
5	05/13/2009	13.03	13.07	12.57	12.73	108138237	69.63	(
6	05/14/2009	12.59	13.00	12.33	12.86	91507642	57.62	
7	05/15/2009	12.89	13.07	12.52	12.68	84562005	39.26	
8	05/18/2009	12.93	13.31	12.90	13.28	77949206	32.04	
9	05/19/2009	13.45	13.77	13.30	13.51	93399522	32.21	
10	05/20/2009	13.84	14.18	13.55	13.58	115136364	42.82	
11	05/21/2009	13.32	13.34	12.86	13.05	106967634	45,49	
12	05/22/2009	13.18	13.22	12.84	12.92	54782537	48.77	
13	05/26/2009	12.84	13.42	12.67	13.20	97279770	48.79	
14	05/27/2009	13.28	13.31	12.68	12.81	108987013	40.92	
15	05/28/2009	13.05	13.10	12.64	13.01	83621254	32.51	
16	05/29/2009	13.14	13.34	12.05	13.29	86101902	37.25	
17	06/01/2009	13.63	13.80	13.32	13.67	87599560	47.17	
18	06/02/2009	13.51	13.66	13.36	13.61	67039102	55.18	
19	06/03/2009	13.46	13.51	13.11	13.31	69318595	65.10	

Gambar 5.5 Tabel GE.xls dalam Microsoft Excel

Buatlah *file* baru pada Microsoft Excel berdasarkan tabel harga saham. Berikan nama Header: Date, Open,

High, Low, Close, Volume, Stochastic Oscilator. Isilah sel seperti gambar [berapa]. Simpan dengan nama GE.xls

Lakukan pemilihan *repository* GE_TEMP_XLSDATA dengan melakukan *drag and drop* yang ditempatkan pada *panel main process* seperti gambar 5.6.



Gambar 5.6 Import Repository

Lakukan pemilihan operator *Neural Network* seperti gambar 5.7. Kemudian *drag and drop ke Main Process* seperti sebelumnya



Gambar 5.7 Operator Neural network

Lakukan pembuatan hubungan antara *repository* dan *operator*, kemudian antara *operator* dengan hasil *output*.



Gambar 5.8 Menghubungkan Seluruh Operator ke Result

klik ikon Play **D**. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 5.9 Ouput Neural Network

Gambar 5.9 merupakan grafik berbentuk *node* yang saling terhubung seperti layaknya sebuah jaringan syaraf dari hasil rules yang telah kita dapatkan

Chapter 6

Market Basket Analysis

Memahami Market Basket Analysis

Retail atau Eceran salah satu cara pemasaran produk meliputi semua aktivitas yang melibatkan penjualan barang secara langsung ke konsumen akhir, konsumen akhir membeli kumpulan produk dengan jumlah yang berbeda di waktu yang berbeda. Namun penjualan secara ritel hari ini bukanlah apa-apa jika insdustrinya tidak mampu berkompetisi dengan baik. Lanskap yang kompleks dan cepat berubah, persaingan yang ketat, dan pelanggan yang semakin menuntut mendorong *retailer* harus memikirkan kembali bagaimana mereka beroperasi. Kemampuan untuk memahami pola pikir konsumen adalah hal yang sangat penting bagi *retailer*.

Teknologi telah membantu *retailer* dengan memungkinkan untuk menyimpan data konsumen dengan volume yang sangat besar dan biaya yang sangat wajar. *Retailer* kini dapat memiliki miliyaran informasi tentang informasi pelanggan mereka. Informasi ini dapat menjawab pertanyaan-pertanyaan penting termasuk: Kapan pelanggan akan membeli? Bagaimana pembayaran dilakukan? Berapa banyak dan apa item tertentu yang dibeli? Apa hubungan antara barang yang dibeli?

Tidak ada keraguan bahwa data point-of-sales (POS) ini yang (ketika digunakan secara efektif) diberdayakan pengecer untuk lebih memahami bisnis mereka dan meningkatkan pengambilan keputusan. Pengecer proaktif menggunakan informasi ini untuk memberikan penawaran yang ditargetkan yang sesuai dengan harapan konsumen dan kemudian memberikan dampak penghasilan positif. Namun pada dasarnya, bagaimanakan *retailer* menggunakan miliyaran informasi ini? Jawabannya adalah menghubungkan produk-produk yang ada.

Sering kali, sebagai konsumen, kita cenderung mengabaikan bagaimana barang secara fisik diatur dalam sebuah toko *retail* atau supermarket. Apa yang mungkin terlihat (bagi kita) hanyalah seperti sebuah 'distribusi acak', namun sebenarnya hal tersebut merupakan pengaturan barang yang direncanakan secara cermat. Pada intinya, toko *retail* menilai pola pembelian pelanggan dan mengatur produk-produk yang akan dibeli secara sesuai. Sehingga menyebabkan pelanggan melakukan kegiatan pembelian beberapa produk sekaligus tanpa disadarinya.

Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan inilah yang dikenal sebagai *Market Basket Analysis* (MBA). Seperti namanya, *Market Basket Analysis* pada dasarnya melibatkan penggunaan data transaksional konsumen untuk mempelajari pola pembelian dan menjelajahi kemungkinan (probabilitas dan) *crossselling*. Tujuan dari MBA adalah untuk memanfaatkan data penjualan efektif untuk meningkatkan taktik pemasaran dan penjualan di tingkat toko.

Contoh yang paling umum dari Market Basket Analysis adalah "Beer dan Diapers". Contoh ini merupakan kasus dari salah satu toko retail besar yang ada di US, Wal-Mart. Seorang manajer toko menemukan hubungan yang kuat antara salah satu merek popok bayi (diapers) dan salah satu merek beer pembeli. pada beberapa Analisa pembelian mengungkapkan bahwa kegiatan pembelian dilakukan oleh laki-laki dewasa pada hari jumat malam terutama sekitar jam enam dan tujuh sore. Setelah beberapa observasi, supermarket mengetahui bahwa:

- Karena bungkus dari popok bayi sangat besar, para istri, dimana dalam banyak kasus adalah seorang ibu rumah tangga, akan menyuruh suaminya untuk membelinya.
- Pada akhir dari minggu, para suami dan ayah akan menghabiskan minggunya dengan membeli beberapa beer.

Jadi, apa yang akan dilakukan supermarket dari pengetahuan ini?

- Mereka menempatkan *premium beer* tepat disebelah *diapers*
- Hasilnya adalah para ayah akan membeli diapers dan yang biasanya membeli beer biasa sekarang

membeli *premium beer* seperti yang sudah diperkirakan.

 Secara signifikan, para pria yang biasanya tidak membeli bir sebelum mulai berbelanja akan membelinya karena itu begitu mudah dilihat dan diambil - hanya sebelah popok (cross-sell)

Istilah Market Basket Analysis sendiri datang dari kejadian yang sudah sangat umum terjadi di dalam swalayan, vakni ketika para konsumen pasar memasukkan semua barang yang merak beli ke dalam keranjang (basket) yang umumnya telah disediakan oleh pihak swalayan itu sendiri. Informasi mengenai produk-produk yang biasanya dibeli secara bersamasama oleh para konsumen dapat memberikan "wawasan" tersendiri bagi para pengelola toko atau swalayan untuk menaikkan laba bisnisnya (Albion Research, 2007).

Metodologi Association Rules

Metodologi Association Rules, atau Analisis Asosiasi adalah sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar. Salah satu penerapan Metode Association rules adalah pada Market Basket Analysis.
Association rule adalah sebuah ekspresi implikasi dari bentuk $X \rightarrow Y$, dimana X dan Y adalah itemset yang saling terpisah (disjoint), dengan kata lain $X \cap Y = \emptyset$. Dalam menentukan Association Rule, terdapat suatu interestingness measure (ukuran ketertarikan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Ada dua ukuran yaitu:

1. Support: Bagian transaksi yang mengandung kedua X dan Y.

 $Support(A) = rac{Jumlah transaksi mengandung A}{Total Transaksi}$

Atau jika terdapat dua buah item dalam X, nilai support diperoleh dari rumus berikut:

 $Support(A \cap B) = \frac{Jumlah transaksi mengandung A dan B}{Total Transaksi}$

2. Confidence: Seberapa sering item dalam Y muncul di transaksi yang mengandung X.

 $Cofidence = P(B|A) = \frac{Jumlah \ transaksi \ mengandung \ A \ dan \ B}{Jumlah \ transaksi \ mengandung \ A}$

Kudua ukuran ini nantinya berguna dalam menentukan interesting association rules, yaitu untuk dibandungkan dengan batasan (threshold) yang ditentukan oleh user. Batasan tersebut umumnya bernama *minimum support* dan *minimum confidence*.

Mengapa menggunakan Support dan Confidence? Support adalah ukuran yang penting karena jika aturan memiliki support yang kecil, maka kejadian bisa saja hanyalah sebuah kebetulan. Aturan Support yang rendah juga cenderung tidak menarik dari mungkin perspektif bisnis karena tidak akan memberikan keuntungan saat mempromosikan barangbarang yang jarang dibeli pelanggan bersamaan. Untuk alasan ini. dukungan sering digunakan untuk menghilangkan ketidak-menarikan ini. Confidence, adalah ukuran kehandalan dari kesimpulan yang dibuat oleh aturan. Semakin besar Confidence, semakin besar kemungkinan untuk Y hadir dalam transaksi yang mengandung X. Confidence juga memberikan probabilitas bersyarat dari Y yang diberikan ke X.

Contoh Association Rules

Untuk lebih memahami Association Rules, mari kita telusuri contoh berikut. Sebuah toko retail telah melakukan transaksi dengan pembeli seperti yang tertulis pada tabel.

Kode Transaksi	Produk yang terjual			
001	Pena, Roti, Mentega			
002	Roti, Mentega, Telur			
003	Buncis, Telur, Susu			

Tabel 6.1 Tabel Transaksi

004	Roti, Mentega
005	Roti, Mentega, Kecap, Telur, Susu

Tahap pertama adalah mencari nilai dari Support sesuai dengan rumus yang telah disebutkan sebelumnya. Misalnya, Untuk transaksi yang memuat {roti, mentega} ada 4, maka nilai supportnya adalah 80%. Lalu jumlah transaksi yang memuat {Roti, Mentega, Susu} ada 2, maka nilai supportnya adalah 40%. Sedangkan transaksi yang memuat {buncis} hanya 1, maka nilai supportnya adalah 20%. Jika kita tentukan bahwa *minimum support*nya adalah 30%, maka rule yang memenuhi adalah sebagai berikut:

Kombinasi Produk	Nilai Support
{roti}	80%
{mentega}	80%
{telur}	60%
{susu}	60%
{roti, mentega}	80%
{mentega, telur, susu}	40%
{roti, mentega, telur, susu}	40%

Tabel 6.2 Kombinasi Produk dan Nilai Support

Setelah semua pola kombinasi dan nilai dari Supportnya ditemukan, barulah dicari Association Rules yang memenuhi syarat minimum untuk confidence. Bila ditentukan syarat minimum untuk confidence sebesar 50% maka Association Rules yang dapat dipakai adalah:

Association Rules	Support	Confidence
{roti} → {mentega}	80%	100%
{roti} → {telur}	40%	50%
{roti} → {susu}	40%	50%
{roti} → {mentega, telur}	40%	50%
{roti} → {mentega, susu}	40%	50%
{roti} → {telur, susu}	40%	50%
{roti} → {mentega, telur, susu}	40%	50%
{mentega, telur} \rightarrow {roti}	40%	100%
{roti, mentega, susu} \rightarrow {telur}	40%	100%
{roti, telur, susu} → {mentega}	40%	100%
{mentega, telur, susu} \rightarrow {roti}	40%	100%

Tabel 6.3 Association Rules dan Nilai Confidence

Assosiation Rule akan dipilih sesuai kebijakan manajer toko, semakin tinggi support dan confidence semakin baik hasilnya. Misalkan kita ambil contoh yaitu {mentega, telur} \rightarrow {roti} yang memiliki nilai Support 80% dan Confidence 100%, artinya adalah: "Seorang konsumen yang membeli mentega dan telur memiliki kemungkinan 100% untuk juga membeli roti. Aturan ini

cukup signifikan karena mewakili 40% dari catatan selama ini."

Frequent Itemset Generation dan Rule Generation

Frequent Itemset Generation

Tujuannya adalah untuk menemukan semua itemset yang memenuhi *minimum support*. Item set ini sering disebut dengan frequent. Namun Masalah utama pencarian Frequent Itemset adalah banyaknya jumlah kombinasi itemset yang harus diperiksa apakah memenuhi minimum support atau tidak. Salah satu cara untuk mengatasinya adalah dengan mengurangi jumlah kandidat itemset yang harus diperiksa.

Apriori adalah salah satu pendekatan yang sering digunakan pada Frequent Itemset Mining. Prinsip Apriori adalah jika sebuah itemset infrequent, maka itemset yang infrequent tidak perlu lagi diexplore supersetnya sehingga jumlah kandidat yang harus diperiksa menjadi berkurang. Kira kira ilustrasinya seperti ini:



Gambar 6.1 Frequent Item Set tanpa Apriori

Pada gambar 36, pencarian Frequent Itemset dilakukan tanpa menggunakan prinsip Apriori. Dengan menggunakan prinsip Apriori, pencarian Frequent Itemset akan menjadi seperti di bawah ini:



Gambar 6.2 Frequent Item Set dengan Apriori

Dapat dilihat bahwa dengan menggunakan Apriori, jumlah kandidat yang harus diperiksa cukup banyak berkurang.

Rule Generation

Tujuannya adalah untuk mengekstrak semua aturan yang memiliki high-confidence dari itemsets yang ditemukan dari langkah sebelumnya. Aturan ini disebut Strong Rules.

Market Basket Analysis pada RapidMiner

Sekali lagi, pencarian Rule pada Association Rules merupakan sebuah proses yang luar biasa panjang. Manusia tidak akan mampu untuk melakukan pengitungan dengan berates-ratus (belum data kombinasi dari seluruh item yang ada). Maka dari itu, untuk mencari seluruh Rules yang ada, RapidMiner menyediakan tools telah untuk mempermudah pengguna. Untuk memahami cara menggunakan tools ini, ikuti manual berikut secara seksama.

Contoh Kasus :

Transaksi Penjualan Sederhana.

Kita mulai dengan menggunakan data sederhana yang kita miliki yang terdapat pada sub bab pengenalan Market Basket Analysis, Tabel 5.1.

	Α	В	С	D	E	F	G	н	
1	TID	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP	
2	001	1	1	1	0	0	0	0	
3	002	0	1	1	1	0	0	0	
4	003	0	0	0	1	1	1	0	
5	004	0	1	1	0	0	0	0	
6	005	0	1	1	1	0	1	1	
7									
									1

Gambar 6.3 Tabel Penjualan Sederhana

Buatlah Table baru pada Microsoft Excel berdasarkan tabel 5.1. Berikan nama Header: TID (Transaction ID), PENA, ROTI, MENTEGA, TELUR, BUNCIS, SUSU, KECAP. Isilah cell seperti gambar 5.3. Simpan dengan nama TransaksiMakanan.xls.

Lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada **Bab 2**. Browse table Microsoft Excel yang telah dibuat, dan masukan kedalam Local Repository, seperti gambar disamping.



Gambar 6.4 Repositori

Lakukan Drag dan Drop Tabel TransaksiMakanan tadi kedalam Process. Sehingga Operator Database muncul dalam Main Proses seperti gambar 5.5.



Gambar 6.5 Database dalam Main Process

Untuk melakukan Market Basket Analysis, kita membutuhkan setidaknya tiga buah operator, antara lain Association Rule, FP-Growth, dan Numerical to Binomial.



Gambar 6.6 Operator Create Association Rules

Assocuation rules dilakukan dengan menganalisis data pada frequent if/then patterns

menggunakan kriteria support dan confidence untuk mengidentifikasikan suatu relasi antar item. *Frequent if/then pattern* digali menggunakan operator FP-Growth. Operator Create Association Rules menggunakan frequent itemsets ini dan menghasilkan association rules.



Gambar 6.7 Operator FP-Growth

Frequent itemsets merupakan kelompok item yang sering muncul bersama-sama dalam data. Operator *FP-Growth* mengkalkulasikan semua frequent itemset dari input yang diberikan menggunakan struktur data FP-tree. Adalah wajib bahwa semua atribut dari masukan merupakan bilangan binominal (true/false).



Gambar 6.8 Operator Numerical to Binominal

Operator *Numerical to Binominal* diperlukan untuk mengubah nilai atribut yang berada pada table TransaksiMakanan menjadi binominal.

Selanjutnya lakukan Pencarian Filter untuk memudahkan kita menemukan operator yang dibutuhkan, lakukan seperti pada gambar berikut.



Gambar 6.9 Pencarian Operator Numerical to Binominal

Untuk Mencari Operator Numerical to Binominal, lakukan pencarian seperti gambar disamping. Operator ini terdapat pada hirarki: Data Transformation \rightarrow Type Conversion



Gambar 6.10 Pencarian Association Rules

Susunlah ketiga operator tersebut menjadi seperti gambar 5.11.



Gambar 6.11 Menghubungan Database TransaksiMakanan pada Operator Numerical to Binomial

Hubungkan Tabel TransaksiMakanan yang kita miliki dengan operator Numerical to Binominal. Proses ini akan membuat nilai dari Tabel Transaksi makan mejadi *Binominal Attributes*.

🛛 🛃 Parameters	🛛 🕘 Context 🔀						
🚨 💀 💀 🦻	🕵 🖶 -						
🍕 Nume	rical to Binominal						
attribute filter type	all 🔻						
invert selection							
include special attributes							
min	0.0						
max	0.0						

Gambar 6.12 Parameter Numerical to Binomial

Data yang kita miliki merupakan data sederhana. Kita hanya memperhitungkan 1 buah penjualan produk pada setiap transaksinya. Maka nilai yang terbaik untuk menjadi *false* adalah ketika tidak ada produk tertentu yang terjual dalam suatu transaksi, jadi kita sini nilai *min* dan *max* menjadi 0, Sehingga yang bernilai *false* adalah ketika sebuah produk tidak terdapat pada sebuah transaksi.

Hubungkan operator *Numerical to Binominal* dengan operator FP-Growth pada example output.



Gambar 6.13 Menghubungkan Operator Numerical to Binomial dengan Operator FP-Growth

Terdapat dua buah output untuk *Numerical to Binominal*, yaitu example dan original.

- Example, *numeric attributes* dikonversikan menjadi *binominal attributes* melalui output ini.
- Original, *numeric attributes* dilewatkan tanpa konversi. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

Lewatkan output pada *example*.

Isilah Parameter FP-Growth seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum support* senilai 30% atau 0.3.

🛛 🛃 Parameters 🕱	🔵 Context 🔀
🏅 🕫 🖻 🦻 🕵	- 🛼 🛨
🎘 FP-	Growth
🖌 find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.3
max items	-1
must contain	

Gambar 6.14 Parameter FP-Growth

Kemudian hubungkan operator *FP-Growth* dengan operator *Association Rules*.



Gambar 6.15 Menghubungkan Operator FP-Growth dengan Operator Create Association Rules

Terdapat dua buah output pada operator FP-Growth, yakni *example* dan *frequent*.

• *Example*, input yang diberikan dilewatkan tanpa adanya perubahan. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

• *Frequent,* frequent itemset dikirimkan melalui output ini.

Lewatkan output pada frequent.

Kemudian isilah Parameter Association Rules seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum confidence* senilai 50% atau 0.5.



Gambar 6.16 Parameter Association Rules

Setelah itu hubungkan Association Rules pada result. Sehingga seluruhnya membentuk seperti gambar 5.17. Ialu klik ikon Play . Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 6.17 Susunan Operator Association Rules

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 152 rules yang ditemukan.

No.	Premises	Conclusion	Support	Confid	LaPI	Gain	p-s	Lift	Convi
131	TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
132	ROTI, SUSU	MENTEGA, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
133	MENTEGA, SUSU	ROTI, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
134	ROTI, MENTEGA, SUSU	TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
135	ROTI, TELUR, SUSU	MENTEGA, KECAP	0.200	1	1	-0.200	0.160	5	00
136	MENTEGA, TELUR, SUSU	ROTI, KECAP	0.200	1	1	-0.200	0.160	5	00
137	ROTI, MENTEGA, TELUR, SUSU	KECAP	0.200	1	1	-0.200	0.160	5	00
138	KECAP	ROTI, MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
139	ROTI, KECAP	MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	∞
140	MENTEGA, KECAP	ROTI, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
141	ROTI, MENTEGA, KECAP	TELUR, SUSU	0.200	1	1	-0.200	0.120	2.500	∞
142	TELUR, KECAP	ROTI, MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
143	ROTI, TELUR, KECAP	MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
144	MENTEGA, TELUR, KECAP	ROTI, SUSU	0.200	1	1	-0.200	0.160	5	00
145	ROTI, MENTEGA, TELUR, KECAP	SUSU	0.200	1	1	-0.200	0.120	2.500	00
146	SUSU, KECAP	ROTI, MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
147	ROTI, SUSU, KECAP	MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
148	MENTEGA, SUSU, KECAP	ROTI, TELUR	0.200	1	1	-0.200	0.120	2.500	00
149	ROTI, MENTEGA, SUSU, KECAP	TELUR	0.200	1	1	-0.200	0.080	1.667	00
150	TELUR, SUSU, KECAP	ROTI, MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
151	ROTI, TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
152	MENTEGA, TELUR, SUSU, KECAP	ROTI	0.200	1	1	-0.200	0.040	1.250	00

Gambar 6.18 Hasil Association Rules Pertama

Tentunya ini akan menyulitkan kita untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Maka dari itu yang harus kita lakukan adalah mengubah nilai *minimum support* dan *minimum confidence*.

Klik ikon Edit 🔟 untuk kembali pada *model view.* Lalu klik Operator FP-Growth.



Gambar 6.19 Operator FP-Growth

Kemudian lihat bagian parameter. Ubah nilai minimum support menjadi 95%, seperti yang sudah dijelaskan pada sub bab Association Rules, semakin tinggi nilai support maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum supportnya terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

🦳 🛃 Parameters 🕱	🔵 Context 🔀
🚨 🗟 🖻 🦻 🕵	: 📭 👻
🎘 FP-	Growth
🖌 find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.95
max items	-1
must contain	

Gambar 6.20 Mengubah Parameter FP-Growth

Sekarang kita beralih pada Operator Create Association Rules.



Gambar 6.21 Operator Create Association Rules

Ubah nilai minimum confidence menjadi 95% atau 0.95, semakin tinggi nilai confidence maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum confidence terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

criterion	confidence 🔻
min confidence	0.95
gain theta	2.0
laplace k	1.0

Gambar 6.22 Mengubah Parameter Association Rules

Klik ikon Play 🖻 untuk menampilkan hasil yang baru.

No	Premises	Conclusion	Support	Confid	LaPla	Gain	D-S	Lift	Convi
1	ROTI	MENTEGA	0.800	1	1	-0.800	0.160	1.250	00
2	MENTEGA	ROTI	0.800	1	1	-0.800	0.160	1.250	00
3	SUSU	TELUR	0.400	1	1	-0.400	0.160	1.667	00
4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400	0.080	1.250	00
5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400	0.080	1.250	00

Gambar 6.23 Hasil Association Rules Kedua

Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

Untuk melihat dalam bentuk grafik. kita dapat memilih opsi Graph View. O Table View O Graph View O Annotations



Gambar 6.24 Hasil dalam bentuk Graph View

Glossarium

- *Algoritma* Kumpulan perintah untuk menyelesaikan suatu masalah.
- Apriori Algoritma untuk frequent itemset mining dan association rule dalam database transaksional. DIhasilkan dengan menidentifikasi setiap buah item, dan memperluasnya menjadi kombinasi kumpulan item yang lebih besar asalkan himpunan item muncul cukup sering dalam database.
- Association Rules Sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar.
- **Binominal Attributes** Atribut dengan tipe Binominal (true dan false).
- Confidence (Market Basket Analysis) Seberapa sering item dalam Y muncul di transaksi yang mengandung X.
- Decision tree Struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan

simpul daun merepresentasikan kelas atau distribusi kelas.

- Disjoint Himpunan terpisah, tidak ada elemen yang berhubungan diantara kedua himpunan yang bersangkutan
- *Flowchart* Sebuah diagram dengan simbol-simbol grafis yang menyatakan aliran algoritma.
- Frequent Itemset Itemset yang memunyai support >= minimum support yang diberikan oleh user dalam Market Basket Analysis.
- Market Basket Analysis Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan.
- MBA Lihat Market Basket Analysis.
- Minimum SupportNilaiSupportTerkecildalamMarketBasket Analysis yang dapat di toleransi.
- Minimum Confidence Nilai Confidence terkecil dalam Market Basket Analysis yang dapat di toleransi.
- Neural Network Jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia.
- **Numeric Attributes** Atribut dengan tipe Numerical (1-9).
- *Operator* suatu tanda atau simbol yang dipakai untuk menyatakan suatu operasi atau manipulasi nilai.
- ParameterNilai yang mengikuti acuan keterangan
atau informasi yang dapat menjelaskan

batas-batas tertentu dari suatu suatu sistem persamaan.

Pruning Teknik dalam machine learning yang mengurangi ukuran pohon keputusan dengan menghapus bagian dari pohon yang memberikan sedikit kekuatan untuk mengklasifikasikan kasus.

- RapidMinerSebuah tool yang digunakan untuk
melakukan analisis terhadap data mining,
text mining dan analisis prediksi.
- Repositori Kumpulan paket yang siap untuk diambil dan digunakan sesuai dengan kebutuhan pengguna.
- Simpul akar Simpul tanpa ayah yang berada pada tingkat tertinggi.
- Simpul daun Semua simpul yang berada pada tingkat terendah.
- *Simpul internal* Semua simpul dari pohon yang memiliki anak tetapi bukan daun.
- Support(Market Basket Analysis) Bagian transaksi
yang mengandung kedua X dan Y.
- Teori graf Cabang kajian yang mempelajari sifat-sifat graf.
- Validasi Tindakan yang membuktikan bahwa suatu proses/metode dapat memberikan hasil yang konsisten sesuai dengan spesifikasi yang telah ditetapkan.

Daftar Pustaka

Akhtar, Fareed dan Caroline Hahne. 2012. *RapidMiner 5 Operator Reference*, [online], (www.rapid-i.com, diakses tanggal 30 Januari 2013).

Amiruddin, dkk. Penerapan Association Rule Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan Untuk Menemukan Pola Sertifikasi Guru. Institut Teknologi Surabaya. Surabaya.

Basuki, Achmad dan Iwan Syarif. *Decision Tree*, [online], (http://lecturer.eepis-its.edu/~entin/Data%20Mining/ Minggu%205%20Decision%20Tree.pdf, diakses tanggal 05 Februari 2013).

Khusnawi. 2007. *Pengantar Solusi Data Mining*. Yogyakarta.

Kusumadewi, Sri. 2003. Artificial Intelligence: Teknik dan Aplikasinya.

Mitchel, Tom M. 1997. *Machine Learning*. New York: McGraw-Hill.

Prasetyo, Bowo. 2011. *Mengenal RapidMiner*, [online], (www.slideshare.net/bowoprasetyo/RapidMiner, diakses tanggal 31 Januari 2013).

Prasetyo, Kokoh Philips. 2006. *APriori*, [online] (http://philips.wordpress.com/2006/06/07/apriori, diakses tanggal 03 Februari 2013)

-----. 2006. Association Rule Mining, [online]. (http://philips.wordpress.com/2006/05/10/associationrule-mining, diakses tanggal 03 Februari 2013).

Rafaida, Ropi. *Decision Tree (Pohon Keputusan)*, [online],

(http://file.upi.edu/Direktori/FPEB/PRODI._MANAJEME N_FPEB/197302052005012-

ROFI_ROFAIDA/MATERI_KULIAH/DECISION_TREE.pdf, diakses tanggal 05 februari 2013).

Ross, Peter. 2000. Data Mining [online]. (http:// www.soc.napier.ac.uk/~peter/vldb/dm/dm.html, diakses tanggal 07 Februari 2013)

Wahono, Romi satria. *Data Mining:Proses Data Mining*, [online], (http://romisatriawahono.net/lecture/dm/ romi-dm-02-proses-june2012.pptx, diakses tanggal 31 Januari 2013).

2012. *RapidMiner 5.0 Manual English*, (online), (www.rapid-i.com, diakses tanggal 30 Januari 2013).

3 tips for Setting up Association Rules using RapidMiner, [online]. (http://www.simafore.com/blog /bid/110113/3-tips-for-setting-up-a-Market-Basket-Analysis-using-RapidMiner, diakses tanggal 08 Maret 2013).

Association Analysis: Basic Concepts and Algorithms, [online]. (http://www-users.cs.umn.edu/ ~kumar/dmbook/ch6.pdf, diakses tanggal 08 April 2013)

Decision Tree (Pohon Keputusan), [online], (http://www.google.co.id/url?sa=f&rct=j&url=http://no vrina.staff.gunadarma.ac.id/Downloads/files/21783/Alg oritma%2BC4.pdf&q=algoritma+c4&ei=6h9gUcbJFIqrrA

fT7IGQAw&usg=AFQjCNG7HbyNPOqa63ZoPexX76TrIIJ7g, diakses tanggal 05 februari 2013).

Landasan Teori Market Basket Analysis, [online]. (http://library.binus.ac.id/eColls/eThesis/Bab2/2010-1-00498-MTIF%20Bab%202.pdf, diakses tanggal 08 April 2013)

Understanding the Concept of Market Basket Analysis, [online]. (http://www.thesmartcube.com/ insights/blog/brand-strategy/understanding-theconcept-of-market-basket-analysis, diakses tanggal 08 Maret 2013)

RapidMiner Resources. (http://RapidMinerresources .com/uploads/videos/tomott/RapidMiner5-Vid1.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%201.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%202.flv, diakses tanggal 02 Februari 2013)



ID WMO Lintang Bujur Elevasi

: 96087 Nama Stasiun : Stasiun Meteorologi Hang Nadim : 1.11667 : 104.11667 : 26

Tanggal	Tn	Тх	Tavg	RH_avg	SS
01-02-2020	26,4	31,8	28,6	76	10,4
02-02-2020	25	31,7	27,2	82	10,1
03-02-2020	25	32,2	28,1	73	3,5
04-02-2020	25,1	31,2	27,7	72	9,3
05-02-2020	25,1	32,1	28	76	8,5
06-02-2020	25,6	31,6	28	77	9
07-02-2020	25,7	32,1	28,3	72	9,1
08-02-2020	25,6	31,6	27,7	71	7,4
09-02-2020	25	29,8	26,4	85	2,5
10-02-2020	23,6	31,3	26,9	84	0
11-02-2020	24	31,4	27,1	88	7,5
12-02-2020	26	31,6	28,1	84	4,9
13-02-2020	25	31,3	27,6	84	6,3
14-02-2020	25	31,4	28	79	9,5
15-02-2020	25	34,5	28	78	9,9
16-02-2020	25,9	31,5	28	81	10,8
17-02-2020	25,9	32,2	28,2	76	4,4
18-02-2020	24,2	31,4	26,8	85	10,2
19-02-2020	24,2	32	28,7	74	6,6
20-02-2020	24,2	31,8	28,1	76	2,2
21-02-2020	24,2	31,7	26,7	87	9,9
22-02-2020	24,9	32,1	28,7	74	9
23-02-2020	26,1	32,4	28,7	73	9,8
24-02-2020	24,2	30,2	27,2	82	9,8
25-02-2020	24,8	31,7	28,2	77	2,2
26-02-2020	25,7	32,4	28,5	76	10,6
27-02-2020	25,2	31,8	28,4	76	11
28-02-2020	26,4	31,8	28,4	78	10,6
29-02-2020	25,8	31,4	28,1	73	8,8

Keterangan :

Tn: Temperatur minimum (°C)

Tx: Temperatur maksimum (°C)

Tavg: Temperatur rata-rata (°C)

RH_avg: Kelembapan rata-rata (%)

ss: Lamanya penyinaran matahari (jam)

ff_x: Kecepatan angin maksimum (m/s)

ff_avg: Kecepatan angin rata-rata (m/s)

ff_x	ff_avg	cuaca
7	6	tidak hujan
9	5	hujan
8	4	tidak hujan
7	4	tidak hujan
8	4	tidak hujan
8	5	tidak hujan
7	4	tidak hujan
4	3	tidak hujan
6	2	hujan
4	3	hujan
6	4	tidak hujan
6	4	hujan
9	4	hujan
8	4	tidak hujan
5	3	tidak hujan
5	3	tidak hujan
7	3	tidak hujan
8	4	tidak hujan
9	5	tidak hujan
8	4	hujan
9	5	hujan
10	5	hujan
9	5	tidak hujan
8	5	hujan
8	5	hujan
7	4	tidak hujan
8	4	tidak hujan
9	5	tidak hujan
9	4	tidak hujan

Tanggal	Tn Tx	Tavg	RH_av	g ss	ff_x	ff_avg	cuaca
01-02-2020	26,4	31,8	28,6	76	10,4	7	6 tidak hujan
02-02-2020	25	31,7	27,2	82	10,1	9	5 hujan
03-02-2020	25	32,2	28,1	73	3,5	8	4 tidak hujan
04-02-2020	25,1	31,2	27,7	72	9,3	7	4 tidak hujan
05-02-2020	25,1	32,1	28	76	8,5	8	4 tidak hujan
06-02-2020	25,6	31,6	28	77	9	8	5 tidak hujan
07-02-2020	25,7	32,1	28,3	72	9,1	7	4 tidak hujan
08-02-2020	25,6	31,6	27,7	71	7,4	4	3 tidak hujan
09-02-2020	25	29,8	26,4	85	2,5	6	2 hujan
10-02-2020	23,6	31,3	26,9	84	0	4	3 hujan
11-02-2020	24	31,4	27,1	88	7,5	6	4 tidak hujan
12-02-2020	26	31,6	28,1	84	4,9	6	4 hujan
13-02-2020	25	31,3	27,6	84	6,3	9	4 hujan
14-02-2020	25	31,4	28	79	9,5	8	4 tidak hujan
15-02-2020	25	34,5	28	78	9,9	5	3 tidak hujan
16-02-2020	25,9	31,5	28	81	10,8	5	3 tidak hujan
17-02-2020	25,9	32,2	28,2	76	4,4	7	3 tidak hujan
18-02-2020	24,2	31,4	26,8	85	10,2	8	4 tidak hujan
19-02-2020	24,2	32	28,7	74	6,6	9	5 tidak hujan
20-02-2020	24,2	31,8	28,1	76	2,2	8	4 hujan
21-02-2020	24,2	31,7	26,7	87	9,9	9	5 hujan
22-02-2020	24,9	32,1	28,7	74	9	10	5 hujan
23-02-2020	26,1	32,4	28,7	73	9,8	9	5 tidak hujan
24-02-2020	24,2	30,2	27,2	82	9,8	8	5 hujan
25-02-2020	24,8	31,7	28,2	77	2,2	8	5 hujan
26-02-2020	25,7	32,4	28,5	76	10,6	7	4 tidak hujan
27-02-2020	25,2	31,8	28,4	76	11	8	4 tidak hujan
28-02-2020	26,4	31,8	28,4	78	10,6	9	5 tidak hujan
29-02-2020	25,8	31,4	28,1	73	8,8	9	4 tidak hujan

Tutorial Menentukan Kondisi Cuaca Menggunakan Rapidminer Dengan Metode Decision Tree

Kita akan menggunakan datasheet berupa file excel sebagai berikut : - terlampir di attachment

Software yang akan kita gunakan yaitu Rapidminer version 9.8

Datasheet ini berasal dari data iklim harian di Stasiun Meteorologi Hang Nadim Batam dari tanggal 1 – 02 – 2020 s/d 29 – 02 – 2020 . Kita akan menentukan kondisi cuaca hujan atau tidak hujan dengan menggunakan beberapa atribut sebagai berikut :

- Tn : Temperatur minimum (°C)
- Tx : Temperatur maksimum (°C)
- Tavg : Temperatur rata-rata (°C)
- RH_avg : Kelembapan rata-rata (%)
- Ss : Lamanya penyinaran matahari (jam)
- ff_x : Kecepatan angin maksimum (m/s)
- ff_avg : Kecepatan angin rata-rata (m/s)

Kemudian buka aplikasi rapidminer, pilih blank document, selanjutnya pada bagian operator, pilih read excel, kemudian klik 2x, maka akan muncul di bagian Process.

🍥 <new process*=""> – RapidMiner Studio Ec</new>	ducational 9.8.000 @ DESKTOP	P-U4V3K7U					
Eile Edit Process View Connection	ns <u>S</u> ettings E <u>x</u> tensions	<u>H</u> elp					
	•	Views:	Design	Results	Turbo Prep	Auto Model	Deployments
Repository ×	Process						
🛟 Import Data 🛛 🖃 🔻	Process				,o p		🛛 🤹 🗲
Training Resources (connected)	Process						
Samples							
Community Samples (connected)) inp						res (
Local Repository (Local)	Read Excel						
DB (Legacy) Operators ×	fil out	•					
Files (3)							
Read (2)							
Read CSV							
Read Excel							
Write (1)							
Write Excel							
We found "Advanced Reporting Extension" and "Spreadsheet Table Extraction" in the Marketplace. <u>Show mel</u>	Leverage the Wisdom	of Crowds to get o	operator recomi	mendations based vate Wisdom of Cr	d on your process de rowds	isign!	

Gambar 1.

Selanjutnya klik read excel di bagian process, pilih import configuration Wizard di bagian parameter,

Parameters	×	
📩 Read Excel		
*	Import Configuration Wizard]@
excel file]0
sheet number	1	D
🖌 first row as nam	es	٩
date format	Enter value	0
Show advanced	parameters ibility (9.8.000)	

Gambar 2.

Setelah di klik, maka akan muncul tampilan seperti Gambar 3. Kemudian kita pilih file datasheet excel yang ada di dalam folder komputer, lalu klik next.

	Select the data loca	tion.			
Downloads			▼ ← 5	井 金 📑	q
Bookmarks	File Name	Size	Туре	Last Modified	
🕇 Last Directory	Compressed		File Folder	Oct 10, 2020	
	- Documents		File Folder	Oct 29, 2020	
	FastStone Capture 9.4 Corporate.kuyhAa		File Folder	Aug 24, 2020	
	Hard Disk Sentinel Pro 5.61 Build 11463.kuyh	Aa	File Folder	Mar 13, 2020	
	IDM.6.37.b.14.kuyhAa.Me		File Folder	May 3, 2020	
	KMSAuto Lite 1.5.6 Portable.kuyhAa.Me		File Folder	Oct 4, 2020	
	Music		File Folder	Jun 7, 2020	
	Programs		File Folder	Oct 29, 2020	
	Video		File Folder	Jun 10, 2020	
	Algorima C4.5 penjurusan.xlsx	87 KB	Microsoft Excel Wor	Oct 29, 2020	
	feb2020.xisx	16 KB	Microsoft Excel Wor	Oct 29, 2020	
	kohkiloyeh (1).xlsx	11 KB	Microsoft Excel Wor	Oct 28, 2020	
	kohkiloyeh.xlsx	11 KB	Microsoft Excel Wor	Oct 28, 2020	
	aporan.xlsx	19 KB	Microsoft Excel Wor	Oct 29, 2020	
	laporan_iklim_harian (1).xlsx	13 KB	Microsoft Excel Wor	Oct 29, 2020	
	lanoran iklim harian /?) vlev	16 KR	Microsoft Excel Wor	0420 2020	
eb2020.xlsx					
eb2020.xisx	별을 kohkiloyeh.xisx 또한 laporan.xisx 또한 laporan_iklim_harian (1).xisx 또한 laporan_iklim_harian /2\.visy	11 KB 19 KB 13 KB 16 KB	Microsoft Excel Wor Microsoft Excel Wor Microsoft Excel Wor Microsoft Excel Wor	Oct 28, 2020 Oct 29, 2020 Oct 29, 2020 Oct 29, 2020	

Gambar 3.

Selanjutnya kita pilih kolom mana yang akan digunakan sebagai atribut, kemudian klik next.

				Select th	ne cells to	import.			
She	et: Data Haria	an - Table 🔻	Cell rar	nge: A9:169	Select All	🗸 Defin	e hea 1	A V	
	A	в	с	D	E	F	G	Н	T
8	18-02-2020	24.200	31.400	26.800	85.000	10.200	8.000	4.000	tidak hujan
9	19-02-2020	24.200	32.000	28.700	74.000	6.600	9.000	5.000	tidak hujan
0	20-02-2020	24.100	31.800	28.100	76.000	5.000	8.000	4.000	hujan
1	21-02-2020	24.200	31.700	26.700	87.000	9.900	9.000	5.000	hujan
2	22-02-2020	24.900	32.100	28.700	74.000	9.000	10.000	5.000	hujan
3	23-02-2020	26.100	32.400	28.700	73.000	9.800	9.000	5.000	tidak hujan
4	24-02-2020	24.200	30.200	27.200	82.000	9.800	8.000	5.000	hujan
5	25-02-2020	24.800	31.700	28.200	77.000	2.200	8.000	5.000	hujan
6	26-02-2020	25.700	32.400	28.500	76.000	10.600	7.000	4.000	tidak hujan
7	27-02-2020	25.200	31.800	28.400	76.000	11.000	8.000	4.000	tidak hujan
8	28-02-2020	26.400	31.400	28.400	78.000	10.600	9.000	5.000	tidak hujan
	29-02-2020	25.800	31.400	28.100	73.000	8.800	9.000	4.000	tidak hujan
9									



Selanjutnya kita akan menentukan format di setiap kolomnya, untuk kolom tanggal, karena tidak termasuk atribut maka kita ganti dengan cara pilih change role, lalu pilih id dan untuk kolom cuaca kita ubah menjadi label.

_					Forma	t you	r column	IS.				_
	Replace er	rors	with missir	gvalues (D							
Ta pol	nggal -	¢ •	Tn real	¢ •	Tx real	¢ •	Tavg real	¢ •	RH_avg integer	¢ •	ss real	¢ v
01	-02-2020	6	Change	ole							× .400	
02	-02-2020										.100	
03	-02-2020		_	Please	nter the new	role:					500	
04	-02-2020										300	
05	-02-2020	16	d								▼ 500	
06	-02-2020	la	abel								000	
07	-02-2020	ic	j 								100	
08	-02-2020		reigni					•		-	100	
09	-02-2020	_	25.000		29.800		26.400		85		2.500	
0 10	-02-2020		23.600		31.300		26.900		84		0.000	
1 11	-02-2020		24.000		31.400		27.100		88		7.500	
			26.000		31.600		28 100		84		4 900	

Gambar 5.

Setelah kita ubah, kolom tanggal akan berubah warna menjadi biru muda, dan kolom cuaca berubah warna menjadi hijau muda. Apabila tidak ada error selanjutnya klik finish.

	Penlace	errore w	ith missing v	alues (i)	Format your	columns.		
,	Tavg real	¢ v	RH_avg	¢ v	ss ¢ v real	ff_x & v	ff_avg	cuaca & v
1	28.200		76		4.900	9	4	tidak hujan
2	27.700		83		4.900	7	4	tidak hujan
3	28.300		78		5.800	8	4	tidak hujan
4	27.900		79		5.700	4	3	tidak hujan
5	28.000		78		4.200	7	4	hujan
6	26.800		84		10.500	5	4	tidak hujan
7	26.200		87		6.200	5	4	hujan
8	28.100		79		2.700	8	5	hujan
9	27.900		80		7.400	7	4	hujan
0	27.700		82		6.900	4	3	hujan
1	27.400		82		1.800	4	3	tidak hujan
2	27 800		78		3.000	6	4	tidak huian

Gambar 6.

Selanjutnya pada bagian operator, kita pilih Decision Tree, maka akan muncul di bagian process

🧶 <new process*=""> – RapidMiner Studio Educ</new>	ational 9.8.000 @ DESKTOP-U	J4V3K7U						
Elle Edit Process View Connections	Settings Extensions E	tiews:	Design	Results	Turbo Prep	Auto Model	Deployme	ents
Repository ×	Process							
🔂 Import Data 🛛 = 💌	Process				jo jo		a 🧧	Ð
Training Resources (connected) Samples	Process							
Community Samples (connected)	Dinp							res
Local Repository (Local)	Read Excel		Decision T	ree				
DB (Legacy) Operators X	fi 📩 out		tra	mod) exa) wei				
decision X								
Vodeling (8)								
▼ 📇 Predictive (8)								
🕶 📇 Trees (8)								
Decision Tree								
💡 Random Forest								
💡 Gradient Boosted Trees 🗸								
<	Leverage the Wisdom of	Crowds to get o	perator recom	mendations base	d on your process d	esign!		
We round "Whibo" and "Weka Extension" in the Marketplace. <u>Show me!</u>			🗸 Act	ivate Wisdom of C	rowds			

Gambar 7.

Selanjutnya, hubungkan out di read excel ke Decision Tree, kemudian di bagian Model dan Example Decision Tree, hubungkan ke Result

<new process*=""> - RapidMiner Studio Educ File Edit. Descent View Operations</new>	ational 9.8.000 @ DESKTOP-U	U4V3K7U				
		Views: Design	Results	Turbo Prep	Auto Model	Deployments
Repository ×	Process					
🕂 Import Data 🛛 = 💌	Process			⊊ر ≎ر	la 🖡 🔒	🗵 🦷 🍒
	Process					res
Local Repository (Local)	Read Excel	Decision T	ree			res
DB (Legacy) Operators X nav X	fi ot	tra 🧯	mod exa wei		1	res
Data Access (5) Data Access (5) Files (2) Write (2) Write CSV Write Excel Database (1) VALUE Access						
< > >	Leverage the Wisdom of	Crowds to get operator recom	mendations based	i on your process d	lesign!	
We found "Tacl" in the Marketplace. <u>Show mel</u>		🗸 Act	ivate Wisdom of Ci	rowds		

Gambar 8.

Kemudian kita klik tombol run , maka akan muncul hasil berupa tampilan akar keputusan dan Example set seperti gambar 9. Untuk melihat deskripsinya klik tanda description.

<new process*=""></new>	– RapidMiner Studio Educational 9.8.000 @ DE	ESKTOP-U4V3K7U					
		Views:	Design Resu	Its Turbo Prep	Auto Model	Deployments	Fin
Result History	💡 Tree (Decision Tree)	× 📒 ExampleS	Set (Read Excel)				
Graph	Zoom P P Tree V Node Labels	> 81	ss >3 RH_avg	≤3 hujan			
Description	✓ Edge Labels	Tavg > 27.150 ≤ 27.1	tidak h	ujan			
Annotations		nujan tida	ik nujan				


In the second secon

		Views: Desig	n Results	Turbo Prep	Auto Model	Deployments	Fil
Result History	💡 Tree (Decision Tree) 🛛 🗙						
Graph	Tree ss > 3 RH avg > 81.500						
Description	Tavg > 27.150: hujan {tida Tavg \leq 27.150: tidak hujan RH_avg \leq 81.500: tidak hujan { ss \leq 3: hujan {tidak hujan=0, huja	k hujan=0, hujan= {tidak hujan=2, tidak hujan=17, h n=4}	4} hujan=1} ujan=1}				
Annotations							

Gambar 10.

Eile Edit Proce	ss <u>V</u> iew <u>C</u> o	nnections <u>S</u>	ettings Exte	nsions <u>H</u> elp)							
	-	•			Views:	Design	Results	Turb	o Prep	Auto Model	Deployments	Fi
Result History	9	Tree (Decis	ion Tree)	×	ExampleSe	t (Read Ex	cel) ×					
	Open in	📑 Turbo Pre	p 🖁	Auto Model						Filter (29 / 29 e	xamples): all	•
Data	Row No.	Tanggal	cuaca	Tn	Тх	Tavg	RH_avg	SS	ff_x	ff_avg		
	1	01-02-20	tidak huj	26.400	31.800	28.600	76	10.400	7	6		^
Σ	2	02-02-20	hujan	25	31.700	27.200	82	10.100	9	5		
Statistics	3	03-02-20	tidak huj	25	32.200	28.100	73	3.500	8	4		
	4	04-02-20	tidak huj	25.100	31.200	27.700	72	9.300	7	4		
I	5	05-02-20	tidak huj	25.100	32.100	28	76	8.500	8	4		
Visualizations	6	06-02-20	tidak huj	25.600	31.600	28	77	9	8	5		
	7	07-02-20	tidak huj	25.700	32.100	28.300	72	9.100	7	4		
	8	08-02-20	tidak huj	25.600	31.600	27.700	71	7.400	4	3		
	9	09-02-20	hujan	25	29.800	26.400	85	2.500	6	2		
Annotations	10	10-02-20	hujan	23.600	31.300	26.900	84	0	4	3		
	11	11-02-20	tidak huj	24	31.400	27.100	88	7.500	6	4		
	12	12-02-20	hujan	26	31.600	28.100	84	4.900	6	4		
	13	13-02-20	hujan	25	31.300	27.600	84	6.300	9	4		
	14	14-02-20	tidak huj	25	31.400	28	79	9.500	8	4		
	15	15-02-20	tidak huj	25	34.500	28	78	9.900	5	3		~
	ExampleSet (29 examples,	2 special attri	butes, 7 regu	lar attributes)							

Interpretation of the state of the state

Gambar 11.

Sumber :

http://perpustakaan.stmikbudidarma.ac.id/downloadebookData%20Mining%20dg%20Rapid%20Miner_tagt.pdf

Nama	: Trada Ayang Pratiwi
NIM	: 2015210046
Tugas	: 06 Advanced Database
Sumber	: Youtube Tutorial Mengolah Data Mining Metode Klasifikasi Algoritma ID3
	Menggunakan RapidMiner link
	https://www.youtube.com/watch?v=M65yJI5ui9M

- 1. Donload aplikasi Rapid Miner di google dan install aplikasi
- 2. Buatlah datasheet di excel
- 3. Import datasheet ke Rapid Miner
 - a. Cari di Operators Read Excel
 - b. Double klik Read Excel atau drag ke area Process
 - c. Double klik Read Excel di area Process dan cari datasheet yang sudah kita buat



4. Cari Multiply di Operators lalu double klik

	View	vs: Design Results	Turbo Prep Auto Model	Deployments	Find data, operatorsetc 🔎 All Studio 🔻
futorials ×	Repository ×	Process			Parameters ×
Operators and Processes 🔒 View All	🕒 Import Data 🛛 = 💌	Process		🗲 🖬 🛋 🔍 🔍	I Multiply
1/5 Welcome to RapidMiner. RapidMiner Studio has a lot of features to offer. In the next few minutes, you will learn how you can use RapidMiner Studio to perform basic data science techniques, such as: accessing data, transforming data, and building statistical models. We will use data describing the Titanic	Samples Gamples DB (Legacy)	Process Read Excel) re T a or (Welcome to Ra We have a super brief RapidMiner on the lef	pidMiner! f introduction to usin f to f your screen. Che	B Ck it ou
EXPLANATION Each tutorial is broken into several steps to teach you the basics of RapidMiner Studio. Each step lists specific actions to perform, plus an explanation of why those actions	Multipl X Utility (1) Multiply This operat	ors multiplies its input object.			Show advanced parameters Help × Multiply RapidMiner Studio Core
matter. Notice the dotted-line preview in the Process Panel; this	Press "F3" f	or focus.			Tags: <u>Copy</u> , <u>Branch</u> , <u>Duplicate</u> , <u>Utility</u> Synopsis
shows you the process you will be		Leverage the Wisdom of Crowds	to get operator recommendations base	ed on your process design!	This Operator creates copies of a
building in that tutorial.				-	RapidMiner Object.
building in that tutorial.	Ø No results were found.		Activate Wisdom of Crowd	5	Jump to Tutorial Process

5. Cari Splid Data di Operators lalu double klik



6. Cari id3 di Operators lalu double klik



7. Cari Apply Model di Operators lalu double klik



- 8. Double klik pada Split Data, tentukkan Rationya
 - a. Tentukan Ratio pertama : Bagi sisa jumlah sisa data yang akan diuji dengan jumlah total data, misalkan jumlah data 6 sisa data 5 berarti 5:6=0.833 untuk ratio pertama
 - b. Tentukkan Ratio kedua : Bagi jumlah data yang ingin diiuji dengan jumlah total datasheet, misalkan 1:6=0.167
- 9. Sambungkan antara
 - a. Read Excel dengan Splid Data,
 - b. Splid Data dengan ID3,
 - c. ID3 dengan Apply Model,
 - d. Apply Model dengan Result,
 - e. ID3 dengan Result,
 - f. Parameter yang ada di Split data dengan unl pada Apply Model
 - g. Parameter yang ada di Split Data dengan Result
 - h. Out Multiply dengan Result



10. Klik tombol Play diatas, secara otomatis Rapid Miner akan membuat pohon keputusan



TUGAS 5

Nama: Vero FalorisNim: 202420032Kelas: MTI 23 Reguler AMk: Advanced Database

LINK RUJUKAN :

https://www.academia.edu/7712860/Belajar Data Mining dengan R apidMiner

https://medium.com/@ksnugroho/menerapkan-model-machinelearning-pada-rapidminer-142259846e13

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Dalam tutorial ini, kita akan membuat tiga model klasifikasi yang berbeda untuk data Titanic kita: pohon keputusan (decision tree),seperangkat aturan (a set of rules), dan model Bayes. Kita akan menjelajahi model-model tersebut dan melihat apakah kita bisa mengetahui lebih banyak tentang peristiwa kecelakaan itu dan lebih memahami siapa yang memiliki kesempatan terbaik untuk bertahan hidup.

Langkah 1. Menyiapkan data pelatihan

Seret data Titanic Training dari Sampel repositori ke dalam proses Anda.



Kita telah menyiapkan data Pelatihan Titanic untuk model pelatihan: dalam artian tidak ada nilai yang hilang serta labelnya telah didefinisikan. Harap diingat bahwa label adalah atribut yang ingin kita prediksi , dalam kasus ini : selamat(survived). kita memerlukan data pelatihan dengan label yang dikenal sebagai masukan untuk metode pembelajaran mesin semacam ini. Inilah sebabnya mengapa kita menyebut metode pembelajaran yang diawasi (*supervised learning*)

Langkah 2. Membangun tiga model yang berbeda.

Process Operators X 100% 🔎 🔑 📮 🏹 📝 Process deci × Modeling (8) Process -Predictive (8) Retrieve Titanic Trai... Decision Tree 🕶 📒 Trees (8) C out tra mod) inp Decision Tree exa Random Forest Naive Gradient Boosted Trees Bayes ID3 Rule Decision Stump Induction Decision Tree (Multiway Decision Tree (Weight-E) Random Tree

1. Seret pada operator **Decision Tree** dan hubungkan ke port "**out**" dari **Retrieve Titanic Training**.

2. Seret di operator **Naive Bayes** dan hubungkan port input sampel set nya dengan output "exa" dari **Decision Tree**.



3. Seret ke operator **Rule Induction** dan hubungkan contohnya dengan set port input dengan output "exa" dari **Naive Bayes**.



4. Hubungkan port "mod" dari operator pemodelan ke port hasil "res" di kanan, lalu jalankan prosesnya.

5. Periksa tiga model yang berbeda.

Hasil prediksi dari Rule Model



Hasil prediksi dari Naive Bayes



Hasil prediksi dari Decision Tree (Pohon Keputusan)



Pohon keputusan dengan jelas menunjukkan bahwa ukuran keluarga lebih menentukan daripada kelas penumpang untuk wanita. Pola perilaku ini tidak bisa dideteksi untuk pria. Secara umum, pria memiliki kemungkinan lebih rendah untuk bertahan hidup ("dahulukan wanita dan anak "). Cara termudah untuk melihat hal ini adalah pada visualisasi Chart model **Naive Bayes**. Meski biasanya Naive Bayes bukan tipe model yang paling akurat, secara umum aturan yang ditetapkan adalah format yang mudah dibaca, di mana bisa berguna saat kita ingin menafsirkan model.

BELAJAR DATA MINING DENGAN RAPID MINE

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana

Identitas

Belajar Data Mining dengan RapidMiner

Penyusun:

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor: Remi Sanjaya

Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C

Kata Pengantar

Dengan mengucapkan puji syukur kepada Tuhan YME atas Berkah Rahmat dan Hidayah-Nya, penulis dapat menyelesaikan buku yang berjudul Belajar Data Mining dengan RapidMiner.

Produk-produk perangkat lunak gratis (freeware) dan bersifat open source yang demikian banyak jumlahnya, telah memudahkan kita dalam melakukan proses pengolahan dan analisis data. Dalam melakukan analis terhadap data mining, RapidMiner merupakan salah satu solusi yang dapat kita gunakan. Keberadaan RapidMiner yang berupa freeware dan dapat dijalankan pada berbagai sistem operasi tidak hanya menguntungkan penyedia aplikasi karena tidak perlu mengeluarkan biaya untuk lisensi perangkat lunak, tetapi juga memudahkan pengembang maupun calon pengembang dalam mempelajari dan mencoba sendiri fitur-fitur yangada. Buku ini diharapkan dapat membantu pembaca mempelajari RapidMiner, melalui rangkaian tutorial bertahap mulai dari proses instalasi hingga pemrograman. Pada buku ini juga dibahas beberapa teori penunjang mengenai data mining seperti, decision tree, neural network dan market basket analysis untuk membuka wawasan pembaca mengenai data mining sebelum melakukan analisis data mining.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu penyelesaian buku ini.

Akhir kata, penulis menyadari masih terdapat kekurangan dalam penyusunan buku ini baik pada teknis penulisan maupun materi, mengingat akan kemampuan yang dimiliki penulis. Untuk itu kritik dan saran dari semua pihak penulis harapkan demi penyempurnaan pembuatan buku ini. Semoga buku ini dapat bermanfaat bagi para pembaca.

Jakarta, April 2013

Penulis

Daftar Isi

Kata Pengantar	. i
Daftar Isii	iii
Daftar Gambar	v
Daftar Tabelv	iii
Kecerdasan Buatan	2
Definisi Kecerdasan Buatan	.2
Ruang Lingkup Kecerdasan Buatan	5
Perbedaan Komputasi Kecerdasan Buatan dan KomputasiKonvensional	.6
RapidMinerError! Bookmark not defined.	8
ApaituRapidMiner?	.8
Instalasi Software 1	1
Pengenalan Interface 1	6
Cara Menggunakan Repositori2	28

iii | Pengantar

Data Mining	
Mengenal Data Mining	
Pengelompokan TeknikData Mining43	
Decision Tree)
Mengenal Decision Tree45)
Algoritma c4.548	
Kelebihan Pohon Keputusan55)
Kekurangan Pohon Keputusan56	i
Decision Treepada RapidMiner56	j
Neural Network	
Market Basket Analysis96)
Memahami MarketBasket Analysis	j
Metodologi Association Rules100	
Contoh Association Rules102	
Frequent Itemset Generation dan Rule Generatio	n
Market Basket Analysis pada RapidMiner 107	,
Glossarium	•
Daftar Pustaka125	;

Daftar Gambar

Gambar 1.1 Proses Kecerdasan Buatan	4
Gambar 2.1 Form Awal Instalasi	14
Gambar 2.2 Form Persetujuan Lisensi	14
Gambar 2.3 Form Pemilihan Lokasi Instalasi	15
Gambar 2.4 Form Proses Instalasi	15
Gambar 2.5 Form Instalasi selesai	16
Gambar 2.6 Tampilan Welcome Perspective	17
Gambar 2.7 Welcome Perspective	19
Gambar 2.8 Header Tab	20
Gambar 2.9 Tampilan Design Perspective	21
Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki	23
Gambar 2.11 Tampilan Parameter View	25
Gambar 2.12 Problem & Log View	27
Gambar 2.13 Kumpulan Sample Data Repository	28
Gambar 2.14 Tampilan Design Perspective Awal	29
Gambar 2.15 Repository berada dalam Main Process	29
Gambar 2.16 Menghubungkan Output Repositori ke Result	30
Gambar 2.17 Isi Sample Golf Data Repository	30
Gambar 2.18 Repository	32
Gambar 2.19 Step 1 of 5 Import Wizard	32
Gambar 2.20 Step 2 of 5 Import Wizard	33
Gambar 2.21 Step 3 of 5 Import Wizard	34
Gambar 2.22 Step 4 of 5 Import Wizard	34
Gambar 2.23 Tipe Data	35

Gambar 2.24 Step 5 of 5 Import Wizard	
Gambar 2.25 Repository yang sudah diimport	
Gambar 2.26 Menghubungkan Output Repositori pada Result	
Gambar 2.27 Tabel Repository	
Gambar4.1Bentuk Decision Tree Secara Umum	
Gambar 4.2 Grafik Entropi	50
Gambar 4.3 Tabel Keputusan dalam Format xls	
Gambar4.4LokasiTabelpadaRepository	
Gambar 4.5 Repository PlayGolfpada Main Process	
Gambar 4.6 Daftar Operator pada View Operators	
Gambar 4.7 Posisi Operator Decision Tree	60
Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator D	ecision
Tree	61
Gambar 4.9 Parameter Decision Tree	
Gambar 4.10 Tipe Criterion	62
Gambar 4.11 lkon Run	
Gambar 4.12 Hasil Berupa Graph Pohon Keputusan	
Gambar 4.13 Hasil Berupa Penjelasan Teks	67
Gambar 4.14 Tabel SakitHipertensi dalam format xls	
Gambar 4.15 Lokasi Tabel pada Repository	
Gambar 4.16 Tabel Sakit Hipertensi pada Main Process	70
Gambar 4.17 Hirarki Operator X-Validation	72
Gambar 4.18 Operator Validation	72
Gambar4.19ParameterX-Validation	74
Gambar 4.20 Hirarki Operator Apply	77
Gambar 4.21 Operator Apply Model	
Gambar 4.22 Parameter Apply Model	
Gambar 4.23 Hirarki Operator Performance	
Gambar 4.24 Operator Performance	
Gambar 4.25 Parameter Performance	
Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Pe	erformance
	82
Gambar 4.27 Susunan Operator Retrieve dengan Operator Validat	ion 83
Gambar4.28Tampilan Decision Tree	
Gambar 6.1 Frequent Item Set tanpa Apriori	106
Gambar 6.2 Frequent Item Set dengan Apriori	106

Gambar6.3TabelPenjualanSederhana	108
Gambar 6.4 Repositori	108
Gambar 6.5 Database dalam Main Process	109
Gambar 6.6 Operator Create Association Rules	109
Gambar 6.7 Operator FP-Growth	110
Gambar 6.8 Operator Numerical to Binominal	110
Gambar 6.9 Pencarian Operator Numerical to Binominal	111
Gambar 6.10 Pencarian Association Rules	112
Gambar 6.11 Menghubungan Database TransaksiMakanan pada C)perator
Numerical to Binomial	112
Gambar 6.12 Parameter Numerical to Binomial	113
Gambar 6.13 Menghubungkan Operator Numerical to Binomial	dengan
Operator FP-Growth	114
Gambar 6.14 Parameter FP-Growth	115
Gambar6.15MenghubungkanOperatorFP-GrowthdenganOperatorFP-Growth	perator
CreateAssociationRules	115
Gambar 6.16 Parameter Association Rules	116
Gambar 6.17 Susunan Operator Association Rules	117
Gambar 6.18 Hasil Association Rules Pertama	117
Gambar 6.19 Operator FP-Growth	118
Gambar 6.20 Mengubah Parameter FP-Growth	119
Gambar 6.21 Operator Create Association Rules	119
Gambar 6.22 Mengubah Parameter Association Rules	120
Gambar 6.23 Hasil Association Rules Kedua	120
Gambar 6.24 Hasil dalam bentuk Graph View	121

Daftar Tabel

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konvensional	7
Tabel 4.1 Keputusan Bermain Tenis	52
Tabel 4.2 Perhitungan Simpul 1	53
Tabel6.1TabelTransaksi	102
Tabel 6.2 Kombinasi Produk dan Nilai Support	103
Tabel 6.3 Association Rules dan Nilai Confidence	104

Bagian Satu Pendahuluan

Pengenalan Kecerdasan Buatan

Pengenalan RapidMiner

Chapter 1

Kecerdasan Buatan

Definisi Kecerdasan Buatan

Manusia memiliki kecerdasan, manusia memiliki kemampuan untuk menganalisa suatu masalah dengan menggunakan pengetahuan dalam otaknya dan

pengalaman yang pernah dilaluinya. Pengetahuan datang ketika manusia belajar, maka dari itu pembelajaran merupakan faktor penting bagi manusia untuk mencapai sebuah kecerdasan. Namun pengetahuan tidak akan cukup untuk menyelesaikan masalah jika tidak memiliki pengalaman, karena pengalaman akan selalu membawa pengetahuan baru. Tetapi akan sia sia, jika seseorang yang memiliki banyak pengalaman tetapi tidak memiliki akal untuk menalar semua pengetahuan dan pengalaman yang ia miliki. Kombinasi dari pengetahuan, pengalaman, dan kemampuan menalar inilah yang membuat manusia menjadi cerdas dan dapat menyelesaikan permasalahan yang ia hadapi.

Berdasarkan konsep diataslah kecerdasan buatan dibuat. Agar mesin dapat bertindak seperti seorang manusia, maka mesin tersebut harus memiliki sejumlah pengetahuan dan pengalaman serta kemampuan menalar yang dapat mengubah pengetahuan dan pengalaman tersebut menjadi sebuah keputusan dalam menyelesaikan sebuah permasalahan.

Komputer awalnya diciptakan hanya untuk melakukan sebuah perhitungan saja. Jaman terus berkembang hingga akhirnya komputer kini diberdayakan manusia untuk membantu pekerjaannya dalam kesehariannya. Maka dari itu komputer diharapkan memiliki kemampuan yang hampir sama dengan manusia agar dapat mengerjakan segala sesuatu yang bisa dikerjakan oleh manusia – Kecerdasan Buatan. The art of creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)

The study of how to make computers do things at which, at the moment, people are better (Rich dan Knight, 1991)

A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes (Schalkoff, 1990)

The branch of computer science that is concerned with the automation of intelligent behavior (Luger dan Stubblefield, 1993)

Jadi apakah kecerdasan buatan itu? Kecerdasan buatan adalah salah satu bagian dari ilmu komputer yang membuat agar mesin dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia. Dengan demikian, untuk menciptakan sebuah aplikasi kecerdasan buatan terdapat dua bagian utama yang sangat dibutuhkan.



Gambar 1.1 Proses Kecerdasan Buatan

Ruang Lingkup Kecerdasan Buatan

Kecerdasan buatan merupakan teknologi yang fleksibel, dan dapat diterapkan di berbagai macam bidang ilmu. Kemampuan kecerdasan buatan menjadi sangat dibutuhkan di bidang ilmu lain, karena konsepnya tak lagi procedural melainkan meniru cara berpikir manusia. Tak heran kecerdasan buatan bisa di gunakan untuk bidang psikologi yang dikenal dengan cognition dan psycolinguistic. Namun yang paling sering dekat dengan kita ialah robotika, yakni kecerdasan buatan di dalam ilmu elektornika.

Semakin banyaknya ilmu yang menggunakan kecerdasan buatan, semakin sulit juga bagi manusia untuk mengkategorikannya, maka dari itu dibentuklah ruang lingkup kecerdasan buatan yang dapat mewakilinya (Turban dan Frenzel, 1992, pp21-26):

 Sistem Pakar. komputer digunakan untuk menyimpan pengetahuan para pakar. Dengan demikian komputer akan memiliki keahlian untuk menyelesaikan permasalahan dengan meniru keahlian yang dimiliki oleh pakar.

- 2. Pengolahan Basa Alami. dengan pengolahan bahasa alami ini diharapkan user dapat berkomunikasi dengan komputer dengan menggunakan bahasasehari-hari.
- 3. Pengenalan Ucapan. Melalui pengenalan ucapan diharapkan manusia dapat berkomunikasi dengan komputer dengan menggunakan suara.
- 4. Robotika dan Sistem Sensor
- 5. Computer Vision. Mencoba untuk dapat menginterpretasikan gambar atau objek-objek tampak melalui komputer.
- Intelligent Computer-aided Instruction. Komputer dapat digunakan sebagai tutor yang dapat melatih danmengajar.
- 7. Game Playing.

Perbedaan Komputasi Kecerdasan Buatan dan Komputasi Konvensional

Komputasi Konvensional merupakan Komputer yang hanya digunakan untuk alat hitung. Sangatlah berbeda, kerja dan konsep dari kedua komputasi ini. Agar dapat memberikan gambaran, table berikut adalah

detail dari perbedaan komputasi kecerdasan buatan dan komputasi konvensional.

Dimensi	Komputasi Kecerdasan	Komputasi
	Buatan	Konvensional
Pemrosesan	Mengandung konsep-	Algoritmik
	konsep simbolik	
Sifat Input	Bisa tidak lengkap	Harus lengkap
Pencarian	Kebanyakan bersifat	Biasanya didasarkan
	heuristic	pada algoritma
Keterangan	Disediakan	Biasanya tidak
		disediakan
Fokus	Pengetahuan	Data dan Informasi
Struktur	Kontrol dipisahkan dari	Kontrol terintegrasi
	pengetahuan	dengan informasi
Kemampuan	Ya	Tidak
menalar		

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konvensional

RapidMiner

Apa itu RapidMiner?

RapidMiner merupakan perangakat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan munggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya dikembangkan pada tahun 2001 mulai oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan mengunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML)Extensible Markup Language) yang mendefenisikan proses analitis keingginan pengguna untuk diterpkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analis secara otomatis. RapidMiner memiliki beberapa sifat sebagai berikut:

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukarandata.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari RapidMiner, antara lain:

- Banyaknya algoritma data mining, seperti decision treee dan self-organization map.
- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation,

loading), data preprocessing, visualisasi, modelling dan evaluasi

- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI
- Mengintegrasikan proyek data mining Weka dan statistika R

Instalasi Software

System Requirement

Sebelum melakukan instalasi software RapidMiner, terdapat beberapa spesifikasi minimal yang yang harus dimiliki komputer pengguna. Spesifikasi minimal bergantung pada komputer dan sistem operasi yang akan diinstal. Berikut ini beberapa spesifikasi minimal yang dibutuhkan software RapidMiner:

1. Sistem Operasi

RapidMiner merupakan software yang multiplatform, sehingga software ini dapat dijalankan pada berbagai sistem operasi. Berikut ini beberapa jenis sistem operasi yang dapat diinstal RapidMiner:

- Microsoft Windows (x86-32) Windows XP, Windows Server 2003, Windows Vista, Windows Server 2008, Windows 7
- Microsoft Windows (x64) 2 Windows XP untuk x64, Windows Server 2003 untuk x64, Windows Vista untuk x64, Windows Server 2008 untuk x64, Windows 7 untuk x64
- 2 Unix sistem 32 atau 64 bit
- Inux sistem 32 atau 64 bit
- Apple Macintosh sistem 32 atau 64 bit Sebagai bahan pertimbangan, kami merekomendasikan untuk penggunaan sistem 64 bit. Hal ini dikarenakan jumlah maksimum yang dapat digunakan oleh RapidMiner terbatas pada sistem operasi dengan sistem 32, yaitu hanya sebasar 2GB.
- 2. Java Runtime Environment versi 6

Selain itu, penggunaan server RapidAnalytics dalam kombinasi dengan RapidMiner dapat memaksimalkan proses analisis pada RapidMiner, meskipun tugas analisis sudah banyak dapat dijalankan dengan RapidMiner desktop client. Dalam hal ini proses analisa dirancang dengan RapidMiner, kemudian dieksekusi oleh server RapidAnalytics.

Instalasi RapidMiner

Seperti yang yang telah dikemukakan sebelumnya bahwa RapidMiner merupakan software gratis yang bersifat terbuka (open source). Software ini dapat dijalankan pada sistem operai Windows, Linux, maupun Mac. RapidMiner dapat diunduh pada situs resminya, yaitu <u>www.rapid-i.com</u>. Pada bagian ini, akan dijelaskan bagaimana cara melakukan instalasi software RapidMiner versi 5.3 pada sistem operasi Microsoft Windows.

Untuk memulai instalasi software RapidMiner pada sistem operasi Microsoft Windows, jalankan file installer RapidMiner-5.3.000x32-install.exe, sehingga akan muncultampilan wizard seperti pada Gambar 2.



Gambar 2.1 Form Awal Instalasi

Klik **Next >** untuk melanjutkan pada form persetujuan dan lisensi seperti pada Gambar 2.3



Gambar 2.2 Form Persetujuan Lisensi

Pilih **I Agree** untuk melanjutkan. Kemudian, wizard akan menampilkan form seperti pada gambar 2.4.

Kapidiviner 3 Setup	
Choose Install Location Choose the folder in which to install RapidMiner 5.	
Please select the folder where RapidViner 5 should be installed. administrator rights you should select a directory into which you your home directory. Installing into a writable directory is especi use the automatic update service of RapidViner 5 without having	If you are a user without can write, e.g. a directory in ally important if you want to g administrator rights.
Destination Folder	Browse
Space required: 187.7MB Space available: 63.8GB	

Gambar 2.3 Form Pemilihan Lokasi Instalasi

Pilih **Install** untuk melakukan proses instalasi. Kemudian wizard akan menampilkan progress dari proses tersebut, seperti yang ditunjukkan pada Gambar 2.5.

3 RapidMiner 5 Setup	
Installation Complete Setup was completed successfully.	
Completed	
Show details	
	< Back Next > Cancel

Gambar 2.4 Form Proses Instalasi
Setelah proses selesai, pilih **Next** > untuk melanjutkan, maka wizard akan menampilkan informasi bahwa proses instalasi telah selesai dilakukan, seperti yang ditunjukkan pada Gambar2.6.



Gambar 2.5 Form Instalasi selesai

Pilih Finish untuk mengakhiri proses instalasi.

Pengenalan Interface

RapidMiner menyediakan tampilan yang *user friendly* untuk memudahkan penggunanya ketika menjalankan aplikasi. Tampilan pada RapidMiner dikenal dengan istilah Perspective. Pada RapidMiner terdapat 3 Perspective, yaitu; Welcome Perspective, Design Perspective dan Result Perspective.

Welcome Perspective

Ketika membuka aplikasi Anda akan disambut dengan tampilan yang disebut dengan Welcome Perspective, seperti yang ditunjukkan pada Gambar 6. Pada bagian toolbar, terdapat toolbar **Perspectives** yang terdiri dari ikon-ikon untuk menampilkan perspective dari RapidMiner. Toolbar ini dapat dikonfigurasikan sesuai dengan kebutuhan Anda. Sedangkan **Views** menunjukkan pandangan (view) yang sedang Anda tampilkan.



Gambar 2.6 Tampilan Welcome Perspective

Jika komputer Anda terhubung dengan internet, maka pada bagian bawah Welcome Perspective akan menampilkan berita terbaru mengenai RapidMiner. Bagian ini dinamakan **News**. Pada bagian tengah halaman terlihat daftar **Last Processes** (Recent Processes), bagian ini menampilkan daftar proses analisis yang baru saja dilakukan. Hal ini akan memudahkan Anda jika ingin melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut. Bagian **Actions** menunjukkan daftar aksi yang dapat Anda lakukan setelah membuka RapidMine. Berikut ini rincian lengkap daftar aksi tersebut:

- New : Aksi ini berguna ntuk memulai proses analis baru. Untuk memulai proses analisis, pertama-tama Anda harus menentukan nama dan lokasi proses dan Data Repository. Setelah itu, Anda bisa mulai merancang sebuah analisis baru.
- 2. **Open Recent Process** : Aksi ini berguna untuk membuka proses yang baru saja ditutup. Selain aksi ini, Anda juga bisa membuka proses yang baru ditutup dengan mengklik dua kali salah satu daftar yang ada pada Recent Process. Kemudian tampilan Welcome Perspective akan otomotasi beralih ke Design Perspective.
- 3. **Open Process**: Aksi ini untuk membuka Repository Browser yang berisi daftar proses. Anda juga bisa memilih proses untuk dibuka pada Design Perspective.
- 4. **Open Template** : Aksi ini menunjukkan pilihan lain yang sudah ditentukan oleh proses analisis.

5. Online Tutorial : Aksi digunakan untuk memulai tutorial secara online (terhubung internet). Tutorial yang dapat secara langsung digunakan dengan RapidMiner ini, memberikan perkanalan dan beberapa konsep data mining. Hal ini direkomendasikan untuk Anda yang sudah memiliki pengetahuan dasar mengenai data mining dan sudah akrab dengan operasi dasar RapidMiner.

RapidMiner dapat menampilkan beberapa view pada saat bersamaan. Seperti yang ditunjukkan pada Gambar 7, pada tampilan Welcome Perspective tedapat **Welcome view** dan **Log View**. Ukuran dari setiap view tersebut dapat diubah sesuai dengan kebutuhan Anda dengan Mengklik dan menarik garis batas diantara keduanya ke atas atau ke bawah.



Gambar 2.7 Welcome Perspective



Gambar 2.8 Header Tab

Anda bisa melakukan beberapa aksi terhadap view, dengan mengklik salah satu ikon yang tampak pada bagian view, seperti yang ditunjukkan pada gambar 2.8. Berikut ini beberapa aksi yang dapat Anda lakukan:

- 1. **Close** : Aksi ini untuk menutup view yang ditampilkan pada perspective. Anda bisa menampilkan view kembali dengan mengklik menu view dan memilih view yang ingin ditampilkan.
- 2. *Maximize* : Aksi ini untuk memperbesar ukuran view pada perspective.
- 3. *Minimize*: Aksi ini untuk memperkecil ukuran view pada perspective.
- Detach : Aksi ini untuk melepaskan view dari perspective menjadi jendela terpisah, kemudian Anda juga dapat memindahkannya sesuai dengan keinginan Anda.

Design Perspective

Design Perspective merupakan lingkungan kerja RapidMiner. Dimana Design Perspective ini merupakan perspective utama dari RapidMiner yang digunakan sebagai area kerja untuk membuat dan mengelola proses analisis. Seperti yang ditunjukkan pada Gambar 2.10, perspective ini memiliki beberapa view dengan fungsinya masing-masing yang dapat mendukung Anda dalam melakukan proses analisis data mining. Anda bisa mengganti perspective dengan mengklik salah satu ikon dari tollbar perspective yang sebelumnya telah dijelaskan. Selain dengan cara tersebut, Anda juga bisa mengganti perspective dengan mengklik menu view, kemudian pilih perspective, lalu pilih perspective yang ingin Anda tampilkan.



Gambar 2.9 Tampilan Design Perspective

Sebagai lingkungan lingkungan kerja, Design Perspective memiliki beberapa view. Berikut ini beberapa view yang ditampilkan pada Design Perspective:

1. Operator View

Operator View merupakan view yang paling penting pada perspective ini. Semua operator atau langkah kerja dari RapidMiner disajikan dalam bentuk kelompok hierarki di Operator View ini sehingga operator-operator tersebut dapat digunakan pada proses analisis, seperti yang ditunjukkan pada Gambar 2.10. Hal ini akan memudahkan Anda dalam mencari dan menggunakan operator yang sesuai dengan kebutuhan Anda. Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut:

- Process Control : Operator ini terdiri dari operator perulangan dan percabangan yang dapat mengatur aliran proses.
- Utility : Operator bantuan, seperti operator macros, loggin, subproses, dan lain-lain.
- Repository Access : Kelompok ini terdiri dari operator-operator yang dapat digunakan untuk membaca atau menulis akses pada repository.
- Import : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk membaca data dan objek dari format tertentu seperti file, database, dan lain-lain.
- Export : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk menulis data dan objek menjadi format tertentu.

- Data Transformation : kelompok ini terdiri dari semua operator yang berguna untuk transformasi data dan meta data.
- Modeling : kolompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru.
- Evaluation : kelompok ini berisi operator yang dapat digunakan untuk menghitung kualitas pemodelan dan untuk data baru.



Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki

2. Repository View

Repository View merupakan komponen utama dalam Design Perspective selain Operator View. View ini dapat Anda gunakan untuk mengelola dan menata proses Analisis Anda menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.

3. Process View

Process View menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Anda dapat menambahkan langkah baru dengan beberapa cara. hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali. Pada dasarnya bekerja dengan RapidMiner ialah mendefinisikan proses analisis, yaitu dengan menunjukkan serangkaian langkah kerja tertentu. Dalam RapidMiner, komponen proses ini dinamakan sebagai operator. Operator pada RapidMiner didefinisikan sebagai beikut:

- Deskripsi dari input yang diharapkan.
- Deskripsi dari output yang disediakan.
- Tindakan yang dilakukan oleh operator pada input, yang akhirnya mengarah dengan penyediaan output.
- Sejumlah parameter yang dapat mengontrol *action performed*.

4. Parameter View

Beberapa operator dalam RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar. Namun terkadang parameter tidak mutlak dibutuhkan, meskipun eksekusi operator dapat dikendalikan dengan menunjukkan nilai parameter tertentu. Parameter view memiliki toolbar sendiri sama seperti view-view yang lain. Pada Gambar 2.12, Anda dapat melihat bahwa pada Parameter View ini terdapat beberapa ikon dan nama-nama operator terkini yang dikuti dengan aktual parameter.



Gambar 2.11 Tampilan Parameter View

Huruf tebal berarti bahwa parameter mutlak harus didefinisikan oleh analis dan tidak memiliki nilai default. Sedangkan huruf miring berarti bahwa parameter diklasifikasikan sebagai parameter ahli dan seharusnya tidak harus diubah oleh pemula untuk analisis data. Poin pentingnya ialah beberapa parameter hanya ditunjukkan ketika parameter lain memiliki nilai tertentu.

5. Help & CommentView

Setiap kali Anda memilih operator pada Operator View atau Process View, maka jendela bantuan dalam Help View akan menunjukkan penjelasan mengenai operator ini. Penjelasn yang ditampilkan dalam Help View meliputi:

- Sebuah penjelasan singkat mengenai fungsi operator dalam satu atau beberapa kalimat.
- Sebuah penjelasan rinci mengenai fungsi operator.
- Daftar semua parameter termasuk deskripsi singkat dari parameter, nilai default (jika tersedia), petunjuk apakah parameter ini adalah parameter ahli serta indikasi parameter dependensi.

Sedangkan Comment View merupakan area bagi Anda untuk menuliskan komentar pada langkahlangkah proses tertentu. Untuk membuat komentar, Anda hanya perlu memilih operator dan menulis teks di atasnya dalam bidang komentar. Kemudian komentar tersebut disimpan bersama-sama dengan definisi proses Anda. Komentar ini dapat berguna untuk melacak langkah-langkah tertentu dalamrancangan nantinya.

6. Problem & LogView

Problem View merupakan komponen yang sangat berharga dan merupkan sumber bantuan bagi Anda selama merancang proses analisis. Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam Problem View, seperti yang ditunjukkan pada Gambar 2.13.



Gambar 2.12 Problem & Log View

Pada kolom Message, Anda akan menemukan ringkasan pendek dari masalah. Kolom Location berisi tempat di mana masalah muncul dalam bentuk nama Operator dan nama port input yang bersangkutan. Kolom Fixes memberikan gambaran dari kemungkinan solusi tersebut, baik secara langsung sebagai teks (jika hanya ada satu kemungkinan Solusi) atau sebagai indikasi dari berapa banyak kemungkinan yang berbeda untuk memecahkan masalah.

Cara Menggunakan Repositori

Repositori merupakan Tabel, database, koleksi teks, yang kita miliki untuk dapat digali datanya untuk mendapatkan informasi yang kita inginkan. Ini merupakan awal dari seluruh proses Data Mining. Maka dari itu adalah penting bagi kita untuk mengetahui cara menggunakan repository.

Sample Data Repository

RapidMiner menyediakan contoh database yang dapat digunakan, berikut cara menggunakan Sample Data Repository.



Gambar 2.13 Kumpulan Sample Data Repository

Pada bagian Repositori terdapat 3 buah lokasi repositori, yakni Samples, DB dan Local Repository.

Untuk mengambil Sample Data Repository, buka hirarki Samples, masuk ke folder Data. Sehingga seperti gambar berikut.



Gambar 2.14 Tampilan Design Perspective Awal

Lakukan Drag dan Drop salah satu Example Repository. Kita ambil contoh Golf. Tarik dan lepaskan repository ke dalam Main Process, sehingga seperti gambar berikut.



Gambar 2.15 Repository berada dalam Main Process



Gambar 2.16 Menghubungkan Output Repositori ke Result

Hubungkan output pada Database ke Result seperti Gambar diatas. Lalu klik ikon Play Gambar 2.17 adalah Sample data repository dari Golf. Coba lakukan untuk memasukkan Sample Repository yang lain.

🛛 🐺 Resu	It Overview	🗙 🖊 📑 Exa	mpleSet (Retri	eve Golf) 🏾	
Data Vie	w 🔘 Meta D	ata View 🔘 F	Plot View 🔘 A	dvanced Cha	rts 🔘 Annota
ExampleSe	t (14 example	s, 1 special at	tribute, 4 regul	ar attributes)	
Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

Gambar 2.17 Isi Sample Golf Data Repository

Import Repository

Dibanyak kesempatan lain, kita akan selalu menggunakan database yang kita miliki. RapidMiner menvediakan lavanan pengguna adar dapat mengimport database miliknya. Namun, tidak seperti kebanyakan tools Data Mining Lain, RapidMiner memiliki kelebihan tersendiri yakni dapat langsung melakukan import file dengan ekstensi .xls atau .xlsx, yakni file dari Microsoft Excel, Program yang relatif sering digunakan oleh pengguna. Berikut adalah cara untuk melakukan import file Microsoft Excel.

Lihat pada bagian Repository. Klik pada ikon import seperti gambar 2.18. Seperti yang dapat kita lihat, ada beberapa ekstensi file yang dapat kita masukkan kedalam repository kita. CSV File, Excel Sheen File, Access Database Table File, Database Table, Binary File. Namun pada Dasarnya cara melakukan import pada semua file ini sama. Sebagai contoh, pilih Import Excel Sheet.



Gambar 2.18 Repository

Setelah itu, akan muncul window baru yakni Step 1 dari 5 Step Data import Wizard. Disini akan diarahkan oleh RapidMiner bagaimana langkah untuk melakukan import data.



Gambar 2.19 Step 1 of 5 Import Wizard

Cari file excel kalian dengan klik pada bagian Look in

vang dibutuhkan lalu Klik tombol Next 🖃 🗠.

Berikutnya pada Step 2 ialah, pilih Sheet yang akan dimasukkan. Pada dasarnya, Repository RapidMiner hanya menyediakan 1 repositori untuk 1 buah table.

8	Data import wizard - Step 2 of 5						
"A	This wizard guides you to import your data. • Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into RapidNiner. Furthermore, you can mark a range of cells to be loaded.						
Sheet1		0		-			
	TEMPEDATI		WINDY	PLAY			
SUNNY			false	Don't Play			
SUNNY			true	Don't Play			
OVERCAST			false	Play			
RAIN				Play			
RAIN				Don't Play			
RAIN				Play			
OVERCAST							
SUNNY							
SUNNY							
RAIN			false	Play			
	Previous Next Finish Scancel						

Gambar 2.20 Step 2 of 5 Import Wizard

Klik tombol Next 🕬 Berikutnya ialah memberikan anotasi. Jika data kita tidak memiliki nama attribute, tidak usah melakukan apa-apa pada step 3 ini.

8	Data import wizard - Step 3 of 5						
"/*	This wizard guides you to import your data. Step 3: In RapidMiner, each attribute can be annotated. The most important annotation of antitrubte is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.						
Annotation	А	в	С	D	E		
Name	OUTLOOK	TEMPERATU	HUMIDITY	WINDY	PLAY		
-	SUNNY	85	85	false	Don't Play		
-	SUNNY	80	90	true	Don't Play PLAY		
-	OVERCAST	83	78	false	Play		
-	RAIN	70	96	false	Play		
-	RAIN	68	80	false	Don't Play		
-	RAIN	65	70	true	Play		
-	OVERCAST	64	65	true	Don't Play		
-	SUNNY	72	95	false	Play		
-	SUNNY	69	70	false	Play		
-	RAIN	75	80	folco	Play		
	ΡΔΙΝΙ	75	revious	halea	Einish Einish	ancel	

Gambar 2.21 Step 3 of 5 Import Wizard

Klik tombol Next 💌 Step ke 4 adalah memberikan tipe data pada tabel kita. Sebenarnya RapidMiner akan memberikan tipe data yang tepat secara otomatis.

Beload data	Guess value types	Preview uses only	inst 100 rows. Date for	mat
Ø	0	Ø	Ø	Ø
OUTLOOK	TENPERATURE	HUVIDITY	WINDY	PLAY
polynominal +	integer +	Integer +	binominal +	binominal -
attribule	attribute -	attribute *	aticibute =	att-bute +
SUNNY	85	80	talse	Don't Play
BUININ/	80	90	tue	Dom't Play
OVERCAST	83	70	abe	21aa
RAIN	70	20	talse	Play
RAN	60	80.	taise	Don't Play
0				D
0 enors.			i i	a longre errors 🔲 Show only error
Bau Course		Error Orional value		Herris

Gambar 2.22 Step 4 of 5 Import Wizard

Namun, jika kita merasa tipe data yang diberikan RapidMiner tidak cocok, kita bisa mengubahnya.



Gambar 2.23 Tipe Data

Klik tombol Next Rep. Step ke 5 adalah memasukkan database kita kedalam repository. Disarankan untuk memasukkannya kedalam Local Repository untuk memudahkan kita mencarinya. Jangan lupa untuk memberikan nama repository kita.



Gambar 2.24 Step 5 of 5 Import Wizard

Kemudian klik tombol finish

Hasil Import Repository akan terlihat pada bagian Repository seperti dalam gambar 2.25.



Gambar 2.25 Repository yang sudah diimport

Untuk melihat isi dari repository kita, hubungkan output pada repository kearah result sepertigambar 2.26.



Gambar 2.26 Menghubungkan Output Repositori pada Result

klik ikon Play Dan berikutnya akan muncul isi dari tabel yang kalian miliki.

🛛 🛒 Resu	lt Overview 🖇	👔 🗐 Exam	npleSet (#Loc	al Repository/	data/golfdatas	et) 🔀 🔪	
Data View O Meta Data View O Plot View O Advanced Charts O Annotations							
ExampleSet	(14 examples	, 0 special attr	ributes, 5 regu	ılar attributes)			
Row No.	OUTLOOK	TEMPERAT	HUMIDITY	WINDY	PLAY		
1	SUNNY	85	85	false	Don't Play		
2	SUNNY	80	90	true	Don't Play		
3	OVERCAST	83	78	false	Play		
4	RAIN	70	96	false	Play		
5	RAIN	68	80	false	Don't Play		
6	RAIN	65	70	true	Play		
7	OVERCAST	64	65	true	Don't Play		
8	SUNNY	72	95	false	Play		
9	SUNNY	69	70	false	Play		
10	RAIN	75	80	false	Play		
11	SUNNY	75	70	true	Play		
12	OVERCAST	72	90	true	Play		
13	OVERCAST	81	75	false	Play		
14	RAIN	71	80	true	Don't Play		

Gambar 2.27 Tabel Repository

Bagian Dua Data Mining

Pengenalan Data Mining

Pengenalan Decision Tree

Pengenalan Neural Network

Pengenalan Market Basket Analysis

Chapter 3

Data Mining

Mengenal Data Mining

Pengertian Data Mining

Sebelum kita mulai, ayo kita coba beberapa eksperimen sebagai berikut.

- Pilih angka antara 1 sampai 10
- Kalikan dengan angka9
- Hasil dari perkalian tersebut jumlahkan masingmasing angkanya
- Kalikan hasil dengan4
- Bagi dengan 3
- Kurangi dengan 2

Jawabannya adalah 2. Kebetulan? Sebagai seorang analis, pasti jawabannya adalah tidak.

Bagaimana dengan kejadian acak lainnya, seperti čler par koiš.č Teštu jika terašru rešeHak secara langsung dan hasil dari kejadian tersebut ternyata tepat seperti yang temanmu tebak, kau pasti akan mengatakan bahwa itu merupakan kebetulan.

Kita ambil satu contoh sederhana lagi. Terdapat kejadian seperti: Seseorang menjatuhkan sebuah gelas dari ketinggian tertentu. Detik pertama orang tersebut menjatuhkan gelasnya, kau pasti akan mengatakan dengan pasti bahwa gelas tersebut akan pecah, padahal hukum fisika belum menunjukkan proses penghancuran gelas tersebut ketika bersentuhan dengan tanah. Dan lagi, tebakanmu itu dikatakan bukanlah kebetulan. Jadi secara logika, bagaimana kau tahu dengan sangat tepat hasil dari kejadian tersebut? Bukankah kondisinya sama seperti kejadia šelera par koi se seHelura segara

Jadi apakah yang kita lakukan dalam otak kita? Kita mempertimbangkan karakteristik-karakteristik dari kejadian ini. Pada kasus gelas yang jatuh, kita dengan cepat mengetahui karakteristik penting dari serangkaian kejadian tersebut, bahan gelas, ketinggian, tipe pijakan, dan lain-lain. Kemudian kita menjawab dengan cepat berdasarkan analogi, contohnya kita kita

membuat perbandingan dengan kejadian gelas atau canokir atau piring yang jatuh sebelumnya. Berarti dua hal diperlukan adalah: vand pertama. kita membutuhkan data dari kejadian-kejadian sebelumnya, dan kedua, seberapa mirip kejadian yang di tempat dengan kejadian sebelumnya. Kita bisa membuat estimasi atau prediksi dengan mencari kejadian yang paling mirip dengan kejadian di tempat. Karena kita lebih sering melihat bahwa benda berbahan kaca dijatuhkan akan pecah, maka secara otomatis inilah yang menjadi prediksikita.

Bagaimanapun, prosedur diatas tidak cocok uituk kejadia derapar koi dise terdapat lebih banyak faktor yang harus dipertimbangkan, ada yang sulit dan ada yang tidak bisa diukur. Belum lagi kita harus dapat memikirkan proses kejadian menuju hasil dengan baik, memikirkan analogi yang paling cocok dengan kejadian untuk melakukan prediksi. Dita Hah derapar koi derariliki kodisi ya gelapat berubah-ubah tiap kejadiannya dan berlangsung cepat, ini berarti perhitungan juga harus dilakukan secara cepat. Mustahil untuk metode data mining.

Data Mining adalah serangkaian proses untuk menggali nlai tambah dari suatu kumpulan data

berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006)

Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya. (Pramudiono, 2006)

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005)

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. (Larose, 2005)

Kata *Mining* merupakan kiasan dari bahasa inggris, mine. Jika mine berarti menambang sumber daya yang tersembunyi di dalam tanah, maka Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu *Data Mining s*ebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), machine learning, statistik dan basis Data.

Pengelompokan Teknik Data Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan popular adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

Association

Digunakan untuk mengenali kelakuan dari kejadiankejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya adalah Market Basket Analysis, yaitu salah sati metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan.

Clustering

Digunakan untuk menganalisis pengelompokkan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokkan belum didefinisikan sebelum dijalankannya tool data mining. Biasanya menggunkan metode *neural network* atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool data mining.

Chapter 4

Decision Tree

Mengenal Decision Tree

Seperti diketahui bahwa manusia selalu menghadapi berbagai macam masalah di dalam kehidupannya sehari-hari. Masalah-masalah yang timbul dari berbagai macam bidang ini memiliki tingkat kesulitan dan kompleksitas yang sangat bervariasi, mulai dari masalah yang sangat sederhana dengan sedikit faktor-faktor terkait hingga masalah yang sangat rumit dengan banyak sekali faktor-faktor yang terkait, sehingga factor-faktor yang berkaitan dengan masalah tersebut perlu untuk diperhitungkan. Seiring dengan perkembangan kemajuan pola pikir manusia, manusia mulai mengembangkan sebuah sistem yang dapat membantu manusia dalam menghadapi masalah-masalah yang timbul sehingga dapat menyelesaikannya dengan mudah.

Pohon keputusan atau yang lebih dikenal dengan istilah *Decision Tree* ini merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebutdengan memperhitungkan berbagai macam faktor yang berkaitan di dalam lingkup masalah tersebut.

Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah sehingga dengan memperhitungkan faktor-faktor tersebut dapat dihasilkan penyelesaian terbaik untuk masalah tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah.

Pohon keputusan dalam analisis pemecahan masalah pengambilan keputusan merupakan pemetaan alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon keputusan juga memperlihatkan faktor-faktor kemungkinan yang dapat mempengaruhi alternative-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut.

Secara umum, pohon keputusan adalah suatu gambaran permodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah kepada solusi yang dihasilkan. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Seiring dengan perkembangannya, pohon keputusan kini telah banyak dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang prediksi. (Han, J., & Kamber, M. (2006). Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.)



Gambar 4.1 Bentuk Decision Tree Secara Umum

Algoritma c4.5

Pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Seperti yang telah dijelaskan sebelumnya, klasifikasi merupakan Suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Metode ini popular karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Banyak algoritma yang dapat digunakan untuk membangun suatu decision tree, salah satunya ialah algoritma C45.

Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C.45 menggunakan rasio perolehan (gain ratio). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropi.

Konsep Entropy

Entropy(S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropi digunakan untuk mengukur ketidakaslian S.



Gambar 4.2 Grafik Entropi

Besarnya Entropy pada ruang sampel S didefinisikan dengan:



Dimana:

- S : ruang (data) sampel yang digunakan untuk pelatihan
- : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu
- : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

- Entropi(S) = 0, jika semua contoh pada S berada dalam kelas yangsama.
- Entropi(S) = 1, jika jumlah contoh positif dan negative dalam S adalah sama.
- 0 > Entropi(S) > 1, jika jumlah contoh positif dan negative dalam S tidak sama.

Konsep Gain

Gain (S,A) merupakan Perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

Dimana:

- A : Atribut
- S:Sampel
- n : Jumlah partisis himpunan atribut A
- Si : Jumlah sampel pada pertisi ke -i
- S: Jumlah sampel dalam S
Untuk memudahkan penjelasan mengenai algoritma C4.5berikut ini disertakan contoh kasus yang dituangkan dalam Tabel 4.1:

No	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Tabel 4.1 Keputusan Bermain Tenis

Tabel 1 merupakan kasus yang akan dibuat pohon keputusan untuk menentukan main tenis atau tida. Data ini memiliki atribut-atribut yaitu, keadaan cuaca (outlook), temperatur, kelembaban (humidity) dan keadaan angin (windy).

Berikut merupakan cara membangun pohon keputusan dengan menggunakan algoritma:

- 1. Pilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut-atribut yang ada.
- 2. Buat cabang untuk masing-masing nilai
- 3. Bagi kasus dalamcabang
- 4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

NODE			JUMLAH	NO	YES	ENTROPY	GAIN
			KASUS	(S ₁)	(S ₂)		
1	TOTAL		14	4	10	0.863120569	
	OUTLOOK						0.258521037
		CLOUDY	4	0	4	0	
		RAINY	5	1	4	0.721928095	
		SUNNY	5	3	2	0.970950594	
	TEMPERATURE						0.183850925
		COOL	4	0	4	0	
		НОТ	4	2	2	1	
		MILD	6	2	4	0.918295834	
	HUMIDITY						0.370506501
		HIGH	7	4	3	0.985228136	
		NORMAL	7	0	7	0	
	WINDY						0.005977711
		FALSE	8	2	6	0.811278124	
		TRUE	6	4	2	0.918295834	

Tabel 4.2 Perhitungan Simpul 1



Dari hasil pada Tabel 4.2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah HUMIDITY yaitu sebesar 0.37. Dengan demikian HUMIDITY dapat menjadi node akar.

Ada 2 nilai atribut dari HUMIDITY yaitu HIGH dan NORMAL. Dari kedua nilai atribut tersebut, nilai atribut NORMAL sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya Yes, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut HIGH masih perlu dilakukan perhitungan lagi hingga semua kasus masuk dalam kelas seperti yang terlihat pada Gambar di sebelah kanan.

Kelebihan Pohon Keputusan

Dalam membuat keputusan dengan menggunankan pohon keputusan, metode ini memiliki kelebihan sebagai berikut:

- Daerah pengambilan keputusan lebih simpel dan spesifik.
- Eliminasi perhitungan-perhitungan tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelastertentu.
- Fleksibel untuk memilih fitur dari internal node yang berbeda. Sehingga dapat meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
- Dengan menggunakan pohon keputusan, penguji tidak perlu melakukan estimasi pada distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Karena metode ini menggunakan kcriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Kekurangan Pohon Keputusan

Pohon keputusan sangat membantu dalam pengambilan keputusan, namun pohon keputusan juga memiliki beberapa kekurangan, diantaranya:

- Kesulitan dalam mendesain pohon keputusan yang optimal.
- Hasil kualitas keputusan yang didapat sangat tergantung pada bagaimana pohon tersebut didesain. Sehingga jika pohon keputusan yang dibuat kurang optimal, maka akan berpengaruh pada kualitas dari keputusan yang didapat.
- Terjadi overlap terutama ketika kelas-kelas dan criteria yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
- Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.

Decision Tree pada RapidMiner

RapidMiner sebagai software pengolah data mining menyediakan tool untuk membuat decision tree. Hal ini tentu akan memudahkan kita membuat decision tree dengan menggunakan RapidMiner dibandingkan membuat decision tree secara manual yaitu dengan melakukan perhitungan menggunakan algoritma C4.5 yang telah dijelaskan sebelumnya.

Contoh Kasus:

Keputusan Bermain Tenis

Pada contoh kali ini, kita akan membuat keputusan bermain tenis atau tidak. Untuk memudahkan dalam menggunakan RapidMiner untuk membuat decision tree, kita gunakan data sederhana yang ada pada sub bab decision tree. Pertama-tama data pada tabel 2 dibuat lagi dalam format excel seperti yang terlihat pada Gambar 4.3.

	А	В	С	D	E	F
1	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY	
2	Sunny	Hot	High	No	Don't Play	
3	Sunny	Hot	High	Yes	Don't Play	
4	Cloudy	Hot	High	No	Play	
5	Rainy	Mild	High	No	Play	
6	Rainy	Cool	Normal	No	Play	
7	Rainy	Cool	Normal	Yes	Play	
8	Cloudy	Cool	Normal	Yes	Play	
9	Sunny	Mild	High	No	Don't Play	
10	Sunny	Cool	Normal	No	Play	
11	Rainy	Mild	Normal	No	Play	
12	Sunny	Mild	Normal	Yes	Play	
13	Cloudy	Mild	High	Yes	Play	
14	Cloudy	Hot	Normal	No	Play	
15	Rainy	Mild	High	Yes	Don't Play	
16						

Gambar 4.3 Tabel Keputusan dalam Format xls

Setelah data yang kita punya dibuat dalam bentuk tabel format xls, selanjutnya lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada Bab 2. Lalu cari table Microsoft Excel yang telah dibuat dan masukan kedalam Local Repository seperti yang terlihat pada Gambar 4.4.



Gambar 4.4 Lokasi Tabel pada Repository

Lakukan Drag dan Drop Tabel PlayGolf kedalam Process view. Sehingga Operator Database muncul dalam View Proses seperti pada Gambar 4.5. Pada view Process, tabel PlayGolf yang dimasukkan ke dalam proses akan dijadikan sebagai Operator Retrieve.



Gambar 4.5 Repository PlayGolf pada Main Process

Untuk membuat decision tree dengan menggunakan RapidMiner, kita membutuhkan operator Decision tree, operator ini terdapat pada View Operators. Untuk menggunakannya pilih Modelling pada View Operator, lalu pilih Classification and Regression, lalu pilih Tree Induction dan pilih Decision Tree.



Gambar 4.6 Daftar Operator pada View Operators

Setelah menemukan operator Decision Tree, seret (*drag*) operator tersebut lalu letakkan (*drop*) ke dalam view Process. Kemudian susun posisinya disamping operator Retrieve, seperti yang tampak pada Gambar 4.7.



Gambar 4.7 Posisi Operator Decision Tree

Selanjutnya, hubungkan operator Retrieve dengan operator Decision Tree dengan menarik garis dari tabel PlayGolf ke operator Decision Tree dan menarik garis lagi dari operator Decision Tree ke result di sisi kanan, seperti yang tampak pada Gambar 4.8. Operator Decision Tree berguna untuk memperdiksikan keputusan dari atribut-aribut yang dimasukkan ke dalam operator retrieve. Dengan mengubah tabel (atribut) yang dimasukkan menjadi sebuah pohon keputusan.



Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator Decision Tree

Pada operator Decision tree terdapat *input* **training set (tra)**, port ini merupakan output dari operator retrieve. Output dari operator lain juga dapat digunakan oleh port ini. Port ini menghasilkan ExampleSet yang dapat diperoses menjadi decision tree. Selain itu pada operator ini juga terdapat output model (mod) dan example set (exa). **Mod** akan mengonversi atribut yang dimasukkan menjadi mpdel keputusan dalam bentuk decision tree. **exa** merupakan port yang menghasilkan output tanpa mengubah inputan yang masuk melalui port ini. Port ini biasa digunakan untuk menggunakan kembali sama ExampleSet di operator lebih lanjut atau untuk melihat ExampleSet dalam Hasil Workspace.

Langkah selanjutnya ialah mengatur parameter sesuai dengan kebutuhan kita. Setelah menghubungkan operator retrieve dengan operator decision tree, atur parameter decision tree seperti pada gambar 4.9.

🛛 🛃 Parameters 🛛	Context 🔀			
🚨 🗟 🖻 🦻	🕺 🖶 -			
💡 Dec	cision Tree			
criterion	gain_ratio 🔹			
minimal size for split 2				
minimal leaf size	1			
minimal gain	0.1			
maximal depth	20			
confidence	0.25			
number of preprunin	.3			
no pre pruning				
no pruning	Y			

Gambar 4.9 Parameter Decision Tree



Gambar 4.10 Tipe Criterion

- Criterion, berguna memilih kriteria untuk menetapkan atribut sebagai akar dari decision tree. kriteria yang dapat dipilih, antara lain
 - 1. **Gain ratio** merupakan varian dari information_gain. Metode ini menghasilkan information gain untuk

setiap atribut yang memberikan nilai atribut yang seragam

- 2. Information_gain, dengan metode ini, semua entropi dihitung. Kemudian atribut dengan entropi minimum yang dipilih untuk dilakukan perpecahan pohon (split). Metode ini memiliki bias dalammemilih atribut dengan sejumlah besar nilai.
- Gini_index merupakan ukuran ketidakaslian dari suatu ExampleSet. Metode ini memisahkan pada atribut yang dipilih memberikan penurunan indeks gini rata-rata yang dihasilkan subset.
- Accuracy, metode ini memimilih beberapa atribut untuk memecah pohon (split) yang memaksimalkan akurasi dari keseluruhan pohon.
- Minimal size of split, Ukuran untuk membuat simpul-simpul pada decision tree. simpul dibagi berdasarkan ukuran yang lebih besar dari atau sama dengan parameter Minimal size of split. Ukuran simpul adalah jumlah contoh dalam subset nya

- Minimal leaf size, Pohon yang dihasilkan sedemikian rupa memiliki himpunan bagian simpul daun setidaknya sebanyak jumlah minimal leaf size.
- Minimal gain merupakan nilai gain minimal yang ditentukan untuk menghasilkan simpul pohon keputusan. Gain dari sebuah node dihitung sebelum dilakukan pemecahan. Node dipecah jika gain bernilai lebih besar dari Minimal Gain yang ditentukan. Nilai minimal gain yang terlalu tinggi akan mengurangi perpaecahan pohon dan menghasilkan pohon yang kecil. Sebuah nilai yang terlalu tinggi dapat mencegah pemecahan dan menghasilkan pohon dengan simpul tunggal.
- Maximal depth, Parameter ini digunakan untuk ٠ membatasi ukuran Putusan Pohon Proses pohon tidak berlanjut generasi ketika kedalaman pohon adalah sama dengan kedalaman maksimal. Jika nilainya diatur ke '-1', parameter kedalaman maksimal menempatkan tidak terikat pada kedalaman pohon, pohon kedalaman maksimum dihasilkan. Jika nilainya diatur ke '1 ' maka akan dihasilkan pohon dengan simpul tunggal.

- Confidence, Parameter ini menentukan tingkat kepercayaan yang digunakan untuk pesimis kesalahan perhitungan pemangkasan.
- number of prepruning alternatives. Parameter ini menyesuaikan jumlah node alternatif mencoba untuk membelah ketika split dicegah dengan prepruning pada simpul tertentu.
 - no prepruning, Secara default Pohon Keputusan yang dihasilkan dengan prepruning. Menetapkan parameter ini untuk menonaktifkan benar prepruning dan memberikan pohon tanpa prepruning apapun.
 - no pruning Secara default Pohon Keputusan yang dihasilkan dengan pemangkasan. Menetapkan parameter ini untuk menonaktifkan benar pemangkasan dan memberikan sebuah unpruned

Setelah parameter diatur, klik ikon Run pada toolbar, seperti pada gambar 40 untuk menampilkan hasilnya. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 4.11 Ikon Run

Setelah beberapa detik maka RapidMiner akan menampilkan hasil keputusan pada view Result. Jika kita pilih Graph view, maka akan ditampilkan hasilnya berbentuk pohon keputusan seperti pada gambar 4.12. Hasil pohon keputusan dapat disimpan dengan mengklik save image pada sisi kiri View Result.



Gambar 4.12 Hasil Berupa Graph Pohon Keputusan

Selain menampilkan hasil decision tree berupa graph atau tampilan pohon keputusan, RapidMiner juga menyediakan tool untuk menampilkan hasil berupa teks view dengan mengklik button Text View seperti yang tampak pada Gambar 4.13.



Gambar 4.13 Hasil Berupa Penjelasan Teks

Contoh Kasus : Keputusan seseorang mempunyai potensi menderita hipertensi

Sebelumnya kita telah mengetahui bagaimana membuat pohon keputusan untuk menentukan bermain tenis dengan menggunakan operator decision tree. Pada pembahasan kali ini kita akan membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi atau tidak. Untuk menambah pengeatahuan kita mengenai kegunaan operator yang ada pada RapidMiner, oleh karena itu untuk membuat pohon keputusan kali ini kita menggunakan operator X-Validation, Apply Model dan Performance. Selain itu, kita juga tetap menggunakan operator decision tree dalam pembuatan pohon keputusan kali ini.

Cara yang digunakan dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi, tidak jauh berbeda dengan cara membuat pohon keputusan yang sebelumnya, yaitu pertama-tama import data ke dalam repository RapidMiner, lalu lakukan drag dan drop data tersebut pada view process untuk mengubah data yang berisi atribut pohon keputusan menjadi operator retrieve. setelah itu, lakukan drag dan drop operator decision tree ke dalam view process dengan cara yang sama seperti penjelasan sebelumnya.

	А	В	С	D	E
1	Usia	Berat	Kelamin	Hipertensi	
2	Muda	Overweight	Pria	Ya	
3	Muda	Underweight	Pria	Tidak	
4	Muda	Average	Wanita	Tidak	
5	Tua	Overweight	Pria	Tidak	
6	Tua	Overweight	Pria	Ya	
7	Muda	Underweight	Pria	Tidak	
8	Tua	Overweight	Wanita	Ya	
9	Tua	Average	Pria	Tidak	
10					

Gambar 4.14 Tabel SakitHipertensi dalam format xls



Gambar 4.15 Lokasi Tabel pada Repository



Gambar 4.16 Tabel SakitHipertensi pada Main Process

Untuk membuat pohon keputsan kali ini kita menggunakan operator X-Validation. Operator ini melakukan validasi silang untuk memperkirakan kinerja statistik operator pembelajaran (biasanya pada set data yang tak terlihat). Operator ini juga digunakan untuk memperkirakan seberapa akurat suatu model yang akan tampil dalam praktek. Operator X-Validasi merupakan operator bersarang yang memiliki dua subproses: *training subprocess* (subproses percobaan) dan *testing subprocess* (subproses pengujian). Subproses percobaan digunakan untuk melatih sebuah model. Model yang terlatih kemudian diterapkan dalam subproses pengujian.

Biasanya proses belajar mengoptimalkan parameter model untuk membuat model sesuai dengan data percobaan. Jika kita kemudian mengambil sampel independen dari data pengujian, umumnya model tersebut tidak cocok dengan data percobaan maupun data pengujian. Hal ini disebut dengan istilah 'over-pas', dan sangat mungkin terjadi ketika ukuran set data training kecil, atau ketika jumlah parameter dalam model besar. Sehingga validasi silang merupakan cara untuk memprediksi kesesuaian model untuk satu set pengujian hipotesis ketika set pengujian eksplisit tidak tersedia.

Untuk menemukan operator X-Validation, pilih Evaluation pada View Operator, lalu pilih Validation, lalu pilih X-Validation .Setelah menemukan operator X-Validation, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.17 Hirarki Operator X-Validation



Gambar 4.18 Operator Validation

Operator X-Validation memiliki port input yaitu, training example set (tra) sebagai port input memperkirakan ExampleSet untuk melatih sebuah model (training data set). ExampleSet yang sama akan digunakan selama subproses pengujian untuk menguji model.

Selain itu, operator ini juga memiliki port output sebagai berikut:

- **model (mod),** Pelatihan subprocess harus mengembalikan sebuah model yang dilatih pada input ExampleSet. Harap dicatat bahwa model yang dibangun ExampleSet disampaikan melalui port ini.
- training example set (tra), The ExampleSet yang diberikan sebagai masukan pada port input pelatihan dilewatkan tanpa mengubah ke output melalui port ini. Port ini biasa digunakan untuk menggunakan kembali ExampleSet sama di operator lebih lanjut atau untuk melihat ExampleSet dalam Workspace Result.
- averagable (ave), subproses pengujian harus mengembalikan Vector Kinerja. Hal ini biasanya dihasilkan dengan menerapkan model dan mengukur kinerjanya. Dua port tersebut diberikan tetapi hanya dapat digunakan jika diperlukan. Harap dicatat bahwa kinerja statistik dihitung dengan skema estimasi hanya perkiraan (bukan perhitungan yang tepat) dari kinerja yang akan dicapai dengan model yang dibangun pada set data yang disampaikan secara lengkap.

🖉 📴 Parameters 🕅	Context 🗶
2	t 📭 🕶
% Validation	n (X-Validation)
🖌 average performar	nces only
leave one out	
number of validations	10
sampling type	stratified sampling 💌
use local random s	eed
📀 Compatibi <u>l</u> ity level (5.3.000

Gambar 4.19 Parameter X-Validation

Operator X-Validation juga memiliki parameter yang perlu diatur, diantaranya:

- average performances only (boolean), ini merupakan parameter ahli yang menunjukkan jika vector kinerja harus dirata-ratakan atau semua jenis dari hasil rata-rata.
- leave one out (boolean) Seperti namanya, leave one out validasi silang melibatkan penggunaan satu contoh dari ExampleSet asli sebagai data pengujian (dalam pengujian subproses), dan contoh-contoh yang tersisa sebagai data pelatihan (dalam pelatihan subprocess). Namun hal ini biasanya sangat mahal untuk ExampleSets besar dari sudut

pandang komputasi karena proses pelatihan diulang sejumlah besar kali (jumlah waktu contoh). Jika diatur dengan benar, parameter number of validations dapat diabaikan.

- number of validations (integer), parameter ini menentukan jumlah subset ExampleSet yang harus dibagi (setiap subset memiliki jumlah yang sama dari contoh). Juga jumlah yang sama dari iterasi yang akan berlangsung. Setiap iterasi melibatkan pelatihan model dan pengujian model. Jika ini ditetapkan sama dengan jumlah contoh dalam ExampleSet, Hal ini akan setara dengan operator X-Validasi dengan parameter leave one out settrue.
- **sampling type** (selection), Operator X-Validasi dapat menggunakan beberapa jenis sampling untuk membangun subset. Sampel yang tersedia, diantaranya:
 - 1. linear_sampling, Linear sampling hanya membagi ExampleSet ke partisi tanpa mengubah urutan contoh yaitu subset dengan contoh-contoh berturut-turut diciptakan.
 - 2. shuffled_sampling, Shuffled Sampling membangun subset acak ExampleSet. Contoh dipilih secara acak untuk membuat subset.
 - 3. stratified_sampling, Stratified Sampling membangun subset acak dan memastikan

bahwa distribusi kelas dalam himpunan adalah sama seperti dalam ExampleSet seluruh.

 use local random seed (boolean), Parameter ini menunjukkan jika local random seed harus digunakan untuk mengacak contoh subset. Dengan menggunakan nilai yang sama dengan local random seed maka akan menghasilkan subset yang sama. Mengubah nilai parameter ini mengubah cara contoh menjadi acak, sehingga subset akan memiliki satu set yang berbeda dari contoh. Parameter ini hanya tersedia jika Shuffled atau Stratified sampling dipilih. Hal ini tidak tersedia untuk pengambilan sampel Linear karena tidak membutuhkan pengacakan, contoh yang dipilih secara berurutan

Iocal random seed (integer), Parameter ini hanya tersedia jika parameteruse local random seed dipilih. parameter ini menentukan local random seed

Seperti yang telah disebutkan sebelumnya bahwa dalam membuat pohon keputusan pada contoh ini, kita menggunakan operator Apply Model. Operator ini menerapkan suatu model terlatih pada sebuah ExampleSet. Sebuah model pertama kali dilatih di sebuah ExampleSet, informasi yang berkaitan dengan ExampleSet dipelajari oleh model. Maka model tersebut dapat diterapkan pada ExampleSet yang lain dan biasanya untuk prediksi. Semua parameter yang diperlukan disimpan dalam objek model. Ini adalah wajib bahwa kedua ExampleSets harus persis nomor yang sama, order, jenis dan peran atribut. Jika sifat meta data dari ExampleSets tidak konsisten, hal itu dapat menyebabkan kesalahan serius.

Untuk menemukan operator Apply Model, pilih Modeling pada View Operator, lalu pilih Model Application, lalu pilih Confidence dan pilih Apply Model .Setelah menemukan operator Apply Model, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.20 Hirarki Operator Apply

Operator ini memiliki port input yaitu, **model** (mod) port ini mengharapkan model. Port ini harus memastikan bahwa nomor, order, jenis dan peran atribut dari ExampleSet pada model yang dilatih konsisten dengan ExampleSet pada port input data unlabeled. **unlabelled data (unl)** port ini mengharapkan suatu ExampleSet. Ini harus memastikan bahwa nomor, order, jenis dan peran atribut ExampleSet ini konsisten dengan ExampleSet pada model yang dikirim ke port input model dilatih.

Operator ini juga memiliki port output, diantaranya, **labeled Data (lab)**, Model yang diberikan dalam input diterapkan pada ExampleSet yang diberikan dan ExampleSet terbaru disampaikan dari port ini. Beberapa informasi akan ditambahkan ke input ExampleSet sebelum dikirimkan melalui port output. Dan **model (mod)**, Model yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui port ini.



Gambar 4.21 Operator Apply Model

Seperti yang terlihat pada gambar 4.22, Operator Apply Model hanya memiliki dua parameter yaitu, **application parameters** (menu) parameter ini merupakan parameter ahli yang berguna memodelkan parameter untuk aplikasi (biasanya tidak diperlukan). Dan **create view** (boolean) Jika model diterapkan pada port input mendukung Views, Hal ini mungkin untuk membuat View bukannya mengubah data yang mendasarinya. Transformasi yang akan biasanya dilakukan langsung di data kemudian akan dihitung setiap kali nilai diminta dan hasilnya dikembalikan tanpa mengubah data. Beberapa model tidak mendukung Views.



Gambar 4.22 Parameter Apply Model

Dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit Hipertensi, kita juga menggunakan operator Performance. Operator ini digunakan untuk evaluasi kinerja. Operator ini memberikan daftar nilai kriteria kinerja. Kriteria kinerja secara otomatis ditentukan agar sesuai dengan jenis tugas belajar. Berbeda dengan operator lain, operator ini dapat digunakan untuk semua jenis tugas belajar. Secara otomatis menentukan jenis tugas belajar dan menghitung kriteria yang paling umum untuk jenistersebut.

Untuk menemukan operator Performance, pilih Evaluation pada View Operator, lalu pilih Performance and Measurement, lalu pilih Performance. Setelah menemukan operator Performance, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.23 Hirarki Operator Performance

Operator Performance memiliki port input yaitu, **labelled data (lab),** Port ini mengharapkan mengharapkan ExampleSet berlabel. Apply Model merupakan contoh yang baik dari operator yang menyediakan data berlabel. Pastikan bahwa ExampleSet memiliki atribut label dan atribut prediksi. **performance (per)** Ini adalah parameter opsional yang membutuhkan Performance Vector.

Selain itu, Operator ini juga memiliki port output yaitu, **performance (per)**, port ini memberikan Peformance Vector (kita menyebutnya outputperformance-vektor untuk saat ini). Peformance Vector adalah daftar nilai kinerja kriteria. **example set (exa)**, ExampleSet yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui portini.



Gambar 4.24 Operator Performance

Operator ini hanya memiliki satu parameter yaitu, **use example weights** (boolean) Parameter ini memungkinkan contoh bobot contoh yang akan digunakan untuk perhitungan kinerja jika memungkinkan. Parameter ini memiliki tidak memiliki efek jika atribut tidak memiliki peran bobot.



Gambar 4.25 Parameter Performance

Selanjutnya, susun dan hubungkan port-port dari operator decision tree, operator Apply Model dan operator Performance seperti yang terlihat pada Gambar 55.



Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Performance

Kemudian hubungkan operator retrieve (tabel SakitHipertensi) dengan operator validation dengan menarikgarispadaportinput dan output yang terdapat pada operator tersebut, seperti yang tampak pada Gambar 56.



Gambar 4.27 Susunan Operator Retrieve dengan Operator Validation

Setelah parameter dari masing-masing operator diatur, dan posisi operator disusun dengan benar, klik Run, lalu tunggu beberapa detik hingga RapidMiner akan menampilkan hasil Keputusan decision tree berupa graph pohon. seperti yang tampak pada Gambar 4.28.



Gambar 4.28 Tampilan Decision Tree

Chapter 5

Neural Network

Apa itu Neural Network?

Dapat dikatakan bahwa neural netwok dapat mempelajari pemetaan input data ke output data. Neural network merupakan model komputasi yang terinspirasi oleh prinsip-prinsip mengenai bagaimana cara otak manusia bekerja. Mereka dapat mempelajarinya dari data, mereka mampu mengeneralisasi dengan baik, dan mereka tahan dengan kebisingan.

Biasanya jaringan saraf digunakan untuk masalah-masalah seperti klasifikasi (classification), prediksi (prediction), pengenalan pola (pattern recognition), pendekatan (approximation), dan asosiasi (association). Mereka hanya perlu belajar dari beberapa data sampel, dan setelah mereka telah mempelajarinya, mereka dapat bekerja dengan input data yang tidak diketahui, atau bahkan input data yang bising maupun tidak lengkap.

Secara umum Neural Network (NN) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia. NN ini merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut.

Secara sederhana NN adalah sebuah alat pemodelan data statistik non-linear. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data. Secara mendasar, sistem pembelajaran merupakan proses penambahan pengetahuan pada NN yang sifatnya kontinuitas sehingga pada saat digunakan pengetahuan tersebut akan dieksploitasikan secara maksimal dalam mengenali suatu objek. Neuron adalah bagian dasar dari pemrosesan suatu Neural Network. Dibawah ini merupakan bentuk dasar dari suatu neuron.

Bentuk Neural Network

Setiap neural network terdiri dari unit pengolahan dasar yang saling berhubungan, yang disebut Neuron. Network belajar dengan memodifikasi bobot hubungan antara neuron selama proses pelatihan. Bentuk dasar arsitektur suatu Neural Network adalah sebagai berikut:



Gambar 5.1 Arsitektur Dasar Neural Network

Secara umum, terdapat tiga jenis Neural Network yang sering digunakanberdasarkan jenis network-nya, yaitu:

- 1. Single-Layer Neural Network
- 2. Multilayer Perceptron Neural Network
- 3. Recurrent Neural Networks

Single-Layer Neural Network

Neural Network jenis ini memiliki koneksi pada inputnya secara langsung ke jaringan output.



Gambar 5.2 Single-layer Neural Network

Jenis Neural Network ini sangatlah terbatas, hanya digunakan pada kasus-kasus yang sederhana.

Multilayer Perceptron Neural Network

Jenis Neural Network ini memiliki layer yang diミaマakaミ さhiddenざ, diteミgah layer input dan output. Hidden ini bersifat variable, dapat digunakan lebih dari satu hidden layer.


Gambar 5.3 Multilayer Perceptron Neural Network

Gambar di atas menunjukkan sebuah jaringan saraf sederhana yang dibuat dengan easyNeurons. Jenis jaringan ini disebut Multi Layer Perception dan itu merupakan salah satu jaringan yang paling umum digunakan.

Recurrent Neural Network

Neural network jenis ini memiliki ciri, yaitu adanya koneksi umpan balik dari output ke input.



Gambar 5.4 Recurrent Network

Kelemahan dari jenis ini adalah Time Delay akibat proses umpan balik dari output ke titik input.

Proses Pembelajaran pada Neural Network

Proses pembelajaran merupakan suatu metoda untuk proses pengenalan suatu objek yang sifatnya kontinuitas yang selalu direspon secara berbeda dari setiap proses pembelajaran tersebut. Tujuan dari pembelajaran ini sebenarnya untuk memperkecil tingkat suatu error dalam pengenalan suatu objek.

Secara mendasar, neural network memiliki sistem pembelajaran yang terdiri atas beberapa jenis berikut:

- 1. Supervised Learning
- 2. Unsupervised Learning

Supervised Learning

Sistem pembelajaran pada metoda Supervised learning adalah system pembelajaran yang mana, setiap pengetahuan yang akan diberikan kepada sistem, pada awalnya diberikan suatu acuan untuk memetakan suatu masukan menjadi suatu keluaran yang diinginkan. Proses pembelajaran ini akan terus dilakukan selama kondisi error atau kondisi yang diinginkan belum tercapai. Adapun setiap perolehan error akan dikalkulasikan untuk setiap pemrosesan hingga data atau nilai yang diinginkan telah tercapai.

Unsupervised Learning

Sistem pembelajaran pada neural network, yang mana sistem ini memberikan sepenuhnya pada hasil komputasi dari setiap pemrosesan, sehingga pada sistem ini tidak membutuhkan adanya acuan awal agar perolehan nilai dapat dicapai. Meskipun secara mendasar, proses ini tetap mengkalkulasikan setiap langkah pada setiap kesalahannya dengan mengkalkulasikan setiap nilai weight yang didapat.

Siapa yang menggunakan Neural Network?

Beberapa aplikasi yang khas adalah gambar (image), sidik jari dan pengenalan wajah (fingerprint and face recognition), prediksi saham (stock prediction), prediksi untuk taruhan (sport bets prediction), klasifikasi pola dan pengakuan (pattern classification and recognition), pengawasan dan pengendalian (monitoring and control). Mereka digunakan dalam industri, kedokteran (diagnosa), aplikasi militer (seperti radar pada pengenalan citra), keuangan dan robotika. Akhir-akhir ini mereka sangat popular di industri game karena berkat mekanisme belajar yang dilakukan, mereka dapat memberikan kontrol adaptif dan pembelajaran untuk karakter yang dikendalikan computer.

Kegunaan Neural Networks

- 1. Pengenalan karakter optikal (Optical character recognition)
- 2. Pengenalan citra (Image recognition)
- 3. Pengenalan sidik jari (Fingerprint recognition)
- 4. Prediksi saham (Stock prediction)
- 5. Prediksi taruhan (Sport bets prediction)
- 6. Kontrol computer untuk karakter game (Computer controlled game characters)
- 7. Model statistical (Statistical modeling)
- 8. Data mining

Neural Network pada RapidMiner

Kita mulai dengan menggunakan data sederhana dalam tabel GE.xls. Data tersebut juga bisa kita dapatkan dengan melakukan pengunduhan melalui salah satu *add-ins* Microsoft Excel yang bernama *DownloaderXL*, dimana data mengenai harga saham yang terjadi dalam rentang waktu tertentu telah dicatat pada sebuah *web hosting*.

Contoh Kasus:

Perkiraan harga saham dengan menggunakan metoda Neural Network.

N N	1icrosoft Exc	el - GE						
:1	Eile Edit	⊻iew Inser	t Format	Tools Da	ta <u>W</u> indow	Monte Carlo	o <u>xi</u> bit <u>H</u> elp Ado <u>b</u> e PDF	-
2.04		alar	A ABC 454	LV D. R		A		0.092
-				00 400 4		2 . (- 21)	😂 Z * Z + A + 🛄 🖓 ·	100 %
A	rial	- 10	- B 1	Ū ≣⊧	콜 콜 🛃	\$ % ,	108 208 📰 💷 - 🗸	» • <u>A</u> •
An	alyzerXL 🕶 🛛 By	IkQuotesXL 🔻	Downloade	rXL - RTQU	iotesXL • Pr	edictorXL •	BacktestingXL • ClassifierXL •	PortfolioXL
	G18	• fi	55.17541	86146191				
	A	B	С	D	E	F	G	Н
1	Date	Open	High	Low	Close	Volume	StochasticOscillator(8,5)	
2	05/08/2009	14.18	14.35	13.95	14.33	129813028	30.00	
3	05/11/2009	14,18	14.22	13.90	13.99	94827338	87.69	
4	05/12/2009	14.04	14.11	13.08	13.49	119778085	83.76	
5	05/13/2009	13.03	13.07	12.57	12.73	108138237	69.63	
6	05/14/2009	12.59	13.00	12.33	12.86	91507642	57.62	
7	05/15/2009	12.89	13.07	12.52	12.68	84562005	39.26	
8	05/18/2009	12.93	13.31	12.90	13.28	77949206	32.04	
9	05/19/2009	13.45	13.77	13.30	13.51	93399522	32.21	
10	05/20/2009	13.84	14.18	13.55	13.58	115136364	42.82	
11	05/21/2009	13.32	13.34	12.86	13.05	106967634	45.49	
12	05/22/2009	13.18	13.22	12.84	12.92	54782537	48.77	
13	05/26/2009	12.84	13.42	12.67	13.20	97279770	48.79	
14	05/27/2009	13.28	13.31	12.68	12.81	108987013	40.92	
15	05/28/2009	13.05	13.10	12.64	13.01	83621254	32.51	
16	05/29/2009	13.14	13.34	12.05	13.29	86101902	37.25	
17	06/01/2009	13.63	13.80	13.32	13.67	87599560	47.17	
18	06/02/2009	13.51	13.66	13.36	13.61	67039102	55.18	
19	06/03/2009	13.46	13.51	13.11	13.31	69318595	65.10	

Gambar 5.5 Tabel GE.xls dalam Microsoft Excel

Buatlah *file* baru pada Microsoft Excel berdasarkan tabel harga saham. Berikan nama Header: Date, Open,

High, Low, Close, Volume, Stochastic Oscilator. Isilah sel seperti gambar [berapa]. Simpan dengan nama GE.xls

Lakukan pemilihan *repository* GE_TEMP_XLSDATA dengan melakukan *drag and drop* yang ditempatkan pada *panel main process* seperti gambar 5.6.



Gambar 5.6 Import Repository

Lakukan pemilihan operator *Neural Network* seperti gambar 5.7. Kemudian *drag and drop ke Main Process* seperti sebelumnya



Gambar 5.7 Operator Neural network

Lakukan pembuatan hubungan antara *repository* dan *operator*, kemudian antara *operator* dengan hasil *output*.



Gambar 5.8 Menghubungkan Seluruh Operator ke Result

klik ikon Play **D**. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 5.9 Ouput Neural Network

Gambar 5.9 merupakan grafik berbentuk *node* yang saling terhubung seperti layaknya sebuah jaringan syaraf dari hasil rules yang telah kita dapatkan

Chapter 6

Market Basket Analysis

Memahami Market Basket Analysis

Retail atau Eceran salah satu cara pemasaran produk meliputi semua aktivitas yang melibatkan penjualan barang secara langsung ke konsumen akhir, konsumen akhir membeli kumpulan produk dengan jumlah yang berbeda di waktu yang berbeda. Namun penjualan secara ritel hari ini bukanlah apa-apa jika insdustrinya tidak mampu berkompetisi dengan baik. Lanskap yang kompleks dan cepat berubah, persaingan yang ketat, dan pelanggan yang semakin menuntut mendorong *retailer* harus memikirkan kembali bagaimana mereka beroperasi. Kemampuan untuk memahami pola pikir konsumen adalah hal yang sangat penting bagi *retailer*.

Teknologi telah membantu *retailer* dengan memungkinkan untuk menyimpan data konsumen dengan volume yang sangat besar dan biaya yang sangat wajar. *Retailer* kini dapat memiliki miliyaran informasi tentang informasi pelanggan mereka. Informasi ini dapat menjawab pertanyaan-pertanyaan penting termasuk: Kapan pelanggan akan membeli? Bagaimana pembayaran dilakukan? Berapa banyak dan apa item tertentu yang dibeli? Apa hubungan antara barang yang dibeli?

Tidak ada keraguan bahwa data point-of-sales (POS) ini yang (ketika digunakan secara efektif) diberdayakan pengecer untuk lebih memahami bisnis mereka dan meningkatkan pengambilan keputusan. Pengecer proaktif menggunakan informasi ini untuk memberikan penawaran yang ditargetkan yang sesuai dengan harapan konsumen dan kemudian memberikan dampak penghasilan positif. Namun pada dasarnya, bagaimanakan *retailer* menggunakan miliyaran informasi ini? Jawabannya adalah menghubungkan produk-produk yang ada.

Sering kali, sebagai konsumen, kita cenderung mengabaikan bagaimana barang secara fisik diatur dalam sebuah toko *retail* atau supermarket. Apa yang mungkin terlihat (bagi kita) hanyalah seperti sebuah 'distribusi acak', namun sebenarnya hal tersebut merupakan pengaturan barang yang direncanakan secara cermat. Pada intinya, toko *retail* menilai pola pembelian pelanggan dan mengatur produk-produk yang akan dibeli secara sesuai. Sehingga menyebabkan pelanggan melakukan kegiatan pembelian beberapa produk sekaligus tanpa disadarinya.

Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan inilah yang dikenal sebagai *Market Basket Analysis* (MBA). Seperti namanya, *Market Basket Analysis* pada dasarnya melibatkan penggunaan data transaksional konsumen untuk mempelajari pola pembelian dan menjelajahi kemungkinan (probabilitas dan) *crossselling*. Tujuan dari MBA adalah untuk memanfaatkan data penjualan efektif untuk meningkatkan taktik pemasaran dan penjualan di tingkat toko. Contoh yang paling umum dari Market Basket Ašalysis adalah & Beer daš Diapers & Coštoh iši merupakan kasus dari salah satu toko retail besar yang ada di US, Wal-Mart. Seorang manajer toko menemukan hubungan yang kuat antara salah satu merek popok bayi (diapers) dan salah satu merek beer pada beberapa pembeli. Analisa pembelian mengungkapkan bahwa kegiatan pembelian dilakukan oleh laki-laki dewasa pada hari jumat malam terutama sekitar jam enam dan tujuh sore. Setelah beberapa observasi, supermarket mengetahui bahwa:

- Karena bungkus dari popok bayi sangat besar, para istri, dimana dalam banyak kasus adalah seorang ibu rumah tangga, akan menyuruh suaminya untuk membelinya.
- Pada akhir dari minggu, para suami dan ayah akan menghabiskan minggunya dengan membeli beberapa beer.

Jadi, apa yang akan dilakukan supermarket dari pengetahuan ini?

- Mereka menempatkan *premium beer* tepat disebelah *diapers*
- Hasilnya adalah para ayah akan membeli diapers dan yang biasanya membeli beer biasa sekarang

membeli *premium beer* seperti yang sudah diperkirakan.

 Secara signifikan, para pria yang biasanya tidak membeli bir sebelum mulai berbelanja akan membelinya karena itu begitu mudah dilihat dan diambil - hanya sebelah popok (cross-sell)

Istilah Market Basket Analysis sendiri datang dari kejadian yang sudah sangat umum terjadi di dalam pasar swalayan, yakni ketika para konsumen memasukkan semua barang yang merak beli ke dalam keranjang (basket) yang umumnya telah disediakan oleh pihak swalayan itu sendiri. Informasi mengenai produk-produk yang biasanya dibeli secara bersamasama oleh para konsumen dapat memberikan č wawasa: č terse:diri Hagi para pe:gelola toko atau swalayan untuk menaikkan laba bisnisnya (Albion Research, 2007).

Metodologi Association Rules

Metodologi Association Rules, atau Analisis Asosiasi adalah sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar. Salah satu penerapan Metode Association rules adalah pada Market Basket Analysis. Association rule adalah sebuah ekspresi implikasi dari bentuk X \square Y, dimana X dan Y adalah itemset yang saling terpisah (disjoint), dengan kata lain X \bigcap Y = \varnothing Dalam menentukan Association Rule, terdapat suatu interestingness measure (ukuran ketertarikan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Ada dua ukuran yaitu:

1. Support: Bagian transaksi yang mengandung kedua X dan Y.

岫岻

Ataujika terdapat dua buah item dalam X, nilai support diperoleh dari rumus berikut:

岫岻

2. Confidence: Seberapa sering item dalam Y muncul ditransaksi yang mengandung X.

岫|岻

Kudua ukuran ini nantinya berguna dalam menentukan interesting association rules, yaitu untuk dibandungkan dengan batasan (threshold) yang ditentukan oleh user. Batasan tersebut umumnya bernama *minimum support* dan *minimum confidence*.

Mengapa menggunakan Support dan Confidence? Support adalah ukuran yang penting karena jika aturan memiliki support yang kecil, maka kejadian bisa saja hanyalah sebuah kebetulan. Aturan Support yang rendah juga cenderung tidak menarik dari perspektif karena mungkin bisnis tidak akan memberikan keuntungan saat mempromosikan barangbarang yang jarang dibeli pelanggan bersamaan. Untuk alasan ini, dukungan sering digunakan untuk menghilangkan ketidak-menarikan ini. Confidence. adalah ukuran kehandalan dari kesimpulan yang dibuat oleh aturan. Semakin besar Confidence, semakin besar kemungkinan untuk Y hadir dalam transaksi yang mengandung X. Confidence juga memberikan probabilitas bersyarat dari Y yang diberikan ke X.

Contoh Association Rules

Untuk lebih memahami Association Rules, mari kita telusuri contoh berikut. Sebuah toko retail telah melakukan transaksi dengan pembeli seperti yang tertulis pada tabel.

Kode Transaksi	Produk yang terjual
001	Pena, Roti, Mentega
002	Roti, Mentega, Telur
003	Buncis, Telur, Susu

Tabel	6.1	Tabel	Transaksi
-------	-----	-------	-----------

004	Roti, Mentega
005	Roti, Mentega, Kecap, Telur, Susu

Tahap pertama adalah mencari nilai dari Support sesuai dengan rumus yang telah disebutkan sebelumnya. Misalnya, Untuk transaksi yang memuat {roti, mentega} ada 4, maka nilai supportnya adalah 80%. Lalu jumlah transaksi yang memuat {Roti, Mentega, Susu} ada 2, maka nilai supportnya adalah 40%. Sedangkan transaksi yang memuat {buncis} hanya 1, maka nilai supportnya adalah 20%. Jika kita tentukan bahwa *minimum support*nya adalah 30%, maka rule yang memenuhi adalah sebagai berikut:

Kombinasi Produk	Nilai Support
{roti}	80%
{mentega}	80%
{telur}	60%
{susu}	60%
{roti, mentega}	80%
{mentega, telur, susu}	40%
{roti, mentega, telur, susu}	40%

Tabel 6.2 Kombinasi Produk dan Nilai Support

Setelah semua pola kombinasi dan nilai dari Supportnya ditemukan, barulah dicari *Association Rules* yang memenuhi syarat minimum untuk confidence. Bila ditentukan syarat minimum untuk confidence sebesar 50% maka Association Rules yang dapat dipakai adalah:

Association Rules	Support	Confidence
{roti} 🛛 {mentega}	80%	100%
{roti} 🛛 {telur}	40%	50%
{roti} ₪ {susu}	40%	50%
{roti} 🛛 {mentega, telur}	40%	50%
{roti} 🛛 {mentega, susu}	40%	50%
{roti} 🛛 {telur, susu}	40%	50%
{roti} ಔ{mentega, telur, susu }	40%	50%
{mentega, telur} 🛛 {roti}	40%	100%
{roti, mentega, susu} ₪{telur}	40%	100%
{roti, telur, susu} ₪{mentega}	40%	100%
{mentega, telur, susu} ₪{roti}	40%	100%

Tabel 6.3 Association Rules dan Nilai Confidence

Assosiation Rule akan dipilih sesuai kebijakan manajer toko, semakin tinggi support dan confidence semakin baik hasilnya. Misalkan kita ambil contoh yaitu {mentega, telur} 2 {roti} yang memiliki nilai Support

8[‡]% da≷ Co≷fide≷Ie ヹ[‡][‡]%, arti≷ya adalah: č ^eeora≷g konsumen yang membeli mentega dan telur memiliki kemungkinan 100% untuk juga membeli roti. Aturan ini cukup signifikan karena mewakili 40% dari catatan selaરa iરi.ર્ટ

Frequent Itemset Generation dan Rule Generation

Frequent Itemset Generation

Tujuannya adalah untuk menemukan semua itemset yang memenuhi *minimum support*. Item set ini sering disebut dengan frequent. Namun Masalah utama pencarian Frequent Itemset adalah banyaknya jumlah kombinasi itemset yang harus diperiksa apakah memenuhi minimum support atau tidak. Salah satu cara untuk mengatasinya adalah dengan mengurangi jumlah kandidat itemset yang harus diperiksa.

Apriori adalah salah satu pendekatan yang sering digunakan pada Frequent Itemset Mining. Prinsip Apriori adalah jika sebuah itemset infrequent, maka itemset yang infrequent tidak perlu lagi diexplore supersetnya sehingga jumlah kandidat yang harus diperiksa menjadi berkurang. Kira kira ilustrasinya seperti ini:



Gambar 6.1 Frequent Item Set tanpa Apriori

Pada gambar 36, pencarian Frequent Itemset dilakukan tanpa menggunakan prinsip Apriori. Dengan menggunakan prinsip Apriori, pencarian Frequent Itemset akan menjadi seperti di bawah ini:



Gambar 6.2 Frequent Item Set dengan Apriori

Dapat dilihat bahwa dengan menggunakan Apriori, jumlah kandidat yang harus diperiksa cukup banyak berkurang.

Rule Generation

Tujuannya adalah untuk mengekstrak semua aturan yang memiliki high-confidence dari itemsets yang ditemukan dari langkah sebelumnya. Aturan ini disebut Strong Rules.

MarketBasketAnalysispada RapidMiner

Sekali lagi, pencarian Rule pada Association Rules merupakan sebuah proses yang luar biasa panjang. Manusia tidak akan mampu untuk melakukan pengitungan dengan berates-ratus data (belum kombinasi dari seluruh item yang ada). Maka dari itu, untuk mencari seluruh Rules yang ada, RapidMiner telah menyediakan tools untuk mempermudah pengguna. Untuk memahami cara menggunakan tools ini, ikuti manual berikut secara seksama.

Contoh Kasus :

Transaksi Penjualan Sederhana.

Kita mulai dengan menggunakan data sederhana yang kita miliki yang terdapat pada sub bab pengenalan Market Basket Analysis, Tabel 5.1.

	Α	В	С	D	Е	F	G	Н	-
1	TID	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP	
2	001	1	1	1	0	0	0	0	
3	002	0	1	1	1	0	0	0	
4	003	0	0	0	1	1	1	0	
5	004	0	1	1	0	0	0	0	
6	005	0	1	1	1	0	1	1	
7									
0									

Gambar 6.3 Tabel Penjualan Sederhana

Buatlah Table baru pada Microsoft Excel berdasarkan tabel 5.1. Berikan nama Header: TID (Transaction ID), PENA, ROTI, MENTEGA, TELUR, BUNCIS, SUSU, KECAP. Isilah cell seperti gambar 5.3. Simpan dengan nama TransaksiMakanan.xls.

Lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada **Bab 2**. Browse table Microsoft Excel yang telah dibuat, dan masukan kedalam Local Repository, seperti gambar disamping.



Gambar 6.4 Repositori

Lakukan Drag dan Drop Tabel TransaksiMakanan tadi kedalam Process. Sehingga Operator Database muncul dalam Main Proses seperti gambar 5.5.



Gambar 6.5 Database dalam Main Process

Untuk melakukan Market Basket Analysis, kita membutuhkan setidaknya tiga buah operator, antara lain Association Rule, FP-Growth, dan Numerical to Binomial.



Gambar 6.6 Operator Create Association Rules

Assocuation rules dilakukan dengan menganalisis data pada frequent if/then patterns

menggunakan kriteria support dan confidence untuk mengidentifikasikan suatu relasi antar item. *Frequent if/then pattern* digali menggunakan operator FP-Growth. Operator Create Association Rules menggunakan frequent itemsets ini dan menghasilkan association rules.



Gambar 6.7 Operator FP-Growth

Frequent itemsets merupakan kelompok item yang sering muncul bersama-sama dalam data. Operator *FP-Growth* mengkalkulasikan semua frequent itemset dari input yang diberikan menggunakan struktur data FP-tree. Adalah wajib bahwa semua atribut dari masukan merupakan bilangan binominal (true/false).



Gambar 6.8 Operator Numerical to Binominal

Operator *Numerical to Binominal* diperlukan untuk mengubah nilai atribut yang berada pada table TransaksiMakanan menjadi binominal.

Selanjutnya lakukan Pencarian Filter untuk memudahkan kita menemukan operator yang dibutuhkan, lakukan seperti pada gambar berikut.



Gambar 6.9 Pencarian Operator Numerical to Binominal

Untuk Mencari Operator *Numerical to Binominal*, lakukan pencarian seperti gambar disamping. Operator ini terdapat pada hirarki: Data Transformation I Type Conversion



Gambar 6.10 Pencarian Association Rules

Susunlah ketiga operator tersebut menjadi seperti gambar 5.11.



Gambar 6.11 Menghubungan Database Transaksi Makanan pada Operator Numerical to Binomial

Hubungkan Tabel TransaksiMakanan yang kita miliki dengan operator Numerical to Binominal. Proses ini akan membuat nilai dari Tabel Transaksi makan mejadi *Binominal Attributes*.

🛛 🛃 Parameters	🛛 🔘 Context 🗶
🚨 💀 💀 🦻	🕵 👼 -
🍕 Numer	rical to Binominal
attribute filter type	all 🔻
invert selection	
include special	attributes
min	0.0
max	0.0

Gambar 6.12 Parameter Numerical to Binomial

Data yang kita miliki merupakan data sederhana. Kita hanya memperhitungkan 1 buah penjualan produk pada setiap transaksinya. Maka nilai yang terbaik untuk menjadi *false* adalah ketika tidak ada produk tertentu yang terjual dalam suatu transaksi, jadi kita sini nilai *min* dan *max* menjadi 0, Sehingga yang bernilai *false* adalah ketika sebuah produk tidak terdapat pada sebuah transaksi.

Hubungkan operator *Numerical to Binominal* dengan operator FP-Growth pada example output.



Gambar 6.13 Menghubungkan Operator Numerical to Binomial dengan Operator FP-Growth

Terdapat dua buah output untuk *Numerical to Binominal*, yaitu example dan original.

- Example, *numeric attributes* dikonversikan menjadi *binominal attributes* melalui output ini.
- Original, *numeric attributes* dilewatkan tanpa konversi. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

Lewatkan output pada *example*.

Isilah Parameter FP-Growth seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum support* senilai 30% atau 0.3.

🕞 Parameters 🏾	🔘 Context 🗶
🚨 🗟 🖻 🦻 🕵	: 📭 👻
🎘 FP-	Growth
✓ find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.3
max items	-1
must contain	

Gambar 6.14 Parameter FP-Growth

Kemudian hubungkan operator *FP-Growth* dengan operator *Association Rules*.



Gambar 6.15 Menghubungkan Operator FP-Growth dengan Operator Create Association Rules

Terdapat dua buah output pada operator FP-Growth, yakni *example* dan *frequent*.

• *Example*, input yang diberikan dilewatkan tanpa adanya perubahan. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

• *Frequent*, frequent itemset dikirimkan melalui output ini.

Lewatkan output pada frequent.

Kemudian isilah Parameter Association Rules seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum confidence* senilai 50% atau 0.5.



Gambar 6.16 Parameter Association Rules

Setelah itu hubungkan Association Rules pada result. Sehingga seluruhnya membentuk seperti gambar 5.17. Ialu klik ikon Play Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 6.17 Susunan Operator Association Rules

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 152 rules yang ditemukan.

No.	Premises	Conclusion	Support	Confid	LaPI	Gain	p-s	Lift	Convi
131	TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
132	ROTI, SUSU	MENTEGA, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
133	MENTEGA, SUSU	ROTI, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
134	ROTI, MENTEGA, SUSU	TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
135	ROTI, TELUR, SUSU	MENTEGA, KECAP	0.200	1	1	-0.200	0.160	5	00
136	MENTEGA, TELUR, SUSU	ROTI, KECAP	0.200	1	1	-0.200	0.160	5	00
137	ROTI, MENTEGA, TELUR, SUSU	KECAP	0.200	1	1	-0.200	0.160	5	00
138	KECAP	ROTI, MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
139	ROTI, KECAP	MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
140	MENTEGA, KECAP	ROTI, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
141	ROTI, MENTEGA, KECAP	TELUR, SUSU	0.200	1	1	-0.200	0.120	2.500	00
142	TELUR, KECAP	ROTI, MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
143	ROTI, TELUR, KECAP	MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
144	MENTEGA, TELUR, KECAP	ROTI, SUSU	0.200	1	1	-0.200	0.160	5	00
145	ROTI, MENTEGA, TELUR, KECAP	SUSU	0.200	1	1	-0.200	0.120	2.500	00
146	SUSU, KECAP	ROTI, MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
147	ROTI, SUSU, KECAP	MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
148	MENTEGA, SUSU, KECAP	ROTI, TELUR	0.200	1	1	-0.200	0.120	2.500	00
149	ROTI, MENTEGA, SUSU, KECAP	TELUR	0.200	1	1	-0.200	0.080	1.667	00
150	TELUR, SUSU, KECAP	ROTI, MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
151	ROTI, TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
152	MENTEGA, TELUR, SUSU, KECAP	ROTI	0.200	1	1	-0.200	0.040	1.250	∞

Gambar 6.18 Hasil Association Rules Pertama

Tentunya ini akan menyulitkan kita untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Maka dari itu yang harus kita lakukan adalah mengubah nilai *minimum support* dan *minimum confidence*.

Klik ikon Edit untuk kembali pada *model view*. Lalu klik Operator FP-Growth.



Gambar 6.19 Operator FP-Growth

Kemudian lihat bagian parameter. Ubah nilai minimum support menjadi 95%, seperti yang sudah dijelaskan pada sub bab Association Rules, semakin tinggi nilai support maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum supportnya terlalutinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

🛛 🛃 Parameters 🕱	🔵 Context 😫
2 5 5 9 9	: 📭 -
🎘 FP-	Growth
🖌 find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.95
max items	-1
must contain	

Gambar 6.20 Mengubah Parameter FP-Growth

Sekarang kita beralih pada Operator Create Association Rules.



Gambar 6.21 Operator Create Association Rules

Ubah nilai minimum confidence menjadi 95% atau 0.95, semakin tinggi nilai confidence maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum confidence terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

criterion	confidence 🔻
min confidence	0.95
gain theta	2.0
laplace k	1.0

Gambar 6.22 Mengubah Parameter Association Rules

Klikikon Play Duntuk menampilkan hasil yang baru.

No.	Premises	Conclusion	Support	Confid	LaPla	Gain	p-s	Lift	Convi
1	ROTI	MENTEGA	0.800	1	1	-0.800	0.160	1.250	00
2	MENTEGA	ROTI	0.800	1	1	-0.800	0.160	1.250	00
3	SUSU	TELUR	0.400	1	1	-0.400	0.160	1.667	80
4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400	0.080	1.250	00
5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400	0.080	1.250	00

Gambar 6.23 Hasil Association Rules Kedua

Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

Untuk melihat dalam bentuk grafik. kita dapat memilih opsi Graph View. O Table View O Graph View: O Text View O Annotations



Gambar 6.24 Hasil dalam bentuk Graph View

Glossarium

- *Algoritma* Kumpulan perintah untuk menyelesaikan suatu masalah.
- Apriori Algoritma untuk frequent itemset mining dan association rule dalam database transaksional. DIhasilkan dengan menidentifikasi setiap buah item, dan memperluasnya menjadi kombinasi kumpulan item yang lebih besar asalkan himpunan item muncul cukup sering dalam database
- Association Rules Sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar.
- **Binominal Attributes** Atribut dengan tipe Binominal (true dan false).
- Confidence (Market Basket Analysis) Seberapa sering item dalam Y muncul di transaksi yang mengandung X.
- Decisiontree Struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan

simpul daun merepresentasikan kelasatau distribusi kelas.

- Disjoint Himpunan terpisah, tidak ada elemenyang berhubungan diantara kedua himpunan yang bersangkutan
- *Flowchart* Sebuah diagram dengan simbol-simbol grafis yang menyatakan aliran algoritma.
- FrequentItemset Itemset yang memunyai support >= minimum support yang diberikan oleh user dalam Market Basket Analysis.

Market Basket Analysis Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan.

MBA Lihat Market Basket Analysis.

Minimum SupportNilai Support Terkecil dalam MarketBasket Analysis yang dapat di toleransi.

Minimum Confidence Nilai Confidence terkecil dalam Market Basket Analysis yang dapat di toleransi.

- Neural Network Jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syarafmanusia.
- Numeric Attributes Attribut dengan tipe Numerical (1-9).
- *Operator* suatu tanda atau simbol yang dipakai untuk menyatakan suatu operasi atau manipulasi nilai.
- ParameterNilai yang mengikuti acuan keterangan
atau informasi yang dapat menjelaskan
batas-batas tertentu dari suatu suatu sistem persamaan.

- **Pruning** Teknik dalam machine learning yang mengurangi ukuran pohon keputusan dengan menghapus bagian dari pohon yang memberikan sedikit kekuatan untuk mengklasifikasikan kasus.
- RapidMinerSebuah tool yang digunakan untuk
melakukan analisis terhadap data mining,
text mining dan analisis prediksi.
- RepositoriKumpulan paket yang siap untuk diambil
dan digunakan sesuai dengan kebutuhan
pengguna.
- *Simpulakar* Simpul tanpa ayah yang berada pada tingkat tertinggi.
- *Simpul daun* Semua simpul yang berada pada tingkat terendah.
- *Simpul internal* Semua simpul dari pohon yang memiliki anak tetapi bukan daun.
- Support(Market Basket Analysis) Bagian transaksi
yang mengandung kedua X dan Y.
- TeorigrafCabang kajian yang mempelajari sifat-sifat
graf.
- ValidasiTindakan yang membuktikan bahwa suatu
proses/metode dapat memberikan hasil
yang konsisten sesuai dengan spesifikasi
yang telah ditetapkan.

Daftar Pustaka

Akhtar, Fareed dan Caroline Hahne. 2012. *RapidMiner 5 Operator Reference*, [online], (www.rapid-i.com, diakses tanggal 30 Januari 2013).

Amiruddin, dkk. Penerapan Association Rule Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan Untuk Menemukan Pola Sertifikasi Guru. Institut Teknologi Surabaya. Surabaya.

Basuki, Achmad dan Iwan Syarif. *Decision Tree*, [online], (http://lecturer.eepis-its.edu/~entin/Data%20Mining/ Minggu%205%20Decision%20Tree.pdf, diakses tanggal 05 Februari 2013).

Khusnawi. 2007. *Pengantar Solusi Data Mining*. Yogyakarta.

Kusumadewi, Sri. 2003. Artificial Intelligence: Teknik dan Aplikasinya.

Mitchel, Tom M. 1997. *Machine Learning*. New York: McGraw-Hill.

Prasetyo, Bowo. 2011. *Mengenal RapidMiner*, [online], (www.slideshare.net/bowoprasetyo/RapidMiner, diakses tanggal 31 Januari 2013).

Prasetyo, Kokoh Philips. 2006. *APriori*, [online] (http://philips.wordpress.com/2006/06/07/apriori, diakses tanggal 03 Februari 2013)

Rafaida, Ropi. *Decision Tree (Pohon Keputusan)*, [online],

(http://file.upi.edu/Direktori/FPEB/PRODI._MANAJEME N_FPEB/197302052005012-

ROFI_ROFAIDA/MATERI_KULIAH/DECISION_TREE.pdf, diakses tanggal 05 februari 2013).

Ross, Peter. 2000. Data Mining [online]. (http:// www.soc.napier.ac.uk/~peter/vldb/dm/dm.html, diakses tanggal 07 Februari 2013)

Wahono, Romi satria. *Data Mining: Proses Data Mining*, [online], (http://romisatriawahono.net/lecture/dm/ romi-dm-02-proses-june2012.pptx, diakses tanggal 31 Januari 2013).

2012. *RapidMiner* 5.0 *Manual English*, (online), (www.rapid-i.com, diakses tanggal 30 Januari 2013).

3 tips for Setting up Association Rules using RapidMiner, [online]. (http://www.simafore.com/blog /bid/110113/3-tips-for-setting-up-a-Market-Basket-Analysis-using-RapidMiner, diakses tanggal 08 Maret 2013).

Association Analysis: Basic Concepts and Algorithms, [online]. (http://www-users.cs.umn.edu/ ~kumar/dmbook/ch6.pdf, diakses tanggal 08 April 2013)

Decision Tree (Pohon Keputusan), [online], (http://www.google.co.id/url?sa=f&rct=j&url=http://no vrina.staff.gunadarma.ac.id/Downloads/files/21783/Alg oritma%2BC4.pdf&q=algoritma+c4&ei=6h9gUcbJFlqrrA

fT7IGQAw&usg=AFQjCNG7HbyNPOqa63ZoPexX76TrIIJ7g, diakses tanggal 05 februari 2013).

Landasan Teori Market Basket Analysis, [online]. (http://library.binus.ac.id/eColls/eThesis/Bab2/2010-1-00498-MTIF%20Bab%202.pdf, diakses tanggal 08 April 2013)

Understanding the Concept of Market Basket Analysis, [online]. (http://www.thesmartcube.com/ insights/blog/brand-strategy/understanding-theconcept-of-market-basket-analysis, diakses tanggal 08 Maret 2013)

RapidMiner Resources. (http://RapidMinerresources .com/uploads/videos/tomott/RapidMiner5-Vid1.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%201.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%202.flv, diakses tanggal 02 Februari 2013)

Nama : Wahyu Putra Adi Wibowo

NIM : 202420041

Dataset : diabetes_data_upload.csv

- 1. Import ke local repository
- 2. Design & drag dataset dan drag operator decision tree ke dalam rapidminer studio

File Edit Process View Connections Settings Extensions Help Design Results Turbo Prep Auto Model -Ŧ - -Views: Deployments Repository Process 🔎 🔎 🐚 🚺 📮 🗸 🖉 🕒 Import Data ≡ ▼ Process Training Resources Samples Retrieve diabetes_d... • 🚨 Community Samples (c res Ċ res Local Repository (Local) res Connections res 🕶 📒 Data diabetes_data_upload (Set Role Process exa exa < 11 mod exa Operators × deci

Hasilnya



Attibute | Weight

<u>File Edit Proces</u>	s <u>V</u> iew <u>C</u> onnections <u>S</u>	ettings E <u>x</u> tensior
	•	
9	Tree (Decision Tree)	×
Result Hi	istory	
Data	attribute	weight
	partial paresis	0.003
	Obesity	0.031
Weight	Irritability	0.041
Visualizations	Gender	0.076
ming	Itching	0.030
	Alopecia	0.127
Annotations	Genital thrush	0.071
	delayed healing	0.091
	sudden weight loss	0.117
	visual blurring	0.117
	Polyuria	0.042
	Polydipsia	0.040
	Age	0.215

Tampilan Pada ExampleSet

<u>File Edit Proces</u>	s <u>V</u> iew <u>C</u> onne	ctions <u>S</u> ettings	Extensions Help						
	-	•	Views:	Design Resul	ts Turbo Prep	Auto Model D	eployments	F	ind data
2	Tree (Decision Tr	ee) ×		ExampleS	et (//Local Repository/dia	betes_data_uplo	ad) ×		
Result H	istory		ExampleSet (Set	Role) ×	💐 At	tributeWeights (D	ecision Tree)	×	
Open in Turbo Prep Auto Model Filter (520 / 520 examples): all Data Itching Irritability delayed healing partial paresis muscle stiffness Alopecia Obesity label									•
	itening		ucia) cu neuling	partial parcolo		Hopeelu	obcony		^
_	Yes	No	Yes	Yes	No	No	No	Positive	~
Σ	No	No	Yes	Yes	Yes	No	No	Positive	
Statistics	No	No	Yes	No	Yes	No	No	Positive	
	No	Yes	Yes	Yes	Yes	Yes	Yes	Positive	

Hasil Visual weight



Hasil Dari Auto Model

<u>Eile Edit Pro</u>	ocess ⊻iew <u>C</u> onne	ections <u>S</u> ettings E	tensions <u>H</u> elp									
	•	•	Viev	vs: Design	Results	Turbo Prep	Auto Model	Deployments	[Find data, operatorsetc	P	All Stu
Auto Model												
			Load Data	Select Task	Prepare Target	Select Inputs	Model Types	Results				
			•									
				// PES								
				1 1/25								
			Predict		Clu	usters		Outliers				
		Wa	nt to predict the values	of a column?	Want to identify	groups in your data?	Want to de	tect outliers in your da	ta?			
Age	Gender	Polyuria	Polydinsia	sudden weig	weakness	Polyphagia	Genital thrush	visual blurring	Itching	Irritability	delave	ed b
Number	Category	Category	Category	Category	Category	Category	Category	Category	Category	Category	Category	
40	Male	No	Yes	No	Yes	No	No	No	Yes	No	Yes	^
58	Male	No	No	No	Yes	No	No	Yes	No	No	No	
41	Male	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	
45	Male	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	
60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	
55	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	
57	Male	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	

Pengelompokkan (Cluster)



Judul : Pendeteksi Diabetes

Dataset : diabetes_data_upload.csv

JIka di konversi Yes=1 dan No=0, Male=1 dan Female=0, ujicoba menggunakan data Diabetes_data1.csv



📭 AttributeWeights (Decision Tree) 🛛 🗙

attribute	weight
Obesity	0.026
Irritability	0.006
Gender	0.026
muscle stiffness	0.033
weakness	0.001
Alopecia	0.045
Genital thrush	0.071
sudden weight loss	0.124
delayed healing	0.259
visual blurring	0.124
Polyuria	0.052
Polydipsia	0.026
Age	0.207

11	AttributeWeights ([Decision Tree)	× 🚦 Ex	ampleSet (Set F	Role) ×	💡 Tree (Decisio	n Tree) $ imes$	Examples	Set (//Local Repos	itory/Data/diabete	∍s_data1)	×
Open in	Turbo Prep	Auto Model							Filter (520 /	520 examples):	all	
Row No.	label	Age	Gender	Polyuria	Polydipsia	sudden wei	weakness	Polyphagia	Genital thrush	visual blurring	Itching	Ir
1	Positive	40	1	0	1	0	1	0	0	0	1	0
2	Positive	58	1	0	0	0	1	0	0	1	0	0
3	Positive	41	1	1	0	0	1	1	0	0	1	0
4	Positive	45	1	0	0	1	1	1	1	0	1	0
5	Positive	60	1	1	1	1	1	1	0	1	1	1
6	Positive	55	1	1	1	0	1	1	0	1	1	0
7	Positive	57	1	1	1	0	1	1	1	0	0	0
8	Positive	66	1	1	1	1	1	0	0	1	1	1
9	Positive	67	1	1	1	0	1	1	1	0	1	1
10	Positive	70	1	0	1	1	1	1	0	1	1	1
11	Positive	44	1	1	1	0	1	0	1	0	0	1

Hasil Dari Decision Tree





Automodel > Predict



NAMA : WIDIA ASTUTI NIM : 202420021 MATA KULIAH : ADVANCED DATABASE

TUGAS 6

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawab :

Data Debitur :

		NAMA		MAKSIMUM	STATUS
ID	JENIS USAHA	DEBITUR	NO KTP	KREDIT	KREDIT
1	PERDAGANGAN	ANI	161004700389001	Rp100.000.000	LANCAR
2	RESTORAN	BUDI	161004700287002	Rp250.000.000	MACET
3	JASA SEWA	CINTA	161004700388001	Rp200.000.000	LANCAR
4	PERDAGANGAN	DONI	161004700285001	Rp150.000.000	LANCAR
5	PERTANIAN	ERNI	161004700888001	Rp25.000.000	MACET
6	PERIKANAN	FADLAN	161004700689001	Rp50.000.000	LANCAR
7	PERTANIAN	GIA	161004700886002	Rp30.000.000	MACET

Langkah-langkah :

1. Klik menu File -> New Process -> Pilih Blank Selanjutnya pada kota dialog Repository pilih Add Data, -> My Computer, cari lokasi file Data yang akan diimport.



	Import D	Data - Select the ce	ils to import.	Select the cell	s to import	_		Questions?
epository X	Sheet	Sheet1 V	Cell range: A:F	Select the cell	elect All	ne header row: 1)	init •
Local Repository (Sony)	A		в	с	D	E	F	
🜉 data mining repository (Sony)	1 1)	JENIS USAHA	NAMA DEBITUR	NO KTP	MAKSIMUM KREDIT	STATUS KREDIT	
📒 data (Sony)	2 1	.000	PERDAGANGAN	ANI	16100470038900	10000000.000	LANCAR	
proses (Sony)	3 2	.000	RESTORAN	BUDI	16100470028700	25000000.000	MACET	
	4 3	.000	JASA SEWA	CINTA	16100470038800	20000000.000	LANCAR	
	5 4.	.000	PERDAGANGAN	DONI	16100470028500	15000000.000	LANCAR	
	6 5	.000	PERTANIAN	ERNI	16100470088800	2500000.000	MACET	
erators ×	7 6.	.000	PERIKANAN	FADLAN	16100470068900	5000000.000	LANCAR	parameters
amb far Onorstam	8 7.	.000	PERTANIAN	GIA	16100470088600	3000000.000	MACET	tibility (7.3.001)
Data Access (13)								
DI 1 (77)								-
Elending (77)								
Cleansing (26)	1							Studio Core
Cleansing (26) Modeling (125)								

2. setelah menemukan file datanya, klik Next, kemudian setelah sampai pada kotak dialog berjudul Format Your Columns dan pada masing-masing atribut pilih logo setelan.

			Format your	r columns.			
8	ate format MMM d.y	m h mm a a a	•	Replace errors with	missingvalues (1)		
	Jumlah MK o + Integer	Absensi (%) o + /ca/	Jemlah SK o + Integer	IPS St o + NW	IPS S2 0 + mar	825.53 e • .ew	N p
1	0 /	0.000	0	0.000	2	9	٦
2	0 /	0.000	0	0.000	7	7	1
3	23	38.002	39	0.290	0.000	0.000	
4	1	53.750	50	2.380	2.490	0.000	1
5	37	66.952	60	2.259	2.820	1.800	
4	20	47.509	60	1.500	2.950	0.100	
7	35	71.845	59	1.070	2.820	2,773	1
	43	54.771	18	2.268	1.810	1,370	1
9	8	6.253	18	0.360	0.000	9	1
10-	8	6.253	60	0.360	0.000	9	1
11	25	75.821	39	3.030	2.710	2.700	1
	11	28.365	0	2.530	0.000	0.000	1.6

Pilih Change Type. Kriterianya sebagai berikut ;

- Polynominal = untuk atribut yang memiliki lebih dari 2 kategori
- **Binominal** = untuk atribut yang memiliki 2 kategori saja
- **Real** = untuk tipe data yang memiliki nilai koma, atau decimal
- Integer = untuk file bernilai integer atau bilangan buat tanpa koma.

3. Kemudian pada atribut kelasnya atur Change Role dan isi valuenya sebagai label,

selanjutnya klik Next sampai dengan finish

ngimport Da 🗙 🔨	20201-MTIK112 >	K 💙 😒 WhatsApp	× 📐 R	apidMiner Stuc 🗙	🖌 📐 Final Stage -	Rai 🗙 📉 Kot	tak Masuk - 🗤 🗙	M Verify Your Ema	RapidMiner Acc ×
I <new process=""></new>	– RapidMiner Studie	o gui.license.rapidm	iner-studio.educat	onal.label 7.3.001 @	Sony-VAIO				
Eile Edit Proce	ess <u>V</u> iew <u>C</u> onne	ctions <u>S</u> ettings	E <u>x</u> tensions						
	-		 • 	Views:	Design	Results			Questions? 🔻
Result History	× 📕 Exa	ampleSet (//data	mining reposito	ry/data/DATA DE	BITUR) ×				Repository ×
	ExampleSet (7 ex	amples, 1 special	attribute, 5 regular	attributes)		Filter (7	7/7 examples):	all 🔻	🕒 Add Data 🛛 🖃 💌
Data	Row No.	STATUS KR	ID	JENIS USAHA	NAMA DEBIT	NO KTP	MAKSIMUM		Samples
Duta	1	LANCAR	1	PERDAGANG	ANI	2147483647	10000000		Local Repository (Sony)
	2	MACET	2	RESTORAN	BUDI	2147483647	250000000		 data mining repository (Sony) data (Sony)
Σ	3	LANCAR	3	JASA SEWA	CINTA	2147483647	20000000		DATA DEBITUR (Sony - v1, 10/2)
Statistics	4	LANCAR	4	PERDAGANG	DONI	2147483647	150000000		proses (Sony)
	5	MACET	5	PERTANIAN	ERNI	2147483647	25000000		
	6	LANCAR	6	PERIKANAN	FADLAN	2147483647	5000000		
Charts	7	MACET	7	PERTANIAN	GIA	2147483647	3000000		
Advanced									
- Charlo									
Annotations									
2									<

4. Setelah finish pada halaman Repository akan muncul data teman-teman yang sudah Terimpor oleh sistem. Kemudian Drag atau seret ke halaman kerja RapidMiner



kemudian klik Run (F11)

gimpo	ort Dai 🗙 🗸 🍒	20201-MTIK112	× 🔍 😒 WhatsApp	× 🖉 🖬	RapidMiner Stuc 🗙	🗸 📐 Final Stage -	Ral 🗙 📉 M Ko	tak Masuk - 🗤 🗙 🚩 M Verify Y	our Emai 🗙 🗸	RapidMiner Acc 🗙 🔽 👛 😐
•	<new process*=""></new>	– RapidMiner Stu	dio gui.license.rapidr	niner-studio.ed	ducational.label 7.3.001	@ Sony-VAIO				
Eil	e <u>E</u> dit <u>P</u> roce	iss <u>V</u> iew <u>C</u> onn	ections <u>S</u> ettings	Extensions						
		-	\hookrightarrow \hookrightarrow	•	Views:	Design	Results			Questions? 🔻
			Exar	npleSet (//dat	a mining repository/da	ata/DATA DEBITU	IR) ×			Repository ×
	Re	sult History	×		Exa	mpleSet (Retrie	ve DATA DEBIT	TUR) ×		Add Data
^		ExampleSet (7 e	examples 1 special	attribute 5 rec	ular attributes)		Filter (7/7 examples): all	•	
		Row No.	STATUS KR	ID	JENIS USAHA	NAMA DEBIT	NO KTP	MAKSIMUM		 Samples Local Repository (Sony)
	Data	1	LANCAR	1	PERDAGANG	ANI	2147483647	10000000		data mining repository (Sony)
		2	MACET	2	RESTORAN	BUDI	2147483647	25000000		data (Sony)
	Σ	3	LANCAR	3	JASA SEWA	CINTA	2147483647	20000000		proses (Sony)
	Statistics	4	LANCAR	4	PERDAGANG	DONI	2147483647	15000000		
		5	MACET	5	PERTANIAN	ERNI	2147483647	25000000		
_		6	LANCAR	6	PERIKANAN	FADLAN	2147483647	5000000		
	Charts	7	MACET	7	PERTANIAN	GIA	2147483647	3000000		
	-									
	Advanced Charts									
\sim	Annotations									<

5. pada kotak dialog **Operators**, pilih menu-menu seperti **Random Tree, Split Data, Apply Model, dan Permormance.** Cara mencarinya tulisan di kotak pencarian pada kotak dialog tersebut.

17 · 10) =	a mandar a la	TUGAS 6 WIDIA ASTUTI 2024	20021 - Microsoft Word	the cardy have been in the Ameridan in	
<pre>// <new process*=""> - RapidMiner Stu</new></pre>	dio qui license rapidminer-studio.edu	cational.label 7.3.001 @ Sonv-VAIO			×
Eile Edit Process View Conn	nections <u>S</u> ettings E <u>x</u> tensions	,			
		Views: Design	Results		Questions? 🔻
Repository ×	Process ×			Parameters ×	
🔂 Add Data 🛛 = 👻	Process		100% 🔑 🔎 🔑 🛃	🗣 💣 🔯 🗏 Performance (Performan	ce (Classification))
mples) inp	out		re A main criterion first	• 0
cal Repository (Sony) ta mining repository (Sony)				re accuracy	©
data (Sony)		Random Tree		classification error	٢
DATA DEBITUR (Sony - v1, 10/29/2 proses (Sony)	•	ra mod) exa		🗌 карра	D
<		Split Data		weighted mean recall	Ð
Operators ×	•	par		weighted mean precision	© ~
dation (18)		Apply Model		Help ×	
Predictive (7)	•	uni mod		Performance	^
Serformance (Binominal Cl V	6	ab @g per		(Classification) RapidMiner Studio Co Tags: Accuracy, Errors, Precisi	re
We found "Model Management" in the Marketplace. <u>Show mel</u>	<	er exa		Squared, Relative, Validations, Metrics, Predictive	Evaluations,

Setelah menemukan semua **Opertors** tersebut, jangan lupa Drag atau seret ke halaman kerja satu persatu.

6. Selanjutnya pada **operator Split Data klik 2x**, kemudian atur Sampling Type pada kota Parameters , ini bertujuan untuk memisah atau membagi menjadi data latih dan data uji guna untuk mengetahui permorma dari model yang dibentuk oleh k-NN

100% 🔎 🔎 📮 🌄 🕯	EH N	Parameters	×	
	-	narffione	Bdit Enumer	Ð
		sampling type	linear sampli 🔻	Ð
Spit Data (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		use local rand	lom seed	Ð
		Hide advance	d parameters	
		neip -		-

7. Setelah itu atur porposisi atau kapasitas antara data latih dan data uji. Klik pada menu edit *Enumeration*. kemudian akan muncul kotak dialog sepeti dibawah ini :

	<u> ヴ・び)</u> =		TUGAS 6 WIDIA ASTUTI 202420021 - Microsoft Word	
He	<new process*=""> - RapidMiner Stu</new>	dio gui.li	cense.rapidminer-studio.educational.label 7.3.001 @ Sony-VAIO	
8	<u>File Edit Process View Conn</u>	ections	Settings Extensions	
V		c)	Viewe: Design Results Edit Parameter List: partitions	Questions? -
	Repository ×	Proc	Edit Parameter List partitions The partitions that should be created.	rs ×
	mples	Pro	ratio	😼 Edit Enumeration (🛈
	cal Repository (Sony)			pe shuffled sampling 💌 🛈
	ta mining repository (Sony)) inp	·	
	data (Sony)			
	DATA DEBITUR (Sony - v1, 10/29/			
	<			
	Operators ×			
	split X			dvanced parameters
	r 🛅 Blending (2) 🔨			~
	Examples (1)			
	Sampling (1)			it Data 📃
	Values (1)			dMiner Studio Core
	<		📑 Add Entry 🔤 Remove Entry 🗸 OK 🗶 Cancel	s, <u>separate</u> , <u>Part</u> , <u>Training</u> , <u>Testing</u> , ubsets, <u>Partitions</u> , <u>Sampling</u>
	We found "Information Extraction" in the		Synopsis	
	Marketplace. Show me!		This opera	tor produces the desired number of 🤍
of 5		_		

8. Add Entry kemudain Ok.

Sumber :

https://www.youtube.com/watch?v=h82NuHDNhKI

https://www.pengalaman-edukasi.com/2020/02/tutorial-rapidminer-menggunakan_6.html

https://www.youtube.com/watch?v=jB_hKsi408

https://www.youtube.com/watch?v=jXRXI66inDM

TUGAS 6 ADVANCDED DATABASE



Di Susun Oleh : AAN NOVRIANTO NIM : 202420010

Dosen Pengasuh : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

Program Pasca Sarjana

Universitas Bina Darma Palembang

2020/2021

Tugas 6

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban :

Pada kasus ini saya mengambil pengolahan data terkait dengan Market Basket yang akan dicoba pada tool rapidminer. Adapun tahapannya sebagai berikut :

1. Create tambel master datanya di Excel kemudian export ke Rapidminer.

TID	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP
001	1	1	1	0	0	0	0
002	0	1	1	1	0	0	0
003	0	0	0	1	1	1	0
004	0	1	1	0	0	0	0
005	0	1	1	1	0	1	1

Export ke Rapidminer dan sesuaikan type datanya :

Х 📒 Ез	kampleSet (//Lo	ocal Repository	/data/Tugas 06/1	Fugas 06_market b	asket analysis _.	_update) ×		
xampleSet (5	examples, 0 spe	cial attributes, 8 re	gular attributes)			I	Filter (5 / 5 examples):	all
Row No.	TID 个	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP
1	001	1	1	1	0	0	0	0
2	002	0	1	1	1	0	0	0
3	003	0	0	0	1	1	1	0
4	004	0	1	1	0	0	0	0
5	005	0	1	1	1	0	1	1

Repository ×		Process ×		
G Add Data	■	Process >		100% 🔎 🔎 📮 ≩
 Tugas 06 (bhijanta.wijayamukti) Tugas 06 (bhijanta.wijayamukti - v1, 10/28/20 8:51 Pi Rapidminer_Tugas 06 (bhijanta.wijayamukti - v1, 1 Tugas 06_market basket analysis (bhijanta.wijaya Tugas 06_market basket analysis (bhijanta.wijaya Tugas 06_market basket analysis (bhijanta.wijaya Tugas 06_market basket analysis_update (bhija Tugas 06_market basket analysis_update (bhija Tugas 06_market basket analysis_update (bhija 	M - 390 byt 0/27/20 9:5 amukti - v1, amukti - v1, nta.wijayan	ocess Retrieve Tugas 06 out	Numerical to Binomi	FP-Growth
Operators Search for Operators Modeling (125)	>			

2. Create Modul Association Rule, FP-Growth, dan Numerical to Binomial.

Numerical to Binomial.

Process \times Parameters × Numerical to Binominal 100% 🔎 🔎 📮 🛃 🗾 Process > • (1) attribute filter type all Process Retrieve Tugas 06_... Numerical to Binominal FP-Growth ٩ invert selection 🚺 exa 👔 exa res) inp exa 🗋 xa Ċ ori res (fn (j) include special attributes 0.0 ٢ min 0.0 max C Create Association ... rul 1 ite ite D Change compatibility (7.1.001)

Sesuaikan dengan parameter – parameter yang ada.

FP-Growth

Process ×		Parameters ×	
Process >	100% 🔑 🔑 📴 🧟 💣 🔃	FP-Growth	
Process		find min number of iter	msets 🛈 🔨
Retrieve Tugas 06 Numerical to Binomi	FP-Growth exa car res	min number of 100	(D)
on	The res (max number of 15	•
		positive value	
		min support 0.3	D
	Create Association	max items -1	
	ite rul	must contain	eters

Sesuaikan dengan parameter-parameter yang ada.

Association Rule

Process Parameters × × 100% 🔎 🔎 📮 🧣 🗾 F Create Association Rules Process > criterion confidence -Process Retrieve Tugas 06_... Numerical to Binomi... FP-Growth min confidence 0.5 exa 鹶 exa res out exa exa) inp C 1 Ţ ori res gain theta 2.0 laplace k 1.0 **Create Association Rules** ite 🛒 rul ite D Hide advanced parameters

Sesuaikan dengan parameter – parameter yang ada.



3. Create hubungan antar modul mulai dari data master sampai dengan Association Rule.

Setelah dilakukan hubungan antar modulnya kita lakukan RUN disistem, Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 152 rules yang ditemukan.

Result History	🗙 🛒 AssociationRul	les (Crea	te Associ	ation Rules) $ imes$		Repository ×
	Show rules matching		No.	Premises	Conclus	🔂 Add Data 🛛 🖃 💌
Data	all of these conclusions:		137	ROTI, MENTEGA, TELUR, SUSU	KECAP 🔨	🕶 🛅 Samples
Data	ROTI		138	KECAP	ROTI, ME	🕨 🥁 data
	MENTEGA TELUR		139	ROTI, KECAP	MENTEG	processes
Z	SUSU		140	MENTEGA, KECAP	ROTI, TE	Tutorials
Graph	BUNCIS		141	ROTI, MENTEGA, KECAP	TELUR	DB
			142	TELUR, KECAP	ROTI, ME	Local Repository (bhijanta.wijayamukti)
			143	ROTI, TELUR, KECAP	MENTEG	 Tugas 06 (bhijanta.wijayamukti)
Description			144	MENTEGA, TELUR, KECAP	ROTI, SL	Tugas 06 (bhijanta.wijayamukti - v1.
		- 1	145	ROTI, MENTEGA, TELUR, KECAP	susu	Rapidminer_Tugas 06 (bhijanta.w
			146	SUSU, KECAP	ROTI, ME	Tugas 06_market basker analysi
Annotations			147	ROTI, SUSU, KECAP	MENTEG	Tugas 06_market basket analysi
			148	MENTEGA, SUSU, KECAP	ROTI, TE	Tugas 06_market basker analysi
			149	ROTI, MENTEGA, SUSU, KECAP	TELUR	processes (bhijanta.wijayamukti)
	Min. Criterion:		150	TELUR, SUSU, KECAP	ROTI, ME	Cloud Repository (disconnected)
	confidence	•	151	ROTI, TELUR, SUSU, KECAP	MENTEG	
	Min. Criterion Value:		152	MENTEGA, TELUR, SUSU, KECAP	ROTI	
			<		>	<

Tentunya ini akan menyulitkan kita untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Maka dari itu yang harus kita lakukan adalah mengubah nilai minimum support dan minimum confidence.

Oleh karena itu kita harus mengubah nilai di FP-Growth dan Create Association Rules untuk dapat mencapai nilai minimum support dan minimum confidence, untuk dapat menghasilkan skema yg lebih sedikit untuk dapat mempermudah dalam penyusunan skema market.



BEFORE(0.3)









BEFORE(0.5)



AFTER(0.95)



Setelah dilakukan perubahan parameter lakukan RUN kembali, Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk

dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

les matching			No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
se conclusio	ons: 🔻	1	1	ROTI	MENTEGA	0.800	1	1	-0.800
		2	2	MENTEGA	ROTI	0.800	1	1	-0.800
		3	3	SUSU	TELUR	0.400	1	1	-0.400
		4	4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400
		5	5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400
				^					
Data	Zoom Mode	₽	<i>,</i> ₽ <u>↓</u>	· · · · · · · · · · · · · · · · · · ·				[Rule 2 (0.800 / 1.000)

SUMBER :

Belajar Data Mining

dengan RapidMiner

Penyusun:

- Dennis Aprilla C
 Donny Aji Baskoro
 Lia Ambarwati
 IWayan

Simri Wicaksana

Editor: Remi

Sanjaya

Hak Cipta © pada Penulis

Hak Guna mengikuti Open

Content model Desain sampul:

Dennis Aprilla C

NAMA : AHMAD ALI MA'MUN NIM : 202420037

Open lembar kerja baru sobat dengan cara double click NEW PROSES pada tools yang telah disediakan oleh RapidMiner. Seperti tampilan berikut ini :

😆 nee plana - NgaMine 53801	Contraction Contraction				ALC: NO. OF CO., NO.
in fit house has her h					
3988 <i>9</i> P	A # M H H 1 2 2 0				
Constant () () () () Constant () () Constant () () Constant () () Constant () ()		4+0		1	anales -
 Surgert (h) Surgert	-		-	ngarinak ngtu	
I Aquaticas 8 d - Q Q Q = 9 300 1 520 1 9 100 Teambor 1				A character ages	Conner 1
	Protection () () () () The protection that Message	Fires	Loater	The root of	mate shiph multipleater
0	1		_	-	

Ok sobat, langkah pertama sudah sobat lalui sekarang kita kelangkah berikutnya, terlebih dahulu import data yang ingin sobat kelompokkan, disini saya menggunakan data tugas akhir mahasiswa untuk dikelompookkan berdasarkan bidang kompetensi. Oke lanjut, silahkan lakukan seperti tampilan dibawah ini : dan untuk pengetahuan sobat, kita melakukan import data ada beberapa tipe atau format file yang digunakan seperti CSV, Excell Sheet, XML disini sobat sesuaikan dengan format file data sobat karena data saya menggunakan format Excell makadari itu saya pilih import data menggunakan format Excell Sheet.



Ok, kalau sudah seperti pada tampilan diatas maka silahkan sobat arahkan pointer sobat kearah tempat dimana data uji coba sobat simpan.

	Construction (Sec.14)	
Classics 2 2 + 0	The second public parts in the input point data. Bay to Prease second the fits that attractive registers	Ginne : Brannes :
Process Careful (27)		
Constant Constant	Baltratis 1 For Series Six San Lacification B 10 (10 formation Forstation Society Soc	
Aussing States		There are a second as a second
	Call of one sporter the	_
	(100 Yearson in an	The tool operator when

Kemudian silahkan sobat tekan tombol next hingga sampai pada step terakhir seperti tampilan berikut ini, kemudian jangan lupa sobat beri nama pada data cluster sobat seperti yang saya lakukan disini saya beri nama "DATA_UJI_COBA" kemudian tekan tombol finish.

	C (en rept aged (he lat)	-
Constant () () () () () () () () () () () () ()	The solard parts out it must be also	
Comparation (Constant) Constant (Constant) Constant (Constant) Constant(Constant) Constant(Constant) Constant(Constant) Constant(Constant) Constant(Constant) Constant(Constant)	a granne oon	
Passature - 1 - 6		taller systyppeder
	gana (kata, ut., 1984) Lautan Asut Republic (201, 104, 008)	- Parm
	(allowed and the first of the f	The cost operator which The costor wood operator

Jika sobat sudah menekan tombol finis maka dari itu akan otomatis akan mengarahkan kita pada data yang sudah kita import tadi seperti tampilan berikut ini jika sobat menemukan tampilan seperti berikut ini itu artinya sobat berhasil melakukan import data pada RapidMiner dan jika belum silahkan dicoba lagi.

1 44		9 0	n 6	b II	8 N	¥ @				
T head		1 Bite		a freedory	Drin, Lui, D	0840				Cill Assessment of Street Barry
Outs new	Cant	ab Ven Or	No line ()	Advanced [he	ra () insut	abora .			B 4	
and the lot	-	an distantal	Arrent To	and a state of				Transferrer Street at		a Stangest
								and an other sets [In		0 18 00
	and states	1014	4444	1.000	1.811	A COLOR				A COLORADO NO.
	name and	UNITA .	18,750	1.740		10.010	1070			a a and the second
	100,007	1014	0.000	178	0.801	10.072	0.798			CATA US CORP.
	100308	2014	1017	1075	0.700	8773	0.001			
	1062011	2014	0.067	1875	0.123	4.756	4/98			
	1002042	2014	1	4400	1.778	1.004	100			
	1002043	2014	4454	4400	0.798	4.796	476			
	100,014	2014	14.047	1792	1.601	4.796	4467			
	100,014	2014	14.80.0	4.708	11754	4478	1007			
	1002017	2014	8.750	4/28	1410	4.757	0.794			
	1002018	2014	4.600	1.000	0.664	4.750	4/99			
2	1002048	2014	1403	4788	0.122	1.000	1479			
3	1003027	2014	8.7%0	4.77%	0.667	16790	4798			
4	1003029	2014	1.017	1.00	1.888	0.054	0.021			
5	1062504	2014	10.0117	1.762	0.035	10.000	4.000			
	10429102	2014	4.405	4.007	0.122	4.797	1.007			
		107.4	4400	0.792	1419	4.175	4798			3 Q

Oke sobat, proses load data sudah selesai sekarang kita lanjut ke proses pengelompokan atau clustering di RapidMiner:

Ikuti langkah berikut ini untuk hasil yang lebih optimal

- a. Langkah pertama yaitu kita memili **retrieve** sebagai tempat penyimpanan data cluster
- b. Langkah kedua pilih **K-Means** untuk proses clustering
- c. Langkah ketiga pilih **cluster distance performance** untuk tampilan output cluster Berikut tampilannya beserta koneksi

Sebelum melakukan koneksi lakukan terlebih dahulu load data pada database sobat yg telah sobat simpan sebelumnya. Atau dapat juga sobat tarik database sobat atau bahasa kasarnya seret kedalam lembar kerja sobat seperti tampilan berikut ini:



Kemudian silahkan sobat include point2 yang telah saya sebutkan diatas tadi. Seperti pada tampilan berikut ini:

🤹 - sec proces' - Repúblice 1333	House K in the second second		korth and
In fit home yes her it	•		
3 4 H H H #	A ■ ■ ■ ■ ■ ■ ■ ■ ■		
# Santa 1			Breese L
1 Gooden C			
0 0 Ausstant			Automotion Teams. Automation of the optimization. All
	American State 1		K Massa Augustine
-	Proceso Provinsion participation of the state of the stat	Tana Laudini Na. 🖗 Na part Na Antonio 🎬 Nativa (1671, 117,	The spectra performs and
			THE OWNER OF TAXABLE PARTY.

Dismping kanan atas ada input jumlah cluster (K) = 5, itu bisa sobat tentukan sendiri. Dan pastikan tidak ada error pada saat pengkoneksian antara database dengan model. Jika sobat telah sukses sama seperti gambar diatas makadari itu silahkan sobat tekan perinta RUN atau tekan tombol F11 pada keyboard sobat.

	- 11 6	19 0	A 4	1 b II	H 5)	草用	-				
		Anna Indiana	Patrian				E foarstelle	A D. A. S. Roy (Restored Del)	A MALOBACI IN		Breatrie + 12 + 11
2.04	Inte Clinic	Call free ()	na nee C	Address the	n Owie	atore .				0.4-	a Stangert und
transp	-	que, l'operar		aprie deltan					1000 Fiber (100 F 100)		+ B.04
days \$		charter .	- 100	-	ana.	artist.	1000.00	1000.00			T Bruce Reported Comm
	18	(hole.)	19421011	4401	1187.4	0.001	1	1		0	N To processo in
2		shader, 2	1002004	4.794	1.700	10.000	100	1.670			- I Della Los COMANNEL CO
b		stanter, 2	-942467	4403	11718	4441	1.772	10.700			· · · · · · · · · · · · · · · · · · ·
	4	-theiler J	100305	4447	1.875	1.761	8.75	1421			
6		shade, 2	1002971	4887	1.675	4.167	41745	18.0%			
		shades, J	100010		1410	47%	100	4.607			
r		-tester_1	100310	0.023	1475	4.788	0.708	11794			
		charatter, A	100204	1017	1.762	1.001	1.76	1.857			
÷		shader, 2	1002118	1403	11.718	4.798	11010	14.857			
1		Outlet, 2	100817	8798	1.728	4404	8.727	18794			
		14000.7	100219-8	1457	1100	0.004	41758	18 1750			
4	-	shoter, 2	100,000	4453	1.734	4.122	1008	4.679			
		shades _2	-	4.794	1.44	4467	8/708	4.754			
14		nater, 2	-9125.0	1000	1.00	1.000	1004	180			
		mater, 2	1010	2443	1.76	416	100	1.000			
		date J	-	1403	1.07	1.00	1.10	1.00		0	()
100	a. 6. 16 a.	A									Martin Barter A. L. & H.
Q											
			-								

Tampilan diatas merupakan cluster masing-masing anggota dan dibawah ini merupakan hasil grafik pengelompokan. Silahkan sobat perlajari dan kembangkan sendiri.



Sumber : http://cuex-me.blogspot.com/2015/12/tutorial-rapidminer-menggunakan-metode.html

TUGAS 06

1. Model Prediksi Ini Menggunakan Tools RapidMiner

RapidMiner memiliki keunggulan tersendiri, RapidMiner merupakan aplikasi data mining berbasis sistem open-source dunia yang terkemuka dan ternama. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai data mining. Solusi yang diusung antara lain Integrasi data, analisis ETL, analisis data dan pelaporan dalam satu suite tunggal. Memiliki antarmuka pengguna grafis yang intuitif untuk desain analisis proses, serta fleksibel terdapat banyak transformasi data, pemodelan data, dan metode visualisasi data.

2. Model Prediksi Elektabilitas Caleg

a. Modeling Decision Tree

Decision tree adalah model prediksi yang digambarkan dengan struktur pohon atau struktur berhirarki. Konsep dari decision tree adalah membantu user dalam mengubah data menjadi pohon keputusan dan syarat-syarat keputusan agar lebih mudah dipahami dan dipetakan.



Deskripsi :

- Jika calon memiliki suara sah lebih dari 3.082 maka akan terpilih menjadi anggota legislatif. Jika kurang dari 3082 akan dipertimbangkan dalam dua klasifikasi yaitu jumlah suara >2418 dan ≤ 2418
- jika suara sah caleg ≤2418 akan dibagi lagi ke dalam kelompok suara sah caleg jika memiliki ≤ 1436.500 maka caleg tidak akan terpilih jika jumlah suara >1436.500 dan terdapat >1468 suara maka tidak akan terpilih dan jika ≤1468 maka akan terpilih menjadi anggota legislatif
- Dalam jumlah suara sebanyak >2418 terbagi menjadi dua klasifikasi yaitu apabila suara yang dimiliki >2935 maka caleg tidak akan terpilih. Dan jika ≤2935 akan ditentukan oleh suara sah partai dan daerah pemilihan untuk menentukan terpilih atau tidaknya menjadi anggota legislatif
- Jika memiliki suara sah partai berjumlah ≤ 16443 maka tidak akan terpilih menjadi anggota legislative. Jika memiliki suara sah partai >16443 maka akan ditentukan dari daerah pemilihannya.
- Jika daerah pemilih memiliki suara ≤1.500 maka caleg tersebut akan terpilih menjadi legislatif
- Kemudian jika memiliki suara daerah pemilihan >1.500 akan ditentukan dari suara sah partai.
- Jika suara sah partai >83822 maka tidak akan terpilih menjadi anggota legislatif. Dan jika suara sah partai ≤83822 akan ditinjau kembali dari suara sah caleg.
- Jika suara sah caleg >2445.500 maka ia akan terpilih menjadi anggota legislatif, dan jika ≤2445.500 maka tidak akan terpilih.

	true TIDAK	true YA	class precision
pred. TIDAK	362	14	96.28%
pred. YA	15	34	69.39%
class recall	96.02%	70.83%	

accuracy: 93.16% +/- 4.45% (mikro: 93.18%)

AUC: 0.915 +/- 0.076 (mikro: 0.915) (positive class: YA)



b. Modeling Naïve Bayes

Naive Bayes merupakan suatu klasifikasi berpeluang sederhana berdasarkan teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen).

SimpleDistribution

Distribution model for label attribute TERPILIH ATAU TIDAK

```
Class TIDAK (0.887)
10 distributions
Class YA (0.113)
10 distributions
```

	true TIDAK	true YA	class precision
pred. TIDAK	326	18	94.77%
pred. YA	51	30	37.04%
class recall	86.47%	62.50%	

accuracy: 83.73% +/- 4.2	5% (mikro: 83.76%)



c. Modeling K-nn

K-Neares Neighbour (K-NN) adalah sebuah metde klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

KNNRegression

```
1-Nearest Neighbour model for regression.
The model contains 425 examples with 10 dimensions.
```

Berdasarkan pernyataan tersebut dapat diartikan bahwa dari data yang dianalisis terdapat 425 data dengan 10 dimensi (atribut) untuk menentukan target "ya" dan "tidak".

	true TIDAK	true YA	class precision
pred. TIDAK	358	25	93.47%
pred. YA	19	23	54.76%
class recall	94.96%	47.92%	

ассштаст	r 89 63% +/- 2	21% (mikro	89 65%)
accurac	. 03.03/0 -1-2		. 03.03/0

AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: YA)



3. Evaluasi dan Akurasi

	Decision Tree (C4.5)	Native Bayers (NB)	K-Nearest Neighbor(K-NN)
Akurasi	93.16 %	83.73 %	89.63 %
Minus	4.45 %	4.25 %	2.21 %
AUC	0.915	0.840	0.500

Kesimpulan :

Berdasarkan tabel diatas, Model *Decision Tree* memiliki akurasi tertinggi dengan nilai sebesar 93.16 % diikuti K-nn dengan nilai 89.63 % dan pada urutan terakhir *Naïve Bayes* dengan akurasi paling rendah sebsar 83.73 %. Dari hasil tersebut dapat ditarik kesimpulan bahwa modeling yang paling baik digunakan untuk dataset tersebut yaitu *Decision Tree* model , karena memiliki akurasi prediksi paling besar dengan nilai presentase sebesar 93.16 % diantara ketiga model, dapat diartikan analisis yang dihasilkan dari Model *Decision Tree* memiliki kemungkinan menjadi kenyataan paling besar dan tepat.
Untuk melakukan validasi terhadap *Tree* yang dihasilkan, digunakan nilai dari area dibawah kurva ROC (AUC/*Area Under Curve*). Menurut Gorunescu (2011), hasil perhitungan yang divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu :

a. Akurasi bernilai 0.90 – 1.00 = excellent classification

- b. Akurasi bernilai $0.80 0.90 = good \ classification$
- c. Akurasi bernilai 0.70 0.80 = fair classification
- d. Akurasi bernilai $0.60 0.70 = poor \ classification$
- *e*. Akurasi bernilai 0.50 0.60 = failure

Berdasarkan tabel tersebut juga dipresentasikan bahwa dengan menggunakan "Decision

Tree, AUC/ Area Under Currve yang dihasilkan lebih besar yaitu 0.91 > 0.90 masuk dalam kategori *excellent classification*", sedangkan untuk *Native Bayers* masuk dalam kategori *good classification*, dan *K-nearest* ada pada kategori *failure*.

Dapat disimpulkan bahwasannya dengan dataset pemilu apabila ingin melakukan pengklasifikasian data dengan hasil yang paling akurat dan baik maka disarankan untuk menggunakan model *Decision Tree*.

Sumber rujukan: <u>https://www.youtube.com/watch?v=VkcgwiSwb10</u>

Nama : Andry Meylani NIM : 202420009 TUGAS 06 Sumber : Belajar Data Mining dengan Rapid Miner Oleh Dennis Aprillia C, dkk.

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban :

Pada kasus ini saya mengambil pengolahan data terkait dengan Market Basket yang akan dicoba pada tool rapidminer. Adapun tahapannya sebagai berikut :

TID	MERAH	KUNING	HIJAU	BIRU	UNGU	JINGGA	NILA
1	1	1	1	0	0	0	0
2	0	1	1	1	0	0	0
3	0	0	0	1	1	1	0
4	0	1	1	0	0	0	0
5	0	1	1	1	0	1	1

1. Create tabel master data di Excel kemudian export ke Rapidminer.

Berikut hasil export data ke RapidMiner :

🧶 <new process=""> –</new>	- RapidMiner Studio	Professional 7.1.001	@ LAPTOP-UIBHC	010					
<u>File Edit Proces</u>	ss <u>V</u> iew <u>C</u> onnec	tions Clou <u>d S</u> e	ttings E <u>x</u> tension	S					
	-		•		Views:	Design	Results]	
Result History	× 📒 Exa	mpleSet (//Local	Repository/Tria	l RapidMiner)	×				
-	ExampleSet (5 exa	amples, O special a	attributes, 8 regula	r attributes)			Filter (5 / 5 e	examples): all	•
Data	Row No.	TID	MERAH	KUNING	HIJAU	BIRU	UNGU	JINGGA	NILA
Data	1	1	1	1	1	0	0	0	0
_	2	2	0	1	1	1	0	0	0
Σ	3	3	0	0	0	1	1	1	0
Statistics	4	4	0	1	1	0	0	0	0
	5	5	0	1	1	1	0	1	1



2. Create Modul Association Rule, FP-Growth, dan Numerical to Binomial, sesuaikan dengan parameter yang ada.

Numerical to Binomial.



FP-Growth





3. Create hubungan antar modul mulai dari data master sampai dengan Association Rule.



Antar modul dihubungkan dan lakukan RUN pada system. Setelah beberapa detik akan muncul sebuah tab Association Rules yang baru seperti pada gambar di bawah ini :

Result History	🗙 🛒 AssociationRules (Crea	ate Assoc	iation Rules) ×		Repository ×	
	Show rules matching	No.	Premises	Conclus	G Add Data	= •
Data	all of these conclusions:	137	ROTI, MENTEGA, TELUR, SUSU	KECAP A	▼ 🛅 Samples	
Data	ROTI	138	KECAP	ROTI, ME	🕨 🧱 data	
	MENTEGA TELUR	139	ROTI, KECAP	MENTEG	processes	
A	SUSU	140	MENTEGA, KECAP	ROTI, TE	 Tutorials 	
Graph	BUNCIS	141	ROTI, MENTEGA, KECAP	TELUR,	DB	
		142	TELUR, KECAP	ROTI, ME	 Local Repository (bhijanta.wija) 	ramukti)
		143	ROTI, TELUR, KECAP	MENTEG	 Gata (chijanta.wijayamutti) Tugas 06 (chijanta.wijaya 	imukti)
Description		144	MENTEGA, TELUR, KECAP	ROTI, SL	Tugas 06 (bhijanta.wi	jayamukti - v1
		145	ROTI, MENTEGA, TELUR, KECAP	SUSU	Rapidminer_Tugas	06 (bhijanta.v
		146	SUSU, KECAP	ROTI, ME	Tugas 06_market ba	asker analys
Annotations		147	ROTI, SUSU, KECAP	MENTEG	Tugas 06_market ba	asket analys
		148	MENTEGA, SUSU, KECAP	ROTI, TE	Tugas 06_market ba	asker analys
		149	ROTI, MENTEGA, SUSU, KECAP	TELUR	Fugas of processes (bhijanta.wijayam)	ukti)
	Min. Criterion:	150	TELUR, SUSU, KECAP	ROTI, ME	Cloud Repository (disconnected	d)
	confidence •	151	ROTI, TELUR, SUSU, KECAP	MENTEG		
	Min. Criterion Value:	152	MENTEGA, TELUR, SUSU, KECAP	ROTI		
	-	<	1	>	<	>

Data yang dihasilkan sebanyak 152 dan iniakan menyulitkan untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Langkah yang diambil adalah mengubah nilai minimum support dan minimum confidence.

Maka kita harus mengubah nilai di FP-Growth dan Create Association Rules untuk dapat mencapai nilai minimum support dan minimum confidence, untuk dapat menghasilkan skema yg lebih sedikit untuk dapat mempermudah dalam penyusunan skema market.

FP-Growth -







Assosiation Rules _





AFTER (0.95)

Setelah dilakukan perubahan parameter lakukan RUN kembali, Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

matching		No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
conclusion	s: 🔻	1	ROTI	MENTEGA	0.800	1	1	-0.800
		2	MENTEGA	ROTI	0.800	1	1	-0.800
		3	SUSU	TELUR	0.400	1	1	-0.400
		4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400
		5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400
sult History	X 🛒 Ass Zoom	ociation	Rules (Create Association Rules)	X				
ult History Data	X Som	ociation p p H	Rules (Create Association Rules)	X	Rule 5 (0.40)/1.000)		Rule 2 (0.800 / 1.000)
Data	X Soom	ociation) A 4	Rules (Create Association Rules)	X Rule 3 (0 400 / 1.000)	Rule 5 (0.40	0/(1000)		Rule 2 (0.800 / 1.000)

TUGAS 06 ADVANCED DATABASE KELAS MTI 23 A



Di Susun Oleh : Ari Hardiyantoro Susanto NIM : 202420015

Dosen Pengasuh : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

Program Pasca Sarjana Universitas Bina Darma Palembang

2020/2021

Soal :

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban :

Pada kasus ini saya mengambil pengolahan data terkait dengan Eksplorasi Data Pemilu yang akan dicoba pada tool rapidminer. Data pemilu yang di analisis adalah sebagai berikut :

到。	latapemi	lukpu																	-
1	А	В	C	D	E	F	G	н	1	J	K	L	M	N	0	р	Q	R	S
329 330	PARTAI KE	HJ.SITI UL	P	LEBAKSIU	34	14308	0	6	3	36	TIDAK								
331	PARTAI PE	ULFATUZ	P	ADIWERN	42	634	0	2	1	49	TIDAK								
333	PARTAI PE	YUSUF RIF	L	TALANG	42	634	0	5	1	368	TIDAK								
334 335	PARTAI BI	ADIKIN	L	PANGKAH	44	11003	1	2	1	559	TIDAK								
337 337	PARTAI BI	DAHURI	L	KRAMAT	44	11003	1	6	1	410	TIDAK								
339	PARTAI BI	ADY KUNT	L.	SURODAD	44	11003	1	6	2	4680	YA								
341	PARTAI BU	NIASRINI	P	KRAMAT	44	11003	1	6	3	123	TIDAK								
343	PARTAI BI	YUDA KUR	RL.	KRAMAT	44	11003	1	6	4	402	TIDAK								
345	PARTAI BU	WAHYUDI	L	KRAMAT	44	11003	1	6	5	3065	TIDAK								
347 I	PARTAI BI	SITI JURIY	P	KRAMAT	44	11003	1	6	6	72	TIDAK								
349 I	PARTAI BI	МОНАМА	L	KRAMAT	44	11003	1	6	7	52	TIDAK								
351	PARTAI BI	MEGENG	L	KRAMAT	44	11003	1	6	8	397	TIDAK								

1. Decision Tree (C4.5)

🗙 % Perfori	manceVector (Performance)	🗙 🚦 ExampleSet (Set Role	e) 🛛 🗙 🤰 Tree (Deci	sion Tree) 🛛 🗙
Criterion	Table View O Plot View			
accuracy				
AUC	accuracy: 93.16% +/- 4.45%	(mikro: 93.18%)		
		true TIDAK	true YA	class precision
	pred. TIDAK	362	14	96.28%
	pred. YA	15	34	69.39%
	class recall	96.02%	70.83%	

• Accuracy

Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran Decission Tree didapatkan hasil dengan tingkat akurasi 93.16% yang artinya tingkat akurasi data sudah sangat baik.Class precission didapat dengan hasil prediksi tidak sebesar 96.28% dan prediksi Ya sebesar 69.39%.

Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, jadi kecocokan antara permintaan dan prediksi tidak adalah 96.28% dan kecocokan dengan prediksi ya adalah 69.39%.

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Jadi dari data diatas dapat disimpulkan bahwa tingkat relevan true Tidak sebesar 96.02% dan True Ya sebesar 70.83%.



• AUC

AUC (the area under curve) dihitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan false positive sebagai garis horizontal dan true positive sebagai garis vertikal. Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran Decission Tree didapatkan hasil AUC 0.915 yang termasuk dalam kategori sangat baik (Excellent Classification) dilihat dari kategori dibawah ini :

- 1. 90 1.00 = Excellent Classification
- 2. 80 0.90 = Good Classification
- 3. 70 0.80 = Fair Classification
- 4. 60 0.70 = Poor Classification
- 5. 50 0.60 = Failure

×	🎇 Performa	nceVector (Perfor	mance)	×	Example	eSet (Set Role)	×	Tree (Decision	Tree) ×	
Exam	oleSet (425 examp	les, 1 special attrib	oute, 9 re	egular attributes)				Filter (42	25 / 425 examples):	all
	TERPILIH AT	NAMA PART	JE	KECAMATAN	NO	SUARA SAH	JUML	DAERAH PE	NO.URUT CA	SUARA SAH
1	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	1	1	594
2	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	1	2	943
3	TIDAK	HATI NURANI	Р	SLAWI	1	18578	1	1	3	1730
4	YA	HATI NURANI	L	DUKUHWARU	1	18578	1	1	4	2508
5	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	2	1	923
6	TIDAK	HATI NURANI	Р	TARUB	1	18578	1	2	2	308
7	TIDAK	HATI NURANI	L	TARUB	1	18578	1	2	3	54
8	TIDAK	HATI NURANI	L	BOJONG	1	18578	1	3	1	1682
9	TIDAK	HATI NURANI	Р	JATINEGARA	1	18578	1	3	2	918
10	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	3	3	87
11	TIDAK	HATI NURANI	L	BALAPULANG	1	18578	1	4	1	728
12	TIDAK	HATI NURANI	L	MARGASARI	1	18578	1	4	2	346
13	TIDAK	HATI NURANI	Р	LEBAKSIU	1	18578	1	4	3	184
14	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	5	1	381
15	TIDAK	HATI NURANI	L	ADIWERNA	1	18578	1	5	2	148
16	TIDAK	HATI NURANI	Р	SLAWI	1	18578	1	5	3	128

• Set Role

Dari hasil analisis diatas, kita dapat melihat bahwa terdapat 10 kolom atribut yang terdiri atas 9 kolom atribut biasa dan 1 kolom spesial atribut dengan nama "TERPILIH ATAU TIDAK", hal ini dikarenakan kolom tersebut menjadi penentu dari hasil analisis data pemilu ini.



Dari hasil analisis diatas dapat ditarik kesimpulan, bahwa :

- Jika suara sah caleg lebih dari 3082 maka Terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082, suara sah caleg kurang dari atau sama dengan 2418 dan kurang dari sama dengan 1436500 maka tidak terpilih.
- ika suara sah caleg kurang dari atau sama dengan 3082, suara sah caleg kurang dari atau sama dengan 2418, lebih dari 1436500 dan lebih dari 1465 maka tidak terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082, suara sah caleg kurang dari atau sama dengan 2418, lebih dari 1436500 dan kurang dari atau sama dengan 1465 maka tidak terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082, suara sah caleg lebih dari 2418 dan lebih dari 2935 maka tidak terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082 & dari atau sama dengan 2418 serta suara sah partai kurang dari atau sama dengan 14663 maka tidak terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082 & dari atau sama dengan 2418 serta suara sah partai lebih dari 14663 dengan daerah pemilihan kurang dari atau sama dengan 1500 maka akan terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082 & dari atau sama dengan 2418 serta suara sah partai lebih dari 14663 dengan daerah pemilihan lebih dari 1500 dengansuara sah partai lebih dari 83822 maka akan tidak terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082 & dari atau sama dengan 2418 serta suara sah partai lebih dari 14663 dengan daerah pemilihan lebih dari 1500 dengansuara sah partai kurang dari atau sama dengan 83822 dan suara sah caleg lebih dari 2445500 maka akan terpilih.
- Jika suara sah caleg kurang dari atau sama dengan 3082 & dari atau sama dengan 2418 serta suara sah partai lebih dari 14663 dengan daerah pemilihan lebih dari 1500 dengansuara sah partai kurang dari atau sama dengan 83822 dan suara sah caleg kurang dari atau sama dengan 2445500 maka akan tidak terpilih.
- 2. Naive Bayes (NB)

ector (Performance)	ExampleSet (Set Role)	🗙 💡 SimpleDistribi	ution (Naive Bayes) 🛛 🛛 🕹
Table View O Plot View			
accuracy: 89.14% +/- 4.79%	(mikro: 89.18%)		
	true TIDAK	true YA	class precision
pred. TIDAK	357	26	93.21%
		00	50.00%
pred. YA	20	22	52.36%

• Accuracy

Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran Naive Bayes didapatkan hasil dengan tingkat akurasi 89.14% yang artinya tingkat akurasi data sudah baik.

Class precission didapat dengan hasil prediksi tidak sebesar 93.21% dan prediksi Ya sebesar 52.38%. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, jadi kecocokan antara permintaan dan prediksi tidak adalah 93.21% dan kecocokan dengan prediksi ya adalah 52.38%.

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Jadi dari data diatas dapat disimpulkan bahwa tingkat relevan true Tidak sebesar 94.69% dan True Ya sebesar 45.83%.Tingkat akurasi data ketika analisis menggunakan Bayes lebih kecil bila dibandingkan ketika menggunakan K-NN atau Decission Tree.



• AUC

AUC (the area under curve) dihitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan false positive sebagai garis horizontal dan true positive sebagai garis vertikal. Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran K-NN didapatkan hasil AUC 0.838 yang termasuk dalam kategori baik (Good Classification) jika dilihat dari kategori dibawah ini :

- 1. 90 1.00 = Excellent Classification
- 2. 80 0.90 = Good Classification

- 3. 70 0.80 = Fair Classification
- 4. 60 0.70 = Poor Classification
- 5. 50 0.60 = Failure

Dapat dikatakan bahwa analisis menggunakan kriteria AUC pada Bayes lebih besar atau jauh lebih baik dibandingkan ketika menggunakan K-NN. Namun menggunakan kriteria AUC pada Bayes masih belum sebaik ketika menggunakan Decission Tree.

× %	PerformanceVector	(Performance)	🛛 📕 Ex	ampleSet (Set R	ole) × 📢	SimpleDistribu	ution (Naive Bayes) ×	
ExampleSet (4	25 examples, 1 speci	ial attribute, 9 regul	ar attributes)			Filter (42	25 / 425 examples):	all	,
Row No.	TERPILIH AT	NAMA PART	JENIS KELA	KECAMATAN	NO.URUT PA	SUARA SAH	JUML.PERO	DAERAH PE	NO.L
1	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	1	1
2	TIDAK	HATI NURANI	L.	SLAWI	1	18578	1	1	2
3	TIDAK	HATI NURANI	Р	SLAWI	1	18578	1	1	3
4	YA	HATI NURANI	L	DUKUHWARU	1	18578	1	1	4
5	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	2	1
6	TIDAK	HATI NURANI	Р	TARUB	1	18578	1	2	2
7	TIDAK	HATI NURANI	L	TARUB	1	18578	1	2	3
8	TIDAK	HATI NURANI	L	BOJONG	1	18578	1	3	1
9	TIDAK	HATI NURANI	Р	JATINEGARA	1	18578	1	3	2
10	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	3	3
11	TIDAK	HATI NURANI	L	BALAPULANG	1	18578	1	4	1
12	TIDAK	HATI NURANI	L	MARGASARI	1	18578	1	4	2
13	TIDAK	HATI NURANI	Р	LEBAKSIU	1	18578	1	4	3
14	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	5	1
15	TIDAK	HATI NURANI	L	ADIWERNA	1	18578	1	5	2
16	TIDAK	HATI NURANI	Р	SLAWI	1	18578	1	5	3

• Set Role

Dari hasil analisis diatas, kita dapat melihat bahwa terdapat 10 kolom atribut yang terdiri atas 9 kolom atribut biasa dan 1 kolom spesial atribut dengan nama "TERPILIH ATAU TIDAK", hal ini dikarenakan kolom tersebut menjadi penentu dari hasil analisis data pemilu ini. Dalam analisis menggunakan Naive Bayes pada Set Role tidak ada perbedaan ketika analisisnya menggunakan Decission Tree maupun menggunakan K-NN.

X	8 PerformanceVector (Performance)	×		ExampleSet (Set Role)	×	9	SimpleDistribution (Naive Bayes)	×	
Sim	pleDistribution								
Distr:	ibution model for label attribut	e TERP	ILIH	H ATAU TIDAK					
Class 9 dist	TIDAK (0.887) tributions								

Class YA (0.113) 9 distributions • Simple Distribution

Dari hasil analisis data di atas dapat diambil kesimpulan bahwa dari data pemilu terdapat dua label "TERPILIH ATAU TIDAK" yang akan dibagi menjadi dua Class yaitu Class TIDAK dengan nilai 0.887 dan Class Ya dengan nilai 0.113.

Class	Nilai
Tidak	0.887
Ya	0.113

3. K-Nearest Neighbor (k-NN)

Criterion	Table View O Plot	/iew		
accuracy				
AUC	accuracy: 89.63% +/- 2.	21% (mikro: 89.65%)		
		true TIDAK	true YA	class precision
	pred. TIDAK	358	25	93.47%
			02	54 76%
	pred. YA	19	23	54.70%

• Accuracy

Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran K-NN didapatkan hasil dengan tingkat akurasi 89.63% yang artinya tingkat akurasi data sudah baik.

Class precission didapat dengan hasil prediksi tidak sebesar 93.47% dan prediksi Ya sebesar 54.76%. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, jadi kecocokan antara permintaan dan prediksi tidak adalah 93.47% dan kecocokan dengan prediksi ya adalah 54.76%.

Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. Jadi dari data diatas dapat disimpulkan bahwa tingkat relevan true Tidak sebesar 94.95% dan True Ya sebesar 47.92%. Tingkat akurasi data ketika analisis menggunakan K-NN kecil bila dibandingkan ketika menggunakan Decission Tree.



• AUC

AUC (the area under curve) dihitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan false positive sebagai garis horizontal dan true positive sebagai garis vertikal. Dari data diatas analisis menggunakan aplikasi Rapid Miner dengan pengukuran K-NN didapatkan hasil AUC 0.500 yang termasuk dalam kategori gagal (Failure) jika dilihat dari kategori dibawah ini :

- 1. 90 1.00 = Excellent Classification
- 2. 80 0.90 = Good Classification
- 3. 70 0.80 = Fair Classification
- 4. 60 0.70 = Poor Classification
- 5. 50 0.60 = Failure

Dapat dikatakan analisis menggunakan kriteria AUC pada K-NN masih belum sebaik ketika menggunakan Decission Tree dan perbandingan nilainya sangat jauh lebih baik ketika menggunakan Decission Tree.

×	% Performance	eVector (Perform	ance)	🗙 📕 Exam	npleSet (Se	t Role) 🛛 🗙	💡 KNNC	lassification	(k-NN) ×	
Example	eSet (425 example	s, 1 special attribut	te, 9 regular	attributes)				Filter (425 / 4	25 examples):	dl
Ro	TERPILIH AT	NAMA PART	JENIS	KECAMATAN	NO.UR	SUARA SAH	JUML.P	DAERA	NO.URUT CA	SUARA SAH
1	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	1	1	594
2	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	1	2	943
3	TIDAK	HATI NURANI	P	SLAWI	1	18578	1	1	3	1730
4	YA	HATI NURANI	L	DUKUHWARU	1	18578	1	1	4	2508
5	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	2	1	923
6	TIDAK	HATI NURANI	P	TARUB	1	18578	1	2	2	308
7	TIDAK	HATI NURANI	L	TARUB	1	18578	1	2	3	54
8	TIDAK	HATI NURANI	L	BOJONG	1	18578	1	3	1	1682
9	TIDAK	HATI NURANI	Р	JATINEGARA	1	18578	1	3	2	918
10	TIDAK	HATI NURANI	L	SLAWI	1	18578	1	3	3	87
11	TIDAK	HATI NURANI	Ĺ	BALAPULANG	1	18578	1	4	1	728
12	TIDAK	HATI NURANI	L	MARGASARI	1	18578	1	4	2	346
13	TIDAK	HATI NURANI	P	LEBAKSIU	1	18578	1	4	3	184
14	TIDAK	HATI NURANI	L	LEBAKSIU	1	18578	1	5	1	381
15	TIDAK	HATI NURANI	L	ADIWERNA	1	18578	1	5	2	148
16	TIDAK	HATI NURANI	P	SLAWI	1	18578	1	5	3	128

• Set Role

Dari hasil analisis diatas, kita dapat melihat bahwa terdapat 10 kolom atribut yang terdiri atas 9 kolom atribut biasa dan 1 kolom spesial atribut dengan nama "TERPILIH ATAU TIDAK", hal ini dikarenakan kolom tersebut menjadi penentu dari hasil analisis data pemilu ini. Dalam analisis pada Set Role tidak ada perbedaan ketika analisisnya menggunakan Decission Tree.



```
The model contains 425 examples with 9 dimensions of the following classes:
TIDAK
YA
```

• K-NN Classification

Dari hasil analisis di atas dapat dilihat bahwa dari data yang telah dianalisis terdapat 1 klasifikasi yang spesial (nearest Neighbour) dan juga terdapat 9 dimensi dari kelaskelas yang terdiri dari "TIDAK" dan "YA". Dari analisis diatas juga disebutkan bahwa dari data yang telah dianalisis terdapat 425 data.

EVALUATION

	Decission Tree	K-Nearest Neighbor (K-NN)	Naive Bayes
Accuration	93.16	89.63	89.14
AUC	0.915	0.5	0.838

Jadi dari tabel diatas dapat disimpulkan bahwa nilai Accuration dan AUC dari Decission Tree lebih baik dibandingkan yang lainnya. Jika dilihat dari Accuration, perbedaan dari ketiga metode tidak terlalu jauh dimana nilai yang paling kecil adalah Naive Bayes 89.14 disusul K-Nearest Neighbor sebesar 89.63 dan nilai yang paling besar/baik adalah Decisison Tree.

Namun ketika dilihat dari AUC, terdapat perbedaan nilai yang cukup jauh antara menggunakan metode K-Nearest Neighbor dibanding dengan Naive bayes dan Decission Tree. AUC K-NN hanya 0.5 yang didefinisikan sebagai analisis yang gagal (failure), sedangkan Naive Bayes memiliki nilai 0.838 yang didefinisikan sebagai analisis yang baik (Good) dan Decission Tree yang memiliki nilai 0.915 yang didefinisikan sebagai analisis sangat baik (Excellent).

Pengolahan data menggunakan metode prediksi Elektabilitas Calon Legislatif

RapidMiner Studio merupakan perangkat lunak yang bersifat terbuka (open source) untuk melakukan analisis terhadap data mining, text mining, dan analisis prediksi. RapidMiner Studio memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing, dan data visualization. RapidMiner Studio dipilih dalam melakukan pengalisisan model prediksi elektabilitas calon legislatif menggunakan data pemilu sebagai data training karena memiliki tampilan antarmuka yang paling ramah pengguna.

C4.5 (Decision Tree)

Pembelajaran pohon keputusan adalah metode yang umum digunakan dalam data mining. Tujuannya adalah untuk menciptakan sebuah model yang memprediksi nilai variabel target berdasarkan beberapa variabel masukan. Setiap simpul interior sesuai dengan salah satu variabel masukan, ada tepi untuk anak-anak untuk setiap nilai yang mungkin dari variabel masukan. Setiap daun merupakan nilai dari variabel target yang diberikan nilai-nilai variabel input diwakili oleh jalan dari akar ke daun. Dalam data mining , pohon keputusan dapat digambarkan juga sebagai kombinasi teknik matematika dan komputasi untuk membantu deskripsi, kategorisasi dan generalisasi dari himpunan data.

Langkah – langkah menggunakan algoritma C4.5 pada RapidMiner Studio yaitu:

- Buka aplikasi RapidMiner Studio
- Pilih New Process, Blank



Pada Operators, pilih Data Access, Files, Read, Read CSV (double click)



 Pada Parameters, pilih Import Configuration Wizard, pilih data pemilu (datapemilukpu.csv), Column Separation – Semicolon, Next, Next, Unchecklist Nama Calon, ganti Attribute menjadi Label pada kolom Sah atau Tidak Sah, Finish.

🏏 Import C	onfiguration Wizard	Ð	^
csv file		D	
column separators	ŕ	D	1
trim lines		٢	
🖌 use quotes		٢	
quotes character		٢	

Subco	nimerius	#			Escape C	haracter:	1			
					Use (Quotes	-			
NAMA PA	NAMA C	JENIS K	KECAMA	NO.URU	SUARA	JUML PE	DAERAH	NO.URU	SUARA	1
HATI NU_	TOTO S	L	LEBAKSIU	1	18578	1	1	1	594	1
HATI NU_	EDI PUR	L	SLAW1	1	18578	1	1	2	943	
HATI NU	ELI RET	Р	SLAWI	1	18578	1	1	3	1730	
HATI NU_	SAHYUD	L	DUKUH	1	18578	1	1	4	2508	
HATINU	H.FAJAR	L	SLAWI	1	18578	1	2	1	923	
		-	Jernin				-			
-			Vicial		1 1000000000000			and the second s		P
Row, Colum	n	Err	or		Original v	alue	1	Message		

a import wizar	rd - Step 4 of 4												×
This v Step 4 Furthe opera	vizard guides 4: RapidMiner ermore, Rapid tors. These ro	you to import Studio uses s Miner Studio oles can be alt	your data. strongly typed assigns roles so defined he	attributes to the att re. Finall	s. In ti tribute y, you	his step, yo es, definin i can renan	ou c g wi ne a	an define the hat they can b attributes or d	data types e used for eselect the	of y by t m e	your attribut he individua antirely.	as. I	
Reload data	Po Gues	s value types	Date form	at Ente	w yak			,					
wiew uses on	ly first 100 row	vs.											
	1	1	1	-		1			4		1		1
NAMA CALO	JENIS KELA	KECAMATAN	NO.URUT P	SUARA	SAH	JUML.PE	RO	DAERAH PE	NO.URU	rc	SUARA SA	H	ATAU 1
polyno 🔻	polyno 🔻	polyno 🔻	integer 🔻	integer	*	integer	٠	integer 🔻	integer	٠	integer	•	polync
attribute *	attribute *	attribute *	attribute *	attribute	e •	attribute	*	attribute *	attribute	۲	attribute		label
TOTO S	L	LEBAKSIU	1	18578		1		1	1		594		TIDA! ^
EDI PUR	L	SLAWI	1	18578		1		1	2		943		TIDAŁ
ELI RET	P	SLAWI	1	18578		1		1	3		1730		TIDAF
SAHYUD	L	DUKUH	1	18578		1		1	4		2508		YA
H.FAJAR	L	SLAWI	1	18578		1		2	1		923		
TOFS.				1				V	Ignore en	ors	Show	on	> ly <u>e</u> rrors
Column		Error			Ori	ginal value			Mess	age	•		
									De la				Stand
	a import wizar Step Furthi opera Beload data eview uses on NAMA CALO polyno	a import wizard - Step 4 of 4 This wizard guides Step 4: RapidMiner Furthermore, Rapid operators. These ro Reload data Wew uses only first 100 row NAMA CALO JENIS KELA polyno	A import wizard - Step 4 of 4 This wizard guides you to import is Step 4: RapidMiner Studio uses is Furthermore, RapidMiner Studio uses is Geload data Colored	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed Furthermore, RapidMiner Studio uses strongly typed Furthermore, RapidMiner Studio uses offined he Reload data Reload data Image: Guess value types Date form Reload data Image: Guess value types Date form NAMA CALO JENIS KELA KECAMATAP NO.URUT P polyno* polyno* Integer * attribute attribute * attribute * TOTO S L LEBAKSIU 1 ELI RET P SLAWI 1 SAHYUD L DUKUH 1 HFAJAR L SLAWI 1 Toros Error Toros 1	A import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes Furthermore, RapidMiner Studio assigns roles to the at operators. These roles can be also defined here. Finall Reload data Reload data Image: Colored to the strongly typed attributes	A import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In t Furthermore, RapidMiner Studio assigns roles to the attribute operators. These roles can be also defined here. Finally, you Reload data Reload data Image: Guess value type Date format Enterval Reload data Image: Guess value type Date format Enterval NAMA CALO JENIS KELA KECAMATAP NO.URUT P SUARA SAH polyno polyno polyno integer integer attribute attribute attribute attribute integer TOTO S L LEBAKSIU 1 18578 ELI RET P SLAWI 1 18578 SAHYUD L DUKUH 1 18578 KHFAJAR L SLAWI 1 18578 Column Error Ori	A import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In this step, yo Furthermore, RapidMiner Studio assigns roles to the attributes, defining operators. These roles can be also defined here. Finally, you can renar Reload data Reload data Image: Guess value type Date format Ender value Reload data Image: Guess value type Date format Ender value NAMA CALO JENIS KELA KECAMATAP NO.URUT P SUARA SAH JUML PE polyno polyno polyno integer integer integer attribute attribute attribute attribute attribute attribute TOTO S L LEBAKSIU 1 18578 1 ELI RET P SLAWI 1 18578 1 SAHYUD L DUKUH 1 18578 1 Toros Error Original value	A import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you of Furthermore, RapidMiner Studio assigns roles to the attributes, defining will operators. These roles can be also defined here. Finally, you can rename a strong data Reload data Reload data Reload types Date format Enter value Reload data Reload data Reload types Date format Enter value NAMA CALO JENIS KELA KECAMATAP NO.URUT P SUARA SAH JUML PERO polyno polyno polyno polyno intribute attribute attribute intribute attribute attribute intribute attribute attribute into S L LEBAKSIU 1 EDI PUR L SLAWI 1 ELI RET P SLAWI 1 SAHYUD L DUKUH 1 INFAJAR L SLAWI 1 INFAJAR INFAJAR 1	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: Rapid/liner Studio uses strongly typed attributes. In this step, you can define the Furthermore, Rapid/liner Studio assigns roles to the attributes, defining what they can bo operators. These roles can be also defined here. Finally, you can rename attributes of Reload data Quess value types Date format Enter value NAMA CALO JENIS KELA KECAMATAP NO.URUT P SUARA SAH JUML.PERO DAERAH PE polyno. polyno	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly type attributes. In this step, you can define the data types Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for operators. These roles can be also defined here. Finally, you can rename attributes or deselect the Action data Counce Error Data format Data format Data format Data f	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of purphermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by to operators. These roles can be also defined here. Finally, you can rename attributes or deselect them attributes uses only first 100 rows. NAMA CALO JENIS KELA KECAMATA/ NO.URUT P. SUARA SAH JUML PERO DAERAH PE NO.URUT C polyno v integer	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely. Reload data	a import wizard - Step 4 of 4 This wizard guides you to import your data. Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. A guided data

Pada Operators, pilih Modeling, Predictive, Trees, Decision Tree (doble click)



• Hubungkan garis dari out pada Read CSV menuju tra pada Decision Tree.



- Pada Operators, pilih Validation, X-Validation (double click)
- Pada Parameters, number of validations = 10
- Hubungkan garis dari exa pada Decision Tree ke tra pada Validation.

Process ×								
Process		100% 🔎	Ø	P	4	2	ø	
Process								
Read CSV	Decision Tree	Validation						
Dinp 🖣 fil 🎽 out D	tra mod	tra % mod						res
		ave						
		ave						

• Hubungkan garis dari mod, tra, ave pada Validation ke res.

Process ×		100% 🔎	P	P	 2	\$	
Process Read CSV	Decision Tree	Validation					res res res (

- Double click pada Validation.
- Hubungkan titik tra ke mod menggunakan Decision Tree, mod dan tes ke res menggunakan Apply Model dan Performance.



Klik Run.



Tree.

•

```
SUARA SAH CALEG > 3082: YA {TIDAK=3, YA=23}
SUARA SAH CALEG ≤ 3082
   SUARA SAH CALEG > 2418
        SUARA SAH CALEG > 2935: TIDAK {TIDAK=5, YA=0}
        SUARA SAH CALEG ≤ 2935
           SUARA SAH PARTAI > 16443
               DAERAH PEMILIHAN > 1.500
                   SUARA SAH PARTAI > 83822: TIDAK {TIDAK=5, YA=1}
                    SUARA SAH PARTAI ≤ 83822
                    | SUARA SAH CALEG > 2445.500: YA {TIDAK=1, YA=
                       SUARA SAH CALEG ≤ 2445.500: TIDAK {TIDAK=1,
                DAERAH PEMILIHAN ≤ 1.500: YA {TIDAK=0, YA=4}
           SUARA SAH PARTAI ≤ 16443: TIDAK {TIDAK=2, YA=0}
        SUARA SAH CALEG ≤ 2418
       SUARA SAH CALEG > 1436.500
    Т
           SUARA SAH CALEG > 1468: TIDAK {TIDAK=47, YA=9}
        1
           SUARA SAH CALEG ≤ 1468: YA {TIDAK=0, YA=3}
        Т
       SUARA SAH CALEG ≤ 1436.500: TIDAK {TIDAK=313, YA=0}
    Т
```

Dari grafik tersebut dapat diketahui bahwa:

Apabila suara sah caleg lebih besar dari 3082 suara maka dikelompokkan pada prediksi ya (terpilih).

Apabila kurang dari nilai tersebut maka sistem akan meninjau kembali nilai besaran suara sah caleg.

- Apabila suara sah caleg lebih dari 2418
 - Jika suara sah caleg lebih dari 2935 maka diprediksikan **tidak** terpilih
 - Jika suara sah caleg kurang dari samadengan 2935 dan kurang dari sama dengan 16433 maka diprediksikan **tidak** terpilih
 - Jika suara sah caleg kurang dari sama dengan 2935 namun lebih besar dari 16433 maka
 - Jika daerah pemilihan kurang dari sama dengan 1500 maka diprediksikan ya (terpilih).
 - Jika daerah pemilihan lebih dari 1500 maka ditinjau kembali dari suara sah partai. Jika suara sah partai lebih dari 83822 maka diperkirakan tidak terpilih. Jika suara sah partai kurang dari sama dengan 83822 maka ditinjau dari perolehan suara sah caleg kembali.
 - Jika lebih besar dari 2445.5 maka diprediksi ya (terpilih).
 - Jika kurang dari sama dengan 2445.5 maka diprediksi tidak terpilih.
- Apabila suara sah caleg kurang dari samadengan 2418.
 - Jika suara sah caleg kurang dari sama dengan 1436.5 atau lebih besar dari 1468 maka diprediksikan **tidak** terpilih. Jika suara sah caleg kurang dari sama dengan 1468 maka diprediksikan ya (terpilih)
 - Jika suara sah caleg kurang dari sama dengan 1468 maka diprediksikan ya (terpilih).

Tugas 06

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban:

Pada kasus ini saya mengambil pengolahan data terkait dengan Market Basket yang akan dicoba pada tool rapidminer. Adapun tahapannya sebagai berikut :

1. Create tambel master datanya di Excel kemudian export ke Rapidminer.

TID	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP
001	1	1	1	0	0	0	0
002	0	1	1	1	0	0	0
003	0	0	0	1	1	1	0
004	0	1	1	0	0	0	0
005	0	1	1	1	0	1	1

Export ke Rapidminer dan sesuaikan type datanya :

× 📕 E:	xampleSet (//Le	ocal Repository/	data/Tugas 06/1	Fugas 06_market b	asket analysis _.	_update) ×		
xampleSet (5	examples, 0 spe	cial attributes, 8 re	gular attributes)			F	filter (5 / 5 examples):	all
Row No.	TID 个	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP
1	001	1	1	1	0	0	0	0
2	002	0	1	1	1	0	0	0
3	003	0	0	0	1	1	1	0
4	004	0	1	1	0	0	0	0
5	005	0	1	1	1	0	1	1

Nama : Bhijanta Wyasa WM

: MTI 23

NIM : 202420019

Kelas



2. Create Modul Association Rule, FP-Growth, dan Numerical to Binomial.

Numerical to Binomial.

Sesuaikan dengan parameter - parameter yang ada.



FP-Growth



Association Rule

Sesuaikan dengan parameter – parameter yang ada.



- Process Parameters × 100% 🔎 🔎 🔎 🛃 🥁 🕅 Process Process > • (j) logverbosity init Process Retrieve Tugas 06_... Numerical to Binomi.. FP-Growth logfile **b** res) inp exa 🕽 C exa exa exa C ori D fre D res **b** resultfile 2001 ٢ random seed Create Association .. rul Dite ▦ • send mail never ite D SYSTEM • encoding Hide advanced parameters Change compatibility (7.1.001)
- 3. Create hubungan antar modul mulai dari data master sampai dengan Association Rule.

Setelah dilakukan hubungan antar modulnya kita lakukan RUN disistem, Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 152 rules yang ditemukan.

Result History	🗙 🛒 AssociationRules (Crea	te Associ	ation Rules) ×		Repository ×
	Show rules matching	No.	Premises	Conclus	🔂 Add Data 🛛 = 👻
Dette	all of these conclusions:	137	ROTI, MENTEGA, TELUR, SUSU	KECAP 🔨	🕶 🔚 Samples
Data	ROTI	138	KECAP	ROTI, ME	🕨 🧱 data
	MENTEGA TELUR	139	ROTI, KECAP	MENTEG	processes
Z	SUSU	140	MENTEGA, KECAP	ROTI, TE	Tutorials
Graph	BUNCIS	141	ROTI, MENTEGA, KECAP	TELUR,	🕨 📜 DB
		142	TELUR, KECAP	ROTI, ME	Local Repository (bhijanta.wijayamukti)
		143	ROTI, TELUR, KECAP	MENTEG	 Gata (bhijanta.wijayamukti) Tugas 06 (bhijanta.wijayamukti)
Description		144	MENTEGA, TELUR, KECAP	ROTI, SL	Tugas 06 (bhijanta.wijayamukti - v1,
		145	ROTI, MENTEGA, TELUR, KECAP	SUSU	Rapidminer_Tugas 06 (bhijanta.w
		146	SUSU, KECAP	ROTI, ME	Tugas 06_market basker analysi
Annotations		147	ROTI, SUSU, KECAP	MENTEG	Tugas 06_market basket analysi
		148	MENTEGA, SUSU, KECAP	ROTI, TE	Tugas 06_market basker analysi
		149	ROTI, MENTEGA, SUSU, KECAP	TELUR	I Ugas Ub_prediksi hypertensi (br processes (bhijanta.wijayamukti)
	Min. Criterion:	150	TELUR, SUSU, KECAP	ROTI, ME	Cloud Repository (disconnected)
	confidence 🔻	151	ROTI, TELUR, SUSU, KECAP	MENTEG	
	Min. Criterion Value:	152	MENTEGA, TELUR, SUSU, KECAP	ROTI	
		<		>	<

Tentunya ini akan menyulitkan kita untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Maka dari itu yang harus kita lakukan adalah mengubah nilai minimum support dan minimum confidence.

Oleh karena itu kita harus mengubah nilai di FP-Growth dan Create Association Rules untuk dapat mencapai nilai minimum support dan minimum confidence, untuk dapat menghasilkan skema yg lebih sedikit untuk dapat mempermudah dalam penyusunan skema market.



BEFORE (0.3)







Nama : Bhijanta Wyasa WM NIM : 202420019

Kelas : MTI 23



BEFORE (0.5)



AFTER (0.95)



Setelah dilakukan perubahan parameter lakukan RUN kembali, Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk

dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

natching	No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
e conclusions: 🔻	1	ROTI	MENTEGA	0.800	1	1	-0.800
۵	2	MENTEGA	ROTI	0.800	1	1	-0.800
	3	SUSU	TELUR	0.400	1	1	-0.400
	4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400
	5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400



SUMBER :

Belajar Data Mining dengan RapidMiner

Penyusun:

- 1. Dennis Aprilla C
- 2. Donny Aji Baskoro
- 3. Lia Ambarwati
- 4. I Wayan Simri Wicaksana

Editor: Remi Sanjaya

Hak Cipta © pada Penulis

Hak Guna mengikuti Open Content model

Desain sampul: Dennis Aprilla C

NAMA: CORNELIA TRI WAHYUNIMATA KULIAH: ADVANCED DATABASENIM: 202420044

TUGAS 6

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawab :

Peserta	Nilai	Ujian Kopetensi	Wawancara	Diterima
P1	tinggi	bagus	baik	уа
P2	tinggi	cukup	baik	уа
P3	tinggi	kurang	buruk	tidak
P4	sedang	cukup	baik	уа
P5	sedang	bagus	baik	ya
P6	sedang	cukup	baik	ya
P7	sedang	kurang	buruk	tidak
P8	rendah	bagus	baik	уа
P9	rendah	cukup	buruk	tidak

Data Penerimaan Siswa Baru (PSB) :

Langkah-langkah :

1. Klik menu File -> New Process -> Pilih Blank

Selanjutnya pada kota dialog Repository pilih **Import Data, -> My Computer,** cari lokasi file Data yang akan diimport.



	Import Data - Select the dat	a location.				× erators_etc	🔎 All Studio 🔻
Repository ×		×					
🕒 Import Data 🛛 = 💌							
Samples	Documents	init .	•				
Local Repository (Local)	Bookmarks	File Name	Size	Туре	Last Modified		
DB (Legacy)		Adobe		File Folder	May 20, 2020	^	
		Apowersoft		File Folder	Dec 4, 2018		
		Assassin's Creed IV Black FL Aure CAPCOM Corel Corel		File Folder	Nov 7, 2017		
				File Folder	Jan 22, 2018		
				File Folder	Nov 11, 2017		
				File Folder	Nov 7, 2017		
				File Folder	Jan 13, 2018		
				File Folder	Apr 3, 2018		
				File Folder	Jan 8, 2018		
				File Folder	Apr 21, 2018		
				File Folder	May 20, 2020		
		Electronic Arts		File Folder	Jan 13, 2018		
		GitHub		File Folder	Oct 6, 2019		
perators ×		Hasil		File Folder	Aug 6, 2019		
	2	KONAMI		File Folder	Jan 13, 2018	d parameters	
Search for Operators		MATERI PROPOSAL		File Folder May 6, 201	May 6, 2019	ALL 10 0 000	
	4	📙 maya		File Folder	May 10, 2018	100111919-0-000	1
Data Access (55)		My Games		File Folder	Nov 7, 2017		
Blending (82)		My Palettes		File Folder	Nov 7, 2017		
Cleansing (28)		oman 📃		File Folder	Aug 3, 2019		
Modeling (167)		Presentation1		File Folder	Jan 8, 2018		
Scoring (14)		Denzi		Eile Folder	0433 3030	V Studio Core	
Sconing (14)						Didulo core	
Validation (30)							
Utility (85)	All Files					which is the	outer most
Extensions (2)		import. process.					
						ancel	
Get more operators from the Marketplace	-	and it must be the ro	tly one operator of ot operator of the				

2. Setelah menemukan file datanya, klik Next, kemudian setelah sampai pada kotak dialog berjudul Format Your Columns dan pada masing-masing atribut pilih logo setelan.



Kriteria :

-	pade format MieM d. y	ow himmissian.					
			•	Replace erors with	missingvalues (1)		
	Jumlah MK o -	Absensi (%) & + /ea/	Jumlah SK o + Integer	IPS S1 e +	IPS S2 0 + rear	825.53 • • real	N
	. /	0.000	0	0.000	2	9	1
2	0	0.000	a	0.000	7	7	Ti
3	23	38.002	39	0.290	0.900	0.000	
4	1	53.750	50	2.380	2.490	0.000	
5	37	60.952	60	2.250	2.820	1.800	
4	25	47.500	60	1.500	2,050	0.100	
7	35	71.845	59	1.070	2.820	2,773	
	-43.	54.771	18	2.268	1.810	1,370	
9	8	6.253	10	0.360	0.000	9	
10	9	6.253	60	0.360	0.000	9	
11	35	75.921	29	3.030	2.710	2.700	
12	17	28.355	0	2.530	0.000	0.000	
	< 10					Cono peop	2

Pilih Change Type. Kriterianya sebagai berikut ;

- **Polynominal** = untuk atribut yang memiliki lebih dari 2 kategori
- Binominal = untuk atribut yang memiliki 2 kategori saja
- **Real** = untuk tipe data yang memiliki nilai koma, atau decimal
- Integer = untuk file bernilai integer atau bilangan buat tanpa koma.
3. Kemudian pada atribut kelasnya atur Change Role dan isi valuenya sebagai label , selanjutnya **klik Nex**t sampai dengan finish.

	-	• •	1	Views:	Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operatorsetc 🔎 .	VI Studio 🔻
t History		ExampleSet (//Lo	ocal Repository	dataset PSB)	× 🚦 Exam	pleSet (//Local F	Repository/datas	et berolahraga)	×		Repository ×	
	Open in	Turbo Prep	Auto Model	1				Filter (9	/9 examples): all	¥	C Import Data	= •
Data	Row No.	Diterima	Peserta	Nilai	Ujian Kopete	Wawancara					 Samples Local Repository (Local) 	
	1	ya	P1	tinggi	bagus	baik					• Connections	
Σ	2	ya	P2	tinggi	cukup	baik					🕨 🚾 data	
stics	3	tidak	P3	tinggi	kurang	buruk					processes dataset berolahraga (11/6/2)	13/23 PM
	4	ya	P4	sedang	cukup	baik					🧬 dataset berolahraga (11/5/2)	3:38 PM - 4
T	5	ya	P5	sedang	bagus	baik					dataset PSB (11/5/20 3:40 PM	- 2 kB)
zations	6	ya	P6	sedang	cukup	baik					DB (Legacy)	
	7	tidak	P7	sedang	kurang	buruk						
	8	ya	P8	rendah	bagus	baik						
	9	tidak	P9	rendah	cukup	buruk						

4. Setelah finish pada halaman Repository akan muncul data teman-teman yang sudah Terimpor oleh sistem. Kemudian **Drag atau seret** ke halaman kerja RapidMiner kemudian klik Run (F11).

File Edit Process View Connection	ons <u>S</u> ettings Extensions <u>H</u> e	lp						
	•	Views: Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operatorsetc 💫 All Studio 🔻
Repository ×	Process							Parameters ×
🕒 Import Data 🛛 🗉 🔻	Process				Q Q	🐚 🖪 📴	G 🔹 🔊	C Retrieve dataset PSB (Retrieve)
🕨 📴 Samples	Process							repository entry Repository/dataset PSB
Local Repository (Local)								
Connections) inp				-		res	
🕨 🚞 data					(res	
processes								
dataset berolahraga (11/5/20 3:2		Retrieve	dataset PSB					
💣 dataset berolahraga (11/5/20 3:3			a 011					
dataset PSB (11/5/20 3:40 PM - 2			-					
DB (Legacy)								
Operators × performace ×								Show advanced parameters
Multi Horizon Performance								
E Sliding Window Validation								Help ×
ation (20)								^
erformance (18)								C Retrieve
Predictive (7)								RapidMiner Studio Core
% Performance (Classification)								Tags: Load, Import, Read, Datasets, Examples, Example Set, Table, Repository, Data Access
% Performance (Binominal Classific								Synansis
% Performance (Regression)								This Operator can access stored information in the
S Performance (Costs)								Repository and load them into the Process.
% Performance (Ranking)	Leverage the Wisdom of Cro	wds to get operator recommendation	tions based on you	ir process design!				Jump to Tutorial Process
< > > No results were found.			Activate Wisdon	of Crowds				
								V

5. Pada kotak dialog **Operators**, pilih menu-menu seperti **Random Tree, Split Data, Apply Model, dan Permormance.** Cara mencarinya tulisan di kotak pencarian pada kotak dialog tersebut.



Setelah menemukan semua **Opertors** tersebut, jangan lupa Drag atau seret ke halaman kerja satu persatu.

6. Selanjutnya pada **Operator Split Data klik 2x**, kemudian atur Sampling Type pada kota Parameters , ini bertujuan untuk memisah atau membagi menjadi data latih dan data uji guna untuk mengetahui permorma dari model yang dibentuk oleh k-NN



7. Setelah itu atur porposisi atau kapasitas antara data latih dan data uji. Klik pada menu edit *Enumeration*. kemudian akan muncul kotak dialog sepeti dibawah ini :

Eile Edit Process View Connection	ns <u>S</u> ettings Extensio	ns <u>H</u> elp										
	•		Views:	Design R	tesults	Turbo Prep	Auto Model	Deployments		Find data,	operatorsetc 🔎	All Studio 🔻
Repository ×	Process Process Process	🕅 Edit Param	eter List: partitic	ms			Q Q		🗵 🖷 🎜 🖡	Parameters P Split Data partitions	× Edit Enumera	ion (0)
Gamples Gamples Gonections Gonections Gata	Dimp		Edit Parameter The partitions t	r List partitions that should be crea	ated.					sampling type	linear sampling	•
dataset berolahraga (11620 3 dataset berolahraga (11620 3 dataset berolahraga (11620 3 dataset PSB (11620 3 da PM - 4) DB (Legero)	< Leverage the Wisdom	n of Growds to g	et operator reco	mmendakons bas	•) • • • • • • • • • • • • • • • • • •	Add Entry	Remove Entry	OK I	Cancel	Show advant Help × Split C RapidMir Some Subset Synopsis This operator p subsets of the g is partitioned in	Sed parameters Data Fitting Fi	ng ber of he
🔮 No results were found.				Activate	e Wisdom of	f Crowds				specified relativ	e sizes. <u>I Process</u>	

8. Add Entry kemudain Ok.

Sumber :

https://www.youtube.com/watch?v=jXRXI66inDM

https://www.youtube.com/watch?v=jB_hKsi408

https://www.youtube.com/watch?v=h82NuHDNhKI

https://www.youtube.com/watch?v=idHF3t4xMak&t=286s

NIM : 202420022

Rapidminer merupakan sebuah *tool* data mining yang dipergunakan dalam membuat sebuah model prediksi. Pengolahan data pada proses ini yaitu Teknik Klasifikasi mengunakan *Decision Tree* dimana menggunakan rangkaian *data training* dalam menentukan pembelian rumah atau tidak dimana terdiri atas 5 atribut yaitu *Age, Income, Employee, Credit_Rating* dan *Buys_House*. Dan

Pada rapid miner, dalam proses pengolahan data terdiri dari beberapa langkah-langkah, yaitu

Pertama-tama buat serangkaian data set yang terdiri dari data training dan data uji.

Age	Income	Employee	Credit_rating	Buys_House
<=30	Low	Yes	Fair	Yes
<=30	Low	Yes	Fair	Yes
3140	Medium	Yes	Excellent	No
3140	Medium	No	Excellent	No
>40	High	Yes	Fair	Yes
>40	High	Yes	Fair	Yes
>40	High	No	Fair	Yes
3140	Medium	No	Excellent	No
<=30	Low	Yes	Fair	Yes
<=30	Low	Yes	Excellent	No
3140	Medium	No	Fair	No
3140	Low	No	Fair	Yes
>40	High	Yes	Excellent	Yes
>40	High	Yes	Excellent	Yes

Tabel 1. Data Training



Age	Income	Employee	Credit_Rating	Buys_House
>40	High	Yes	Excellent	?

Kemudian dari data set yang telah dibuat dilanjutkan dengan membuka aplikasi rapid miner, lalu pilih opeartor *read excel* seperti ditunjukkan pada Gambar 1 dibawah ini.



Gambar 1. Tampilan Operator Read Excel

Setelah dari pemilihan operator *read excel* dilanjutkan dengan memghubungkan ke file *data training*. Kemudian pilih *Import Configuration Wizard* pada bagian Parameter dan akan muncul sebuah tampilan berupa *import data* dari file *excel* seperti pada Gambar 2 dibawah ini.

	36		port.	
heet: Data Training	• C	ell range: A:E Select A	II 🖌 Define hea.	1 *
A	В	с	D	E
1 Age	Income	Employee	Credit_rating	Buys_House
2 <=30	Low	Yes	Fair	Yes
3 <=30	Low	Yes	Fair	Yes
1 3140	Medium	Yes	Excellent	No
5 3140	Medium	No	Excellent	No
6 >40	High	Yes	Fair	Yes
7 >40	High	Yes	Fair	Yes
3 >40	High	No	Fair	Yes
3140	Medium	No	Excellent	No
0 <=30	Low	Yes	Fair	Yes
1 <=30	Low	Yes	Excellent	No
2 3140	Medium	No	Fair	No
3 3140	Low	No	Fair	Yes
4 ≻40	High	Yes	Excellent	Yes

Gambar 2. Import Data File Excel Data Training

Begitupun juga pada operator *read excel 2*. Pilih *import configuration wizard* dan lakukan langkah yang sama seperti opeartor *read excel* sebelumnya dengan pilihan data uji seperti pada gambar 3 dibawah ini.

Import Data - Select the cells to import.									
Select the cells to import.									
Sheet: Data Uji Cell range: A.E Select All I Define hea 1 -									
	A	В	с	D	E				
1	Age	Income	Employee	Credit_Rating	Buys_House				
2	≻40	High	Yes	Excellent	?				

Gambar 3. Import Data File Excel Data Uji

Pada step ketiga dibagian import data file dari *excel* ubah tipe atribut *buy_house* pada data training dan data uji menjadi label dimana akan digunakan untuk menciptakan model prediksi seperti ditunjukkan pada Gambar 4 dan Gambar 5 dibawah ini.

Age & v	polynominal	Employee 🔹 🔻	Credit_rating	Buys_House
<=30	Low	Yes	Fair	Yes
<=30	Low	Yes	Fair	Yes
3140	Medium	Yes	Excellent	No
3140	Medium	No	Excellent	No
>40	High	Yes	Fair	Yes
>40	High	Yes	Fair	Yes
>40	High	No	Fair	Yes
3140	Medium	No	Excellent	No
<=30	Low	Yes	Fair	Yes
<=30	Low	Yes	Excellent	No
3140	Medium	No	Fair	No
3140	Low	No	Fair	Yes



Imp	oort Data - Format ye	our colu	mns.					×			
			Form	aat vour oolu	mne						
			Forn	nat your colu	mns	•					
	\square Replace errors with missing values \oplus										
	Age polynominal	¢ v	Income 🔹 🔻	Employee polynominal	¢ v	Credit_Rating	Buys_House polynominal label	¢ v			
1	>40		High	Yes		Excellent	?				
							🥑 no pro	blems.			
						← Previous	Einish	ancel			

Gambar 5. Perubahan Atribut menjadi Label pada Buys_House Data Uji

Setelah dari *import data file excel* untuk operator *read excel* dilanjutkan dengan pemilihan operator *decision tree*. Lalu, hubungkan antara operator *read excel* dan *read excel* 2 menuju operator *decision tree* yang ditunjukkan pada Gambar 6 dibawah ini.



Gambar 6. Pemilihan Operator Decision Tree

Kemudian dilanjutkan dengan cara yang sama seperti sebelumnya pada pemilihan operator *Apply Model* seperti pada Gambar 7 yang ditunjukkan dibawah ini.

Repository ×	Process		
🔂 Import Data 🛛 🗉 💌	Process		P
Training Resources (connected)	Process		
Samples			
Community Samples (connected)	Dinp		
Local Repository (Local)	Read Excel	Decision Tree	Apply Model
DB (Legacy) Operators X Parameters X	Read Excel (2)	tra umod exa wei	med lab uni med
apply model 🗙			
Modeling (1) Time Series (1) Forecasting (1) Apply Forecast Scring (1)			
Anniv Model			
a upply model	Recommended Operators ①		
No results were found.	C Retrieve	75% 🚺 Set Role 🛃 43%	Select Attributes 👫 34%

Gambar 7. Pemilihan Operator Apply Model



Gambar 8. Tampilan Result

Gambar 8 merupakan tampilan tahap terakhir setelah node telah dihubungkan. Langkah selanjutnya yaitu pilih tombol eksekusi yang akan tampil dalam bentuk *graph* dan *apply model* yang akan ditunjukkan pada Gambar 9 dan 10 dibawah ini.



Gambar 9. Tampilan Hasil Berupa Decision Tree

Result History	2	Tree (Decision Tree	e) × 📒	ExampleSet (A	oply Model)	×			
	Open in	Turbo Prep	Auto Model				Filter (14 / 14 e)	kamples): all	
Data	Row No.	Buys_House	prediction(B	confidence(confidence(Age	Income	Employee	Credit_rat
	1	Yes	Yes	0.800	0.200	<=30	Low	Yes	Fair
Σ	2	Yes	Yes	0.800	0.200	<=30	Low	Yes	Fair
Statistics	3	No	No	0	1	3140	Medium	Yes	Excellent
	4	No	No	0	1	3140	Medium	No	Excellent
(5	Yes	Yes	1	0	≻40	High	Yes	Fair
Visualizations	6	Yes	Yes	1	0	≻40	High	Yes	Fair
	7	Yes	Yes	1	0	>40	High	No	Fair
	8	No	No	0	1	3140	Medium	No	Excellent
	9	Yes	Yes	0.800	0.200	<=30	Low	Yes	Fair
Annotations	10	No	Yes	0.800	0.200	<=30	Low	Yes	Excellent
	11	No	No	0	1	3140	Medium	No	Fair
	12	Yes	Yes	0.800	0.200	3140	Low	No	Fair
	13	Yes	Yes	1	0	≻40	High	Yes	Excellent
	<	10 IV 144 IV	06 1900-090 1907 - 20 - 47						>

Gambar 10. Tampilan Hasil Berupa Apply Model Data

Sumber: Youtuber Payo Belajar

Ada dua teknik validasi yang lumrah digunakan untuk memvalidasi model:

- 1. *Split validation*: melakukan validasi sederhana dengan membagi dataset secara acak menjadi dua data terpisah data latih & data uji.
- 2. *Cross validation*: melakukan validasi berulang di mana dataset dibagi menjadi banyak subset (himpunan) data latih & validasi. Setiap iterasi memvalidasi (menguji) satu subset data dengan subset yang tersisa sebagai data latih. Pada cross validation, # subset data adalah jumlah iterasi.

1. Split Validation (Training Error & Test Error)

Salah satu cara mengetahui kinerja model adalah dengan mengukur akurasinya (meskipun akurasi bukan satu-satunya parameter yang digunakan untuk mengukur kinerja suatu model). Ada dua konsep pada *split validation: training error* dan *test error*.

- 1. *Training error* didapatkan dengan menghitung kesalahan klasifikasi model pada data yang sama dengan model yang dilatih.
- 2. *Test error* didapatkan dengan menggunakan dua data yang sepenuhnya terpisah. Satu untuk melatih model (data latih) dan lainnya untuk menghitung kesalahan klasifikasi (data uji). Kedua dataset harus memiliki nilai label yang sama.

Traning Error

Nilai *traning error* didapatkan dengan menggunakan data yang sama untuk melatih dan menguji model. Perhatikan proses pada gambar 1



Gambar 1. Proses mencari nilai training error.

Pada praktik ini, menggunakan data sonar yang telah tersedia di *repository* RapidMiner. Operator *performance* digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan. Misalkan untuk klasifikasi, kriteria yang diberikan adalah *accuracy*, *precision* dan *recall*.

Jika proses dijalankan maka menghasilkan performance metrics seperti Gambar 2.

accuracy: 86.54%									
	true Rock	true Mine	class precision						
pred. Rock	78	9	89.66%						
pred. Mine	19	102	84.30%						
class recall	80.41%	91.89%							

Gambar 2. Performance metrics traning error.

Dari *confusion matrix* pada Gambar 2, nilai akurasi yang didapatkan sebesar 86.54%. Apakah model tersebut merupakan model yang terbaik karena memiliki akurasi tinggi ?

Nama : Efrik kartono Ahsa NIM : 202420030

Telah dijelaskan bahwa *training error* merupakan kesalahan klasifikasi yang didapatkan menggunakan data yang sama untuk melatih dan menguji model. Karena data yang digunakan sama, maka kemungkinan hasil prediksi "benar" yang diberikan model sangatlah besar. Oleh karena itu disarankan untuk menggabaikan nilai *training error* meskipun memiliki akurasi yang tinggi karena selalu memberikan estimasi yang terlalu "optimis". Sayangnya, banyak referensi yang memberikan nilai *training error* sebagai hasil akhirnya. Sebenarnya ini adalah praktik yang buruk dan harus dihindari.

Hal terbaik yang bisa dilakukan adalah memberikan estimasi seberapa baik kinerja model di masa depan. Jika dilakukan dengan cara yang benar, estimasi ini akan mendekati apa yang dapat dicapai tetapi tidak ada jaminan bahwa kinerja yang diperkirakan akan persis seperti yang diharapkan.

Test error: data latih & data uji

Test error adalah perkiraan yang jauh lebih baik tentang seberapa baik kinerja model untuk kasus-kasus baru dan tak terlihat di masa depan. Diatas telah disinggung untuk mendapatkan *test error* diperlukan dua data yang sepenuhnya terpisah yaitu data latih dan data uji. Kebanyakan dataset yang kita miliki belum memiliki data uji. Untuk mengatasi hal tersebut, pada RapidMiner telah disediakan operator *split data* yang dapat digunakan untuk membagi dataset menjadi partisi data latih dan data uji sesuai porsi yang ditentukan. Perhatikan proses pada Gambar 3.



Gambar 3. Proses membagi dataset menjadi data uji & data latih.

Untuk membaginya, harus menambahkan rasio pada paramater *partition*. Jumlah rasio dari semua partisi harus bernilai total 1.0 (satu). Karena kita ingin melatih model sebaik mungkin maka rasio partisi untuk data latih harus lebih besar dari rasio partisi data uji. Pada praktik ini kita membagi 80% dari dataset sebagai data latih dan sisanya 20% dari dataset sebagai data uji. Perhatikan gambar 4 dibawah ini.



Gambar 4. Rasio 0.8 untuk partisi data latih dan 0.2 untuk partisi data uji.

Nama : Efrik kartono Ahsa NIM : 202420030

Sedangkan operator *store* pada Gambar 3 digunakan untuk menyimpan objek hasil partisi dari operator *split data* ke *repository* RapidMiner. Objek yang disimpan dapat digunakan untuk proses lain dengan menggunakan operator *retrieve*.

Gunakan data latih dan data uji yang telah disiapkan sebelumnya untuk mendapatkan nilai *test error*. Perhatikan proses pada Gambar 5.



Gambar 5. Proses mencari nilai test error.

accuracy: 56.29%								
	true Rock	true Mine	class precision					
pred. Rock	38	33	53.52%					
pred. Mine	40	56	58.33%					
class recall	48.72%	62.92%						

Gambar 6. Performance metrics test error

Dari *confusion matrix* pada Gambar 6, nilai akurasi yang dihasilkan *test error* sebesar 56.29%. Jauh berbeda dengan nilai akurasi yang didapatkan *training error* pada gambar 2 sebesar 86.54%. Karena menggunakan data yang terpisah untuk melatih dan menguji model, bisa jadi data yang diuji tidak dilatih pada model sehingga menghasilkan banyak prediksi "salah". Oleh karena itu, sangat penting untuk memahami perbedaan antara *training error* dan *test error*.

Biasanya, model akan semakin baik jika menggunakan lebih banyak data latih. Hal ini akan mempengaruhi kita untuk menggunakan data sebanyak mungkin untuk melatih model. Pada saat yang sama, kita ingin menggunakan data sebanyak mungkin untuk menguji model sehingga mendapatkan kesalahan prediksi pada pengujian model yang jauh lebih baik. Lalu, apakah ada cara yang lebih baik untuk memvalidasi sebuah model ?

2. Cross Validation

Dari dua percobaan diatas kita setuju bahwa menggunakan data uji dari dataset untuk menghitung *test error* merupakan cara yang lebih baik untuk mendapatkan estimasi yang lebih handal pada akurasi model di masa mendatang. Menggunakan data uji juga pendekatan yang efisien untuk memvalidasi model. Tetapi pada praktiknya masih ada potensi masalah yang timbul: bagaimana mengetahui data uji tersebut tidak terlalu mudah untuk model? Bisa jadi sampel acak yang dipilih tidak begitu acak, terutama jika hanya memiliki dataset yang sedikit. Dalam kasus tersebut, *test error* yang dihasilkan mungkin kurang mewakili akurasi model.

K-fold merupakan salah satu metode *cross validation*. Konsep *k-fold cross validation* tidak hanya membuat beberapa sampel data uji berulang kali, tetapi membagi dataset menjadi bagian terpisah dengan ukuran yang sama. Model dilatih oleh subset data latih dan divalidasi oleh subset validasi (data uji) sebanyak *k*. Dengan *k-fold cross validation* dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi model. Gambar 7 menunjukkan bagaimana prinsip *cross validation*.



Gambar 7. Konsep cross validation.

Pada RapidMiner, operator *cross validation* adalah operator yang bersarang yang memiliki dua subproses: subproses *training* yang digunakan untuk melatih model dan subproses *testing* untuk pengujian sekaligus mengukur kinerja model. Perhatikan proses pada Gambar 8 dan Gambar 9.



Gambar 8. Proses cross validation.

Nama : Efrik kartono Ahsa NIM : 202420030



Gambar 9. Subproses pada operator cross validation.

Operator *cross validation* memiliki parameter yang dapat digunakan. *Number of folds* digunakan untuk memberikan nilai *k* (jumlah iterasi). *Sampling type* digunakan untuk memilih teknik *sampling* yang membagi dataset.

S Cross Validation					
split on batch attribute					
leave one out		١			
number of folds 💙	10	١			
sampling type 💙	automatic	•			
use local random seed					
<pre> enable parallel execution </pre>					

Gambar 10. Parameter pada operator cross validation.

accuracy: 80.71% +/- 13.32% (micro average: 80.77%)								
	true Rock	class precision						
pred. Rock	72	15	82.76%					
pred. Mine	25	96	79.34%					
class recall	74.23%	86.49%						

Gambar 11. Performance metrics dari 10 fold cross validation.

Dari *confusion matrix* pada Gambar 11, nilai akurasi yang dihasilkan dari 10 *fold cross* validation sebesar 80.71% dengan standar deviasi (+/-) sebesar 13.32%. Ketika mengukur kinerja model menggunakan *cross validation* maka akurasi yang dihasilkan memiliki standar deviasi atau simpangan baku yaitu ukuran penyebaran data yang menunjukan jarak rata-rata dari nilai tengah ke suatu titik nilai. Semakin besar simpangan baku yang dihasilkan, maka penyebaran dari nilai tengahnya juga besar, begitu pula sebaliknya. Tujuan simpangan baku adalah melihat jarak antara rata-rata akurasi dengan akurasi setiap percobaan (iterasi).



TUGAS 06 ADVANCED DATABASE

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

JAWAB:

Penerapan Algoritma C.45 dalam Prediksi Kelulusan Tapat Waktu Dengan Rapidminer.

Data Mahasiswa								
	ІРК	PRESTAS I	ΕΤΙΚΑ	SKS	KELAYAKA N			
	TINGGI	ADA	BAIK	PENUH	LULUS			
•	SEDAN							
_	G	ADA	BAIK	PENUH	LULUS			
	SEDAN	TIDAK						
	G	ADA	BAIK	PENUH	LULUS			
	RENDA				TIDAK			
	Н	ADA	BAIK	PENUH	LULUS			
	RENDA			TIDAK	TIDAK			
	Н	ADA	BAIK	PENUH	LULUS			
	TINGGI	ADA	BAIK	PENUH	LULUS			
-		TIDAK	TIDAK	TIDAK	TIDAK			
	TINGGI	ADA	BAIK	PENUH	LULUS			
		TIDAK						
	TINGGI	ADA	BAIK	PENUH	LULUS			
	SEDAN		TIDAK					
-	G	ADA	BAIK	PENUH	LULUS			
	RENDA				TIDAK			
	Н	ADA	BAIK	PENUH	LULUS			
	RENDA				TIDAK			
•	Н	ADA	BAIK	PENUH	LULUS			
	RENDA	TIDAK		TIDAK	TIDAK			
-	Н	ADA	BAIK	PENUH	LULUS			
	SEDAN	TIDAK	TIDAK	TIDAK	TIDAK			
-	G	ADA	BAIK	PENUH	LULUS			
	SEDAN	TIDAK	TIDAK	DENUU				
-	G		BAIK	PENUH	LULUS			
	TINGGI		DVIK		111110			
		ADA	DAIK					
	TINGGI	٨٩٨	BVIK	DENITIO				
•	RENDA		DAIN					
	H	ADA	BAIK	PENUH	LUIUS			
•	TINGGI	ADA	BAIK	PENUH	LULUS			
-	SEDAN							
	G	ΔΠΔ	RAIK	PENIIH	10005			
-	NGGI		BAIK	PENIIH				
	NGGI	ADA	BAIK	PENUH	LULUS			
י ר		TIDAK						
(BAIK					
	п	ADA	DAIN	PENUT	LULUS			



SEDAN				
G	ADA	BAIK	PENUH	LULUS
			TIDAK	TIDAK
TINGGI	ADA	BAIK	PENUH	LULUS
SEDAN	TIDAK	TIDAK	TIDAK	TIDAK
G	ADA	BAIK	PENUH	LULUS
RENDA	TIDAK	TIDAK		TIDAK
Н	ADA	BAIK	PENUH	LULUS
TINGGI	ADA	BAIK	PENUH	LULUS
		TIDAK		
TINGGI	ADA	BAIK	PENUH	LULUS
TINGGI	ADA	BAIK	PENUH	LULUS
SEDAN		TIDAK		
G	ADA	BAIK	PENUH	LULUS
SEDAN		TIDAK		
G	ADA	BAIK	PENUH	LULUS
SEDAN	TIDAK			
G	ADA	BAIK	PENUH	LULUS
TINGGI	ADA	BAIK	PENUH	LULUS
RENDA		TIDAK	TIDAK	TIDAK
Н	ADA	BAIK	PENUH	LULUS
TINGGI	ADA	BAIK	PENUH	LULUS

proses perhitungan jumlah data, entropy dan gain. Hasil sebagai berikut :

no			Jum	Lul	Tid ak Lul	Entro	Coin
de			lan	us	us	pny	Gain
1	total		25	10	16	0,994	
	lotai		35	19	10	0950	0.4702
	אחו						12522
	IPK	Tine				0 721	12532
		ring	15	10	2	0,721	
		gi	15	12	3	9281	
		Sed	11	0	h	0,684	
		ang	11	9	Z	0384	
		Ken	0	0	0	0	
	Dues	uan	9	U	9	0	0.0241
	Pres						0,0241
	tasi						//905
					•	0,965	
		Ada	23	14	9	6361	
		Tida					
		k		_	_	0,979	
		Ada	12	5	7	8688	
	Etik						0,0021
	а						3128
						0,989	
		Baik	25	14	11	5875	
		Tida					
		k					
		Baik	10	5	5	1	
							0,5416
	SKS						15542



Data diatas, nilai gain tertinggi terdapat pada SKS dibandingkan dengan atribut lainnya terlihat gain tertinggi yaitu sks, sks menjadi sebuah akar karena memiliki gain tertinggi pertama. Perhatikan gambar dibawah ini:



Mengapa pohon keputusan sampai pada IPK, karena nilai nya diantara lulus dan tidak lulus terdapat nilai **0**, maka kelayakan keputusan nya langsung didapat. Kemudian terlihat etika dan prestasi tidak termasuk yang mempengaruhi kelulusan mahasiswa tepat waktu.

RAPIDMINER :

selanjutnya adalah melakukan pengujian terhadap data sampel berupa tabel yang berada diexcel melalui tools **Rapidminer** dimulai dari proses koneksi antara basis data sampel, operator dan validasi seperti gambar berikut ini :

Dari proses koneksi pada gambar diatas didapatkan pohon keputusan yang sama dengan hasil manual diatas yang telah dibuat. Pohon keputusan yang terdapat pada tools rapidminer dapat dilihat pada gambar dibawah ini :



Berikut adalah screenshot hasil pengukuran akurasi terhadap kinerja Algoritma C.45 dalam memprediksi kelulusan tepat waktu yaitu sebesar 92,60% +\- 1.60% :

oniarion oaraon	🕘 Multiclass Classification	Performance () Annotations							
accuracy AUC	Table View Plot Vi	() Table View () Phot View							
	accuracy: 92.60% +1-1.36	% (mikro: 92.60%)							
		true negative	tue positive	class precision					
	pred. negative	430	32	93.07%					
	pred. positive	5	33	86.84%					
	dass recall	98.85%	50.77%						

Kemudian setelah proses ini dilakukan penarikan informasi yang akan menjadi hasil dari **kesimpulan dalam memprediksi kelulusan tepat waktu berdasarkan pohon keputusan** pada tools Rapidminer.

Tree

SKS	= PENUH
1	IPK = RENDAH: TIDAK LULUS {LULUS=0, TIDAK LULUS=4}
1	IPK = SEDANG: LULUS {LULUS=9, TIDAK LULUS=0}
1	IPK = TINGGI: LULUS {LULUS=12, TIDAK LULUS=0}
SKS	= TIDAK PENUH: TIDAK LULUS {LULUS=0, TIDAK LULUS=10}

KESIMPULAN :

Dari hasil analis pada penerapan Algoritma C.45 yang telah dilakukan, maka dapat diambil beberapa kesimpulan diantaranya sebagai berikut :

1. Pada sistem data mining menggunakan

metode *Algoritma C.45* dapat diimplementasikan didunia pendidikan dalam memprediksi kelulusan tepat waktu pada perguruan tinggi.

2. Dari hasil pengujian dapat diketahui bahwa faktor yang mempengaruhi kelulusan mahasiswa yang pertama kali adalah SKS dan IPK.



3. Dari hasil pengujian pada tools Rapidminer 5 menghasilkan sebuah

akurasi dengan tingkat kinerja Algoritma C.45 yaitu 92,60% +\-1.60%. jadi dapat diketahui bahwa kelulusan mahasiswa tepat waktu di UBD Kisaran yang didapatkan dari data 35 mahasiswa aktif, mendapatkan tingkat kelulusan tepat waktu berjumlah 21 mahasiswa dari 35 mahasiswa.

- 4. Mahasiswa yang tidak memiliki SKS penuh akan diberi toleransi dan bimbingan agar dapat lulus pada tepat waktu.
- 5. Mahasiswa yang memiliki IPK rendah akan dibina dan diberi arahan untuk terus meningkatkan kualitas diri agar dapat tercapai kelulusan sesuai masa studi.

SUMBER :

https://osf.io/unqt4/download

Selesai

ada 3 tool yang digunakan pada rapidminer yaitu : Decision tree, Naive Bayes, k-Nearest Neighbor

1. Tutorial Metode Decision Tree pada Rapid Miner

https://fnsfind16.wordpress.com/2013/11/11/metode-decision-tree-pada-rapid-miner/

Tools Rapid Miner. Tools ini digunakan sebagai alat bantu untuk melakukan penambangan data alias Data Minning. Untuk percobaan, penulis menggunakan data klasmen Liga Inggris. Tujuannya adalah mencari tahu siapa bakal juara Liga Inggris berdasarkan poin hasil pertandingan sampai bulan Oktober :). Data yang diambil berdasarkan 10 musim terakhir. Data tersebut diambil dari situs http://www.premierleague.com. Berikut adalah contoh datanya.

POS 10	GA 10	GC 10	PTS 10	CHAM
1	19	8	24	YES
2	20	9	23	NO
3	18	6	22	NO
4	9	5	19	NO
5	20	13	18	NO
6	10	5	16	NO
7	20	12	15	NO
8	15	12	15	NO
9	14	13	15	NO
10	15	11	14	NO
1	31	12	26	NO
2	16	3	26	YES
3	14	10	23	NO
4	18	13	21	NO
5	19	14	18	NO
6	18	10	17	NO
7	11	9	17	NO
8	22	19	16	NO
9	15	12	15	NO
10	14	13	15	NO
1	28	6	31	YES

Dengan

Pos10: adalah posisi tim pada pekan ke 10

GA10: jumlah memasukan sebuah tim pada pekan ke 10

GC10: jumlah kebobolan tim pada pekan ke 10

PTS10: jumlah poin tim sampai pada pekan ke 10

CHAM: apakah tim tersebut pada akhirnya menjadi juara (Yes/No)

Nah, sekarang kita coba olah data tersebut dengan menggunakan aturan Decision Tree pada rapid miner. Caranya adalah sebagai berikut

Buka aplikasi RapidMiner, Klik File lalu new process hingga terdapat tampilan sebagai berikut:

Advertisement

REPORT THIS AD



Bagian proses ini adalah untuk menambahkan proses-proses yang digunakan untuk mengolah data tersebut. Hal ini dikarenakan metode digunakan untuk mengolah data berdasarkan metode input -> proses -> output. output akan keluar sebuah diagram yang merepresentasikan sebuah informasi tertentu.

Untuk mengimpor data tersebut, klik ikon impor file pada menu reprositories lalu pilih import Excel Sheet karena data tersebut disimpan pada Microsoft Excel. Pilihlah dokumen yang menyimpan data tersebut:



Klik Next hingga terlihat data yang diimpor tersebut pada aplikasi RapidMiner

A	.8	C	0	E
105.10	GA 10	66.10	P15 19	CHW
				10 115
				10 102

Klik next untuk menambahkan anotasi, bila tidak akan menambahkan sebuah anotasi, klik next kembali.

Pada tampilan pengubahan variabel, ubah tipe kolom CHAM yang berisi data binomial (yes/no) ke tipe label. Untuk Decision Tree, tipe kolom label wajib ada agar data dapat diolah.

			- C						
			10	PT5 10	10		GA 10	2	POS 10
		m. *	+ 0	integer	per +		intege		inte per
		- P	• 6	attrioute	tute +	- 10	attribu	- 10	attribut
6			11	24		-	19		1
			N	23			29		2
			19	22			18		3
			14	19			9		4
			19	18			20		6
			14	16			10		8
			19	15			20		7
D									0
🛃 genore errors 🔲 Strow onto genore								1918	0 ek
Evalue Message	Original vi		WIE			une .	w. Colu	Ro	
e value Illesisage	Ongenal vi		WW.	,		(mat	w. Colu	Ro	

Klik Next, lalu beri nama data tersebut dan data tersebut disimpan pada local Reprository

tawae (asta6PL				
Location /Local Pascellery/GalaEPL				
	Surger Danions	**	Pat Enten	💥 Gancel

Klik tombol Finish, proses berlanjut untuk mengatur proses dari data tersebut.

Berikut adalah cara untuk menyusun sebuah proses untuk mengolah data:

1. Pilih folder Local Reprository. Tarik data yang tadi telah diimpor dari Excel ke bagian Proses.

2. Carilah operasi Decision Tree pada menu Operators. Drag operasi tersebut ke bidang proses. Sehingga pada proses terlihat ebagai berikut:

Transformation (1) Compare raise Discongroutilies Compare raise Discongroutilies Compare raise Discongroutilies Compare raise discongrout Compare raise discongro		
Recentricities Could be an		
Composition of the control of t	A Poolerna II. D Log II.	
- Cata banga Ma zu - un 19693 16 22 AN - 2 M	30 2 potential problema	
	Unitage A Contemptor Innocellar and American a susception in many (i) and (ii) the suite ii	1
CONTRACTOR DECEMBER OF CONTRACTOR	O standadov issue meeting at post Decision Tree Statistics set	

Hubungkan Retrieve dataEPL dengan operasi Decision Tree, pastikan ketika dihubungkan tidak melihat pesan error. Hubungkan juga output dari decision tree ke sisi kanan, atau ke tonjolan res Berikut adalah proses yang telah dihubungkan.



Setelah selesai, tekan F11, bila tidak ada error, maka akan keluar hasil seperti berikut:



Dengan data tersebut, maka didapatkan sebuah informasi bahwa:

1. Tim dengan poin lebih besar dari 26,5 pada saat memasuki pekan ke-10, dalam 10 tahun terakhir dipastikan menjadi juara, dan seterusnya.

2.<u>Tutorial Rapid Miner dengan Metode Naive Bayes</u> http://cuex-me.blogspot.com/2015/12/tutorial-rapidminer-dengan-metodenaive_13.html

langkah-langkah menggunakan software RapidMiner dengan metode Naïve Bayes. Mohon maaf bila dalam penulisan tutorial ini masih kurang lengkap karena saya juga dalam keadaan belajar dan inilah hasil dari kerja keras saya selama belajar RapidMiner. Oke baiklah kita langsung saja ☺ Langkah pertama open RapidMiner 5.0 karena disini saya menggunakan software RapidMiner versi 5.0 dan jika sobat2 menggunakan versi yang berbeda yah silahkan disesuaikan saja. Ok kemudian double klik new proses untuk membuka lembar kerja baru.



Ok, ficture diatas merupakan tampilan lembar kerja baru pada RapidMiner 5.0. ok next langkah selanjutnya ketik read pada bagian operator kemudian double klik read excel bias sobat lihat pada gambar dibawah ini :



Ok kalau sudah tampil seperti ficture diatas silahkan sobat next kelangkah berikutnya 😊

kemudian langkah selanjutnya klik import configuration wizard akan tampil gambar seperti dibawah ini:

name (and the second se	1 111	ander vorde regist vor beig ander vorde regist vor beig ander the the final strand to reported			A P & S. A Dattor
	Reserverse Reserverse Reserverse	Parson # Annel # Salvan Pro # Salvan Pro	the Tape For Some Par Some Par Some Par Some Par Some Par Some Par Some Par Some Par Some	Lat Webber Rec.1.2014 407 (2):2014 405 (2):2	
	Tear Seasthing on		(a	19 X (max)	and Faced Standards of Campelier

pada gambar diatas cari data yang akan anda klasifikasikan dan untuk percobaan, disini saya menggunakan data klasmen Liga Inggris, dan shobat bisa menggunakan data sesuai dengan keinginan shobat.

	-	The super stand in the	and and track		-	h	Contract of the second
Contributing Tare		4.05				America	Hypother Hoard.
		Robertaria Carl Denney	CLUB. Information (1) It on	700 344 142 Month Mark 143 Month Mark	La Rober an Al 1994 an An 1994		1 at at at at at at at at at at
		EXIL OF APPEL PR					department.
	a 11	for bashing in		(m	1000 X 2000	And the loss of	ts an Example for Example

Kemudian klik tombol next sampai di step ke empat , dan prediksi klasifikasi harus ada labelnya dan langsung double klik tombol finish seperti gambar dibawah ini :

	1	Tes etc.			ine di Stania a da c da stania da c	ng, pa cardolar be bit lanc. Ng cardonar (fir ty farindada anal har antas)	Coordinates Tyderrows, Copyrights These cards		Ginne -
Costruores Leve			-		-			Anesto	Hprator Hard.
Scientification (1)	0.00		de parts		C. P. Marrier and P. M.	Canada Sanana			
S Create Archite France	0	63		-63	-			L .	
144.75	100.0	0110	00 H	PT1-10	0.04				
A Real The	steps	1 Internet	1 steps	1 mager	· Depris · ·				141
d factor at fung	414	-	· anna	1 400	- 144				and the second s
And to a	1	10		24	at the D		0		
A fast scient	1	10		15	and the second s				
and the second se	1	18		20	Cone III				
		-	-		1400				and the second s
	1000	14		14					
	1	29	10	. 10	other Bill			anne (NARTICE)
Name Report of the American State	a						D		
	0.000					10.00	the same () there are from		
	· · · · ·			_		Collect rates	Records.	ad Inch	and have from
							Martine Martine		to an Exampleir

Ok, setelah melewati beberapa kali tahap next dan next hehe maka Dari langkah diatas didapat kan hasil seperti gambar dibawah ini:



Kemudian tambahkan metode naivebyes, kalau disini kita gak perlu lagi hitung-hitungannya, karena disini kita dianggap memakai. seperti gambar dibawah ini :



Jika sudah selesai output pada read excel kita gandengkan ke training (naïve bayes) jika warna nya berubah menjadi warna ungu berarti benar dan jika merah itu berarti salah seperti gambar dibawah ini :



Kemudian langkah selanjutnya tekan F 11 dan pilih SimpleDistribution untuk melihat hasil outputnya seperti gambar dibawah ini :

0 >

3. Tutorial RapidMiner Menggunakan Algoritma k-NN (k-Nearest Neighbor)

Pertama admin menggunakan *Tools atau Sofware RapidMiner Studio Free Versi 8.1*. jika teman-teman belum punya bisa juga download dan regristasi versi free di *website resminya*

Kedua, sebelum teman-teman masuk ke proses rapid miner, kalian harus dan sangat harus memiliki Dataset atau Data Latih. Karena itu adalah inti dari mengolah data menjadi informasi menggunakan ilmu data mining.

Setelah semuanya sudah siap, Bukalah aplikasi RapidMiner

Klik menu File -> New Process -> Pilih Blank

Selanjutnya pada kota dialog Repository pilih Add Data, -> My Computer, cari lokasi file Data latih teman-teman.



Langkah Ketiga setelah teman-teman sudah menemukan file datanya, klik Next, kemudian setelah sampai pada kotak dialog berjudul Format Your Columns dan pada masing-masing atribut pilih logo setelan.

			Format your	columns.			
	Quite format MINIM d. y	mputers	•	Replace errors with	missing values (1)		
	Jumiah MK o +	Absensi (%) o + /ea/	Jerslah SK o + Integer	IPS S1 • • Jow	IPS S2 0 + mar	8PS S3 0 + read	N p
1	0	0.000	0	0.000	2	9	
2	0	0.000	0	0.000	7	7	1
3	23	38.002	39	0.290	0.000	0.000	
4	1	53.750	50	2.350	2.490	0.000	
5	37	66.952	60	2.259	2.820	1.800	
4	25	47.509	60	1.589	2.950	0.100	
7	35	71.845	59	1.970	2.820	2,778	
	40.	54.771	18	2.268	1.810	1,370	
9	8	6.253	18	0.360	0.000	9	
10	8	6.253	60	0.360	0.000	9	
11	35	75.821	39	3.030	2.710	2.700	
12	17	28.365	0	2.530	0.000	0.000	
	<		55.C II. 5				2

Pilih Change Type. Kriterianya sebagai berikut ;

- Polynominal = untuk atribut yang memiliki lebih dari 2 kategori
- Binominal = untuk atribut yang memiliki 2 kategori saja
- Real = untuk tipe data yang memiliki nilai koma, atau decimal
- Integer = untuk file bernilai integer atau bilangan buat tanpa koma.

Kemudian pada atribut kelasnya atur Change Role dan isi valuenya sebagai label , selanjutnya klik Next sampai dengan finish

				Format yo	ur columns.	<u>.</u>			
Ū.a	e format	MMM d. yyr	rhmmss az	•	Replace error	s with missing valu	0 20		
•	Absers	(%) 0 -	Juniah SK	• - IPS S1	• + IPS S2	e + iPS S3	۰.	Kelas	• •
		谢 Change	rente				-		
	0.000		Danna ante	a the new sole.				Keluar	
2	0.000		Presservers	The new tool.			- 18	Relation	
	38,002	-						Kelsar	
	63.750	Isbel					•	Katuar	
	60.952						_	Rotuar	
	47.509					~ ~ ~ ~		Keluar	
	11.845				Y	~~~		Kiele (r	
	54,778	_				1.010	_	KONGK	
	0.250		18	9.910	0.000			40693	
	0.250		00	0.000	0.000	*		4,612	
	15.821		20	3,030	2.710	2.700		KAN A	
	28.345		0	2.530	0.000	0.900		Kelsar	

Setelah finish pada halaman Repository akan muncul data teman-teman yang sudah Terimpor oleh sistem. Kemudian **Drag atau seret** ke halaman kerja RapidMiner



Selanjtnya pada kotak dialog **Operators** , pilih menu-menu seperti **Random Tree, Split Data, Apply Model, dan Permormance.** Cara mencarinya tulisan di kotak pencarian pada kotak dialog tersebut.

perators	×	
nn	-	×
Modeli	ng (1)	
• Pre	dictive (1)	
	Lazy (1)	
	k-NN	and the second se

Setelah menemukan semua **Opertors** tersebut, jangan lupa Drag atau seret ke halaman kerja satu persatu.

Selanjutnya pada **operator Split Data klik 2x**, kemudian atur Sampling Type pada kota Parameters , ini bertujuan untuk memisah atau membagi menjadi data latih dan data uji guna untuk mengetahui permorma dari model yang dibentuk oleh k-NN



Setelah itu atur porposisi atau kapasitas antara data latih dan data uji. Klik pada menu edit *Enumeration*. kemudian akan muncul kotak dialog sepeti dibaawah ini



Add Entry kemudain Ok. Maksud dari 0.9 adalah data latih diambil 90% dan 0.1 adalah Data Uji diambil 10% dari perbandiingan Seluruh Dataset.

Langkah Terakhir adalah sambungan semua operators seperti gambar dibawah ini.



Setelah semua sudah tersambung, klik **tombol Play atau Start** seperti arah panah yang ada Digambar atas itu. Kemudian jika proses sudah selesai akan mencul seperti

X Example	Set (Apply Mindul)	Performance/Vector (Perform	nance) 🛪 🔳 Eranyud	int (M. et all Repository) Claim Latin)	Repositor
tarton Sanco Idalen Sali	Table View Phot View eccarrecy: 190.00%	te.		† 3	
C loptimistici		Yee Kenar	the Lotes	class practision	1
	pred Keluar	-	0	100.00%	*
ic patamato	pred Lutes	0	29	100.00%	
	class recall	100.00%	100.00%		1 m

Penjelasan arah panah dengan nomor

- 1. Hasil total yang sudah dilakukan proses Split data
- 2. Hasil dari nilai Performa menggunakan Algoritma k-NN
- 3. Total Dataset asli
- 4. Beberapa Kriteria Evaluasi untuk mengukur performa Suatu Algoritma

1. Model Prediksi Ini Menggunakan Tools RapidMiner

RapidMiner memiliki keunggulan tersendiri, RapidMiner merupakan aplikasi data mining berbasis sistem open-source dunia yang terkemuka dan ternama. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai data mining. Solusi yang diusung antara lain Integrasi data, analisis ETL, analisis data dan pelaporan dalam satu suite tunggal. Memiliki antarmuka pengguna grafis yang intuitif untuk desain analisis proses, serta fleksibel terdapat banyak transformasi data, pemodelan data, dan metode visualisasi data.

2. Model Prediksi Elektabilitas Caleg

a. Modeling Decision Tree

Decision tree adalah model prediksi yang digambarkan dengan struktur pohon atau struktur berhirarki. Konsep dari decision tree adalah membantu user dalam mengubah data menjadi pohon keputusan dan syarat-syarat keputusan agar lebih mudah dipahami dan dipetakan.



Deskripsi :

- Jika calon memiliki suara sah lebih dari 3.082 maka akan terpilih menjadi anggota legislatif. Jika kurang dari 3082 akan dipertimbangkan dalam dua klasifikasi yaitu jumlah suara >2418 dan ≤ 2418
- jika suara sah caleg ≤2418 akan dibagi lagi ke dalam kelompok suara sah caleg jika memiliki ≤ 1436.500 maka caleg tidak akan terpilih jika jumlah suara >1436.500 dan terdapat >1468 suara maka tidak akan terpilih dan jika ≤1468 maka akan terpilih menjadi anggota legislatif
- Dalam jumlah suara sebanyak >2418 terbagi menjadi dua klasifikasi yaitu apabila suara yang dimiliki >2935 maka caleg tidak akan terpilih. Dan jika ≤2935 akan ditentukan oleh suara sah partai dan daerah pemilihan untuk menentukan terpilih atau tidaknya menjadi anggota legislatif
- Jika memiliki suara sah partai berjumlah ≤ 16443 maka tidak akan terpilih menjadi anggota legislative. Jika memiliki suara sah partai >16443 maka akan ditentukan dari daerah pemilihannya.
- Jika daerah pemilih memiliki suara ≤1.500 maka caleg tersebut akan terpilih menjadi legislatif
- Kemudian jika memiliki suara daerah pemilihan >1.500 akan ditentukan dari suara sah partai.
- Jika suara sah partai >83822 maka tidak akan terpilih menjadi anggota legislatif. Dan jika suara sah partai ≤83822 akan ditinjau kembali dari suara sah caleg.
- Jika suara sah caleg >2445.500 maka ia akan terpilih menjadi anggota legislatif, dan jika ≤2445.500 maka tidak akan terpilih.

	true TIDAK	true YA	class precision
pred. TIDAK	362	14	96.28%
pred. YA	15	34	69.39%
class recall	96.02%	70.83%	

accuracy: 93.16% +/- 4.45% (mikro: 93.18%)

AUC: 0.915 +/- 0.076 (mikro: 0.915) (positive class: YA)



b. Modeling Naïve Bayes

Naive Bayes merupakan suatu klasifikasi berpeluang sederhana berdasarkan teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen).

SimpleDistribution

Distribution model for label attribute TERPILIH ATAU TIDAK

```
Class TIDAK (0.887)
10 distributions
Class YA (0.113)
10 distributions
```

	true TIDAK	true YA	class precision
pred. TIDAK	326	18	94.77%
pred. YA	51	30	37.04%
class recall	86.47%	62.50%	



c. Modeling K-nn

K-Neares Neighbour (K-NN) adalah sebuah metde klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

KNNRegression

```
1-Nearest Neighbour model for regression.
The model contains 425 examples with 10 dimensions.
```

Berdasarkan pernyataan tersebut dapat diartikan bahwa dari data yang dianalisis terdapat 425 data dengan 10 dimensi (atribut) untuk menentukan target "ya" dan "tidak".

	true TIDAK	true YA	class precision
pred. TIDAK	358	25	93.47%
pred. YA	19	23	54.76%
class recall	94.96%	47.92%	

ассштаст	r 89 63% +/- 2	21% (mikro	89 65%)
accurac	. 03.03/0 -1-2		. 03.03/0

AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: YA)



3. Evaluasi dan Akurasi

	Decision Tree (C4.5)	Native Bayers (NB)	K-Nearest Neighbor(K-NN)
Akurasi	93.16 %	83.73 %	89.63 %
Minus	4.45 %	4.25 %	2.21 %
AUC	0.915	0.840	0.500

Kesimpulan :

Berdasarkan tabel diatas, Model *Decision Tree* memiliki akurasi tertinggi dengan nilai sebesar 93.16 % diikuti K-nn dengan nilai 89.63 % dan pada urutan terakhir *Naïve Bayes* dengan akurasi paling rendah sebsar 83.73 %. Dari hasil tersebut dapat ditarik kesimpulan bahwa modeling yang paling baik digunakan untuk dataset tersebut yaitu *Decision Tree* model , karena memiliki akurasi prediksi paling besar dengan nilai presentase sebesar 93.16 % diantara ketiga model, dapat diartikan analisis yang dihasilkan dari Model *Decision Tree* memiliki kemungkinan menjadi kenyataan paling besar dan tepat.
Untuk melakukan validasi terhadap *Tree* yang dihasilkan, digunakan nilai dari area dibawah kurva ROC (AUC/*Area Under Curve*). Menurut Gorunescu (2011), hasil perhitungan yang divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu :

- a. Akurasi bernilai 0.90 1.00 = excellent classification
- b. Akurasi bernilai 0.80 0.90 = good classification
- c. Akurasi bernilai 0.70 0.80 = fair classification
- d. Akurasi bernilai $0.60 0.70 = poor \ classification$
- *e*. Akurasi bernilai 0.50 0.60 = failure

Berdasarkan tabel tersebut juga dipresentasikan bahwa dengan menggunakan "Decision

Tree, AUC/ Area Under Currve yang dihasilkan lebih besar yaitu 0.91 > 0.90 masuk dalam kategori *excellent classification*", sedangkan untuk *Native Bayers* masuk dalam kategori *good classification*, dan *K-nearest* ada pada kategori *failure*.

Dapat disimpulkan bahwasannya dengan dataset pemilu apabila ingin melakukan pengklasifikasian data dengan hasil yang paling akurat dan baik maka disarankan untuk menggunakan model *Decision Tree*

Tahapan installasi aplikasi Rapidminer

- 1. Download https://rapidminer.com/get-started/
- 2. Install dan register (selesai)
- 3. Jalankan aplikasi Rapidminer
- 4. Siapkan dataset Dataset : diabetes_data_upload.csv

https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.

- 5. Import ke local repository
- 6. Mode design , drag dataset dan drag operator decision tree



Result

🎒 //Local Reposito	ry/Process/Early Stage Diabe	tes* – RapidMiner Studio E	ducational 9.8.000 @ OKI6600	and the state	- Section	-		-
<u>File Edit Proce</u>	ss <u>V</u> iew <u>C</u> onnections <u>S</u>	<u>S</u> ettings E <u>x</u> tensions <u>H</u>	əlp					
			Views: Design	Results	Turbo Prep	Auto Model	Deployments	Find data
V	Tree (Decision Tree)	×		ExampleSet (//Lo	ocal Repository/di	abetes_data_uplo	ad) \times	
Result H	listory	Exan	npleSet (Set Role) $ imes$		1	AttributeWeights	s (Decision Tree)	×
Graph	Zoom	V	hum	հիրդես	Aqua	. Sets Mg	coldine weight here	a be Agest
Description	✓ Node Labels ✓ Edge Labels	Abguste angen Star Bage Abguste Bagetter Bagetter	Name and Appendix App	Negalite Para	allfren.h Ve Pen.lfre	Failur Failur	Pestler Kapeller Pestler	Ruchy Pacha Adaptificada Rucha Pacha
Annotations			ref ref Registrie Analte					

Attibute weight

<u>File Edit</u> Proces	s <u>V</u> iew <u>C</u> onnections	<u>Settings</u> Extension
	-	•
Q	Tree (Decision Tree)	×
Result Hi	story	
Data	attribute	weight
	partial paresis	0.003
	Obesity	0.031
Weight	Irritability	0.041
Visualizations	Gender	0.076
	Itching	0.030
	Alopecia	0.127
Annotations	Genital thrush	0.071
	delayed healing	0.091
	sudden weight loss	0.117
	visual blurring	0.117
	Polyuria	0.042
	Polydipsia	0.040
	Age	0.215

Example set (display)

🎒 //Local Repositor	y/Process/Early Stag	ge Diabetes* – Rapi	idMiner Studio Educational 9.	8.000 @ OKI6600	And Description of the	-			
<u>File Edit Proces</u>	ss <u>V</u> iew <u>C</u> onneo	tions <u>S</u> ettings	E <u>x</u> tensions <u>H</u> elp						
	-	•	Views:	Design Resul	ts Turbo Prep A	uto Model De	eployments	F	ind data
9	Tree (Decision Tr	ee) ×		ExampleS	et (//Local Repository/diab	etes_data_uploa	nd) ×		
Result H	istory		ExampleSet (Set	Role) $ imes$	📑 Attr	ibuteWeights (De	cision Tree)	×	
Data	Open in Turbo Prep Auto Model Filter (520 / 520 examples): all							•	
Data	Itching	Irritability	delayed healing	partial paresis	muscle stiffness	Alopecia	Obesity	label	
	Yes	No	Yes	Yes	No	No	No	Positive	^
Σ	No	No	Yes	Yes	Yes	No	No	Positive	
Statistics	No	No	Yes	No	Yes	No	No	Positive	
	No	Yes	Yes	Yes	Yes	Yes	Yes	Positive	

Visualisasi weight



Auto Model

勴 //Local Reposi	itory/Process/Early Stage	e Diabetes* – Rapidł	Miner Studio Educatior	nal 9.8.000 @ OK1660	0	****				-		
<u>Eile Edit Pro</u>	ocess ⊻iew <u>C</u> onnecti	ions <u>S</u> ettings Ex	tensions <u>H</u> elp									
	-	•	View	/s: Design	Results	Turbo Prep	Auto Model D	eployments	FI	ind data, operatorsetc	P	All Stud
Auto Model												
			Load Data	Select Task	Prepare Target	Select Inputs	Model Types	Results				
		Wan	Predict at to predict the values	of a column?	Clu Want to identify g	sters roups in your data?	Want to dete	Dutliers ct outliers in your da	ta?			
Age _{Number}	Gender Category	Polyuria _{Category}	Polydipsia Category	sudden weig _{Category}	weakness Category	Polyphagia Category	Genital thrush Category	visual blurring Category	Itching Category	Irritability Category	delayed	ŀ
40	Male	No	Yes	No	Yes	No	No	No	Yes	No	Yes	^
58	Male	No	No	No	Yes	No	No	Yes	No	No	No	
41	Male	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	
45	Male	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	
60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	
55	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	
57												
41 45 60 55	Male Male Male	Yes Yes	No No Yes Yes	Yes Yes No	Yes Yes Yes	Yes Yes Yes	Yes No No	No Yes Yes	Yes Yes Yes	No Yes No	Yes Yes Yes	

Cluster

//Local Repository/Process/Ea	rly Stage Diabetes* – RapidMiner Studio Educational 9.8.000 @ OKI6600
<u>File Edit Process View</u>	2onnections <u>S</u> ettings Extensions <u>H</u> elp
	Views: Design Results Turbo Prep Auto Model Deployments
Auto Model	
	Load Data Select Task Prepare Target Select Inputs Model Types Results
Results	k-Means - Summary
 K-Means Summary Heat Map 	Number of Clusters: 2 Distance Measure: Squared Euclidean Distance Average Cluster Distance: 2425.903
Cluster Tree	Davies-Bouldin Index: -**
Centroid Chart	Cluster 0 279 Average Distance: ?
Centroid Table	visual blurring = No is on average 219 61% smaller. Polyuria = No is on average 165 76% smaller, Age is on average 151 76% smaller.
Clustered Data	
 Image: Second Data Image: Second Data Image: Second Data 	Cluster 1 241 Average Distance: 2425.903 partial paresis = No is on average 247.92% smaller, Alopecia = No is on average 210.54% smaller, delayed healing = No is on average 205.63% smaller

Judul : Deteksi dini diabetes

Dataset : diabetes_data_upload.csv

NIM : 202420040 ISTIARSO B

Advance DB (Data Mining) Penggunaan Aplikasi Rapidminer

https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.

- Attribute
- Age : Umur
- Gender : Jenis Kelamin
- Polyuria : Sering kencing
- Polydispia : sering haus
- Sudden Weight loss: berat bada turun drastis
- Weakness: lemah
- Polyphagia: sering lapar
- Genital Thrush: gatal alat kelamin
- Visual blurring: penglihatan buram/kabur
- Itching: gatal
- Irritability: sifat lekas marah
- Delayed Healing: susah sembuh (luka)
- Partial Paresis: kelumpuhan sebagian
- Muscle stiffness: otot kaku
- Alopecia: rambut rontok
- Obesity: Kegemukan
- Label : (evaluasi positif atau Negatif)

JIka di konversi Yes=1 dan No=0, Male=1 dan Female=0, ujicoba menggunakan data Diabetes_data1.csv



×

📑 AttributeWeig	ghts (Decisio
attribute	weight
Obesity	0.026
Irritability	0.006
Gender	0.026
muscle stiffness	0.033
weakness	0.001
Alopecia	0.045
Genital thrush	0.071
sudden weight loss	0.124
delayed healing	0.259
visual blurring	0.124
Polyuria	0.052
Polydipsia	0.026
Age	0.207

1	AttributeWeights (Decision Tree)	× 📕 E	xampleSet (Set F	Role) ×	💡 Tree (Decisio	n Tree) 🛛 🗙	ExampleSet (//Local Repository/Data/diabetes_data1)				
Open in	Turbo Prep	Auto Model							Filter (520 /	520 examples):	all	
Row No.	label	Age	Gender	Polyuria	Polydipsia	sudden wei	weakness	Polyphagia	Genital thrush	visual blurring	Itching	Ir
1	Positive	40	1	0	1	0	1	0	0	0	1	0
2	Positive	58	1	0	0	0	1	0	0	1	0	0
3	Positive	41	1	1	0	0	1	1	0	0	1	0
4	Positive	45	1	0	0	1	1	1	1	0	1	0
5	Positive	60	1	1	1	1	1	1	0	1	1	1
6	Positive	55	1	1	1	0	1	1	0	1	1	0
7	Positive	57	1	1	1	0	1	1	1	0	0	0
8	Positive	66	1	1	1	1	1	0	0	1	1	1
9	Positive	67	1	1	1	0	1	1	1	0	1	1
10	Positive	70	1	0	1	1	1	1	0	1	1	1
11	Positive	44	1	1	1	0	1	0	1	0	0	1

Hasil Decision Tree



Automodel > Predict



		Load Dat	ta Select Task Pro	epare Target Select	Inputs Model Typ	es Results					
	Selected: 16 / Total: 16 Deselect Yellow Select All Select All Select All										
Selected	Status 个	Quality	Name	Correlation	ID-ness	Stabillity	Missing	Text-ness			
✓	•	C I S M	Polyuria	44.35%	0.38%	50.38%	0.00%	0.00%			
✓	•	C I S M	Polydipsia	42.09%	0.38%	55.19%	0.00%	0.00%			
√	•		Age	1.18%	9.81%	5.77%	0.00%	0.00%			
✓	•	C I S M	Gender	20.18%	0.38%	63.08%	0.00%	0.00%			

Auto Model						
	Load Data	Select Task Prepare Target	Select Inputs Model Typ	pes Results		
	•	•				
		Ҝ RESTART 🔏 BAC				
Execution	Models				Data Prepa	aration
Execute on: Local Computer	Naive Bayes				Remove Colu	umns with Too Man
(this machine)	Generalized Line	ear Model			Maximum Number of V	/alues: 50 🖡
Queue: No queues availa	Use Regularization	Calculate p-Values			Extract Date	Information
Select Folder for Storing Result The results of this run will be sto	ed in the Cogistic Regres	sion			Extract Text	Information
folder selected below. We recom use an empty folder in the select	ed Al Hub	ain			Select Text Columns	(0)
Local Repository (Local)	✓ Automatically Optimiz	e			Number of Extracted F	eatures: 1,000 🔹
	Deep Learning				Automatic Fe	eature Selection
	Decision Tree				Additional Minutes (Ma	ximum): 60 🌲
	Automatically Optimiz	e Maximal Depth: 20 🛊			Final Feature Set shou	Ild be Accurate
Load « Overview	Data Select Task Prepar	e Target Select Inputs	Model Types Re PORT Ø DEPLO	esults		
Load « Overview	Data Select Task Prepar	e Target Select Inputs	Model Types Re PORT OFFLO	esults		
Load « Overview Classification El	Data Select Task Prepar	e Target Select Inputs	Model Types Re PORT © DEPLO lodels: 205 Runtimes (ms)	esults Y		
Load (((Overview Classification El (().0% ()	Data Select Task Prepar RESTART & BACK Do O	e Target Select Inputs	Model Types Re PORT ODEPLO lodels: 205 Runtimes (ms)	esults	and the second sec	20.000 (Mill Energy) 20.000 (Mill Energy) 0
Load Coverview Classification Er	Data Select Task Prepar RESTART & BACK D= 0 TOT TOT Select Task Prepar BACK D= 0	e Target Select Inputs	Model Types Re PORT DEPLO lodels: 205 Runtimes (ms) Standard Deviation	ssults	new contraction of the second	20.000 Ming Time (1998 Bares)
Load Classification End Ulassification End Classification End Classification End	Data Select Task Prepar RESTART & BACK D= 01	e Target Select Inputs	Model Types Re PORT DEPLO Iodels: 205 Runtimes (ms) Standard Deviation ± 3.8%	ssults	Total Time 23 s	10,000 yring Time (Millions) 20,000 to 0 10,000 Trainin 965 m s
Load Classification End Classification Error V	Data Select Task Prepar RESTART & BACK Do O TOP TOP TOP Model Naive Bayes Generalized Linear Model	e Target Select Inputs	Model Types Re PORT DEPLO lodels: 205 Runtimes (ms) 100,000 1	sults	Total Time 23 s 9 s	20.000 Ming Time (20.000 Ming 1) 20.000 Ming (20.000 Ming 1) 10.000 Ming (20.000 Ming 1) 0
Load Classification Error	Data Select Task Prepar RESTART & BACK D= 01 TOT South State St	e Target Select Inputs	Model Types Re PORT DEPLO Iodels: 205 Runtimes (ms) 10,000	ssults	sector of the se	20.000 Million

Hasilnya

BELAJAR DATA MINING DENGAN RAPID MINER

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati Wayan Simri Wicaksana

BELAJAR DATA MINING DENGAN RAPID MINER

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana

Identitas

Belajar Data Mining dengan RapidMiner

Penyusun:

Dennis Aprilla C Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor: Remi Sanjaya

Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C

Kata Pengantar

Dengan mengucapkan puji syukur kepada Tuhan YME atas Berkah Rahmat dan Hidayah-Nya, penulis dapat menyelesaikan buku yang berjudul Belajar Data Mining dengan RapidMiner.

Produk-produk perangkat lunak gratis (freeware) dan bersifat open source yang demikian banyak jumlahnya, telah memudahkan kita dalam melakukan proses pengolahan dan analisis data. Dalam melakukan analis terhadap data mining, RapidMiner merupakan salah satu solusi yang dapat kita gunakan. Keberadaan RapidMiner yang berupa freeware dan dapat dijalankan pada berbagai sistem operasi tidak hanya menguntungkan penyedia aplikasi karena tidak perlu mengeluarkan biaya untuk lisensi perangkat lunak, tetapi juga memudahkan pengembang maupun calon pengembang dalam mempelajari dan mencoba sendiri fitur-fitur yang ada.

Buku ini diharapkan dapat membantu pembaca mempelajari RapidMiner, melalui rangkaian tutorial bertahap mulai dari proses instalasi hingga pemrograman. Pada buku ini juga dibahas beberapa teori penunjang mengenai data mining seperti, decision tree, neural network dan market basket analysis untuk membuka wawasan pembaca mengenai data mining sebelum melakukan analisis data mining.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu penyelesaian buku ini.

Akhir kata, penulis menyadari masih terdapat kekurangan dalam penyusunan buku ini baik pada teknis penulisan maupun materi, mengingat akan kemampuan yang dimiliki penulis. Untuk itu kritik dan saran dari semua pihak penulis harapkan demi penyempurnaan pembuatan buku ini. Semoga buku ini dapat bermanfaat bagi para pembaca.

Jakarta, April 2013

Penulis

Daftar Isi

Kata Pengantar	i
Daftar Isi ii	i
Daftar Gambar	v
Daftar Tabelvii	i
Kecerdasan Buatan	2
Definisi Kecerdasan Buatan	2
Ruang Lingkup Kecerdasan Buatan	5
Perbedaan Komputasi Kecerdasan Buatan dan Komputasi Konvensional	6
RapidMinerError! Bookmark not defined.	8
Apa itu RapidMiner?	3
Instalasi Software1	1
Pengenalan Interface10	5
Cara Menggunakan Repositori28	3

Data Mining	39
Mengenal Data Mining	39
Pengelompokan Teknik Data Mining	43
Decision Tree	45
Mengenal Decision Tree	45
Algoritma c4.5	48
Kelebihan Pohon Keputusan	55
Kekurangan Pohon Keputusan	56
Decision Tree pada RapidMiner	56
Neural Network	84
Market Basket Analysis	96
Memahami Market Basket Analysis	96
Metodologi Association Rules	100
Contoh Association Rules	102
Frequent Itemset Generation dan Rule Generati	on
	105
Market Basket Analysis pada RapidMiner	107
Glossarium	122
Daftar Pustaka	125

Daftar Gambar

Gambar 1.1 Proses Kecerdasan Buatan	4
Gambar 2.1 Form Awal Instalasi	14
Gambar 2.2 Form Persetujuan Lisensi	14
Gambar 2.3 Form Pemilihan Lokasi Instalasi	15
Gambar 2.4 Form Proses Instalasi	15
Gambar 2.5 Form Instalasi selesai	16
Gambar 2.6 Tampilan Welcome Perspective	17
Gambar 2.7 Welcome Perspective	19
Gambar 2.8 Header Tab	20
Gambar 2.9 Tampilan Design Perspective	21
Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki	23
Gambar 2.11 Tampilan Parameter View	25
Gambar 2.12 Problem & Log View	27
Gambar 2.13 Kumpulan Sample Data Repository	28
Gambar 2.14 Tampilan Design Perspective Awal	29
Gambar 2.15 Repository berada dalam Main Process	29
Gambar 2.16 Menghubungkan Output Repositori ke Result	30
Gambar 2.17 Isi Sample Golf Data Repository	30
Gambar 2.18 Repository	32
Gambar 2.19 Step 1 of 5 Import Wizard	32
Gambar 2.20 Step 2 of 5 Import Wizard	33
Gambar 2.21 Step 3 of 5 Import Wizard	34
Gambar 2.22 Step 4 of 5 Import Wizard	34
Gambar 2.23 Tipe Data	35

Gambar 2.24 Step 5 of 5 Import Wizard	.35
Gambar 2.25 Repository yang sudah diimport	.36
Gambar 2.26 Menghubungkan Output Repositori pada Result	.36
Gambar 2.27 Tabel Repository	.37
Gambar 4.1 Bentuk Decision Tree Secara Umum	.48
Gambar 4.2 Grafik Entropi	.50
Gambar 4.3 Tabel Keputusan dalam Format xls	.57
Gambar 4.4 Lokasi Tabel pada Repository	.58
Gambar 4.5 Repository PlayGolf pada Main Process	.59
Gambar 4.6 Daftar Operator pada View Operators	.59
Gambar 4.7 Posisi Operator Decision Tree	.60
Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator Decision	
Tree	.61
Gambar 4.9 Parameter Decision Tree	.62
Gambar 4.10 Tipe Criterion	.62
Gambar 4.11 Ikon Run	.66
Gambar 4.12 Hasil Berupa Graph Pohon Keputusan	.66
Gambar 4.13 Hasil Berupa Penjelasan Teks	.67
Gambar 4.14 Tabel SakitHipertensi dalam format xls	.69
Gambar 4.15 Lokasi Tabel pada Repository	.69
Gambar 4.16 Tabel SakitHipertensi pada Main Process	. 70
Gambar 4.17 Hirarki Operator X-Validation	.72
Gambar 4.18 Operator Validation	.72
Gambar 4.19 Parameter X-Validation	.74
Gambar 4.20 Hirarki Operator Apply	. 77
Gambar 4.21 Operator Apply Model	.78
Gambar 4.22 Parameter Apply Model	. 79
Gambar 4.23 Hirarki Operator Performance	.80
Gambar 4.24 Operator Performance	.81
Gambar 4.25 Parameter Performance	.82
Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Performar	ice
	.82
Gambar 4.27 Susunan Operator Retrieve dengan Operator Validation	.83
Gambar 4.28 Tampilan Decision Tree	.83
Gambar 6.1 Frequent Item Set tanpa Apriori	106
Gambar 6.2 Frequent Item Set dengan Apriori	106

Gambar 6.3 Tabel Penjualan Sederhana	108
Gambar 6.4 Repositori	108
Gambar 6.5 Database dalam Main Process	109
Gambar 6.6 Operator Create Association Rules	109
Gambar 6.7 Operator FP-Growth	110
Gambar 6.8 Operator Numerical to Binominal	110
Gambar 6.9 Pencarian Operator Numerical to Binominal	111
Gambar 6.10 Pencarian Association Rules	112
Gambar 6.11 Menghubungan Database TransaksiMakanan pada C	perator
Numerical to Binomial	112
Gambar 6.12 Parameter Numerical to Binomial	113
Gambar 6.13 Menghubungkan Operator Numerical to Binomial de	ngan
Operator FP-Growth	114
Gambar 6.14 Parameter FP-Growth	115
Gambar 6.15 Menghubungkan Operator FP-Growth dengan Opera	itor
Create Association Rules	115
Gambar 6.16 Parameter Association Rules	116
Gambar 6.17 Susunan Operator Association Rules	117
Gambar 6.18 Hasil Association Rules Pertama	117
Gambar 6.19 Operator FP-Growth	118
Gambar 6.20 Mengubah Parameter FP-Growth	119
Gambar 6.21 Operator Create Association Rules	119
Gambar 6.22 Mengubah Parameter Association Rules	120
Gambar 6.23 Hasil Association Rules Kedua	120
Gambar 6.24 Hasil dalam bentuk Graph View	121

Daftar Tabel

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konven	sional7
Tabel 4.1 Keputusan Bermain Tenis	52
Tabel 4.2 Perhitungan Simpul 1	53
Tabel 6.1 Tabel Transaksi	102
Tabel 6.2 Kombinasi Produk dan Nilai Support	
Tabel 6.3 Association Rules dan Nilai Confidence	104

Bagian Satu Pendahuluan

Pengenalan Kecerdasan Buatan

Pengenalan RapidMiner

Kecerdasan Buatan

Definisi Kecerdasan Buatan

Manusia memiliki kecerdasan, manusia memiliki kemampuan untuk menganalisa suatu masalah dengan menggunakan pengetahuan dalam otaknya dan

pengalaman yang pernah dilaluinya. Pengetahuan ketika manusia belajar, maka dari itu datang pembelajaran merupakan faktor penting bagi manusia untuk mencapai sebuah kecerdasan. Namun pengetahuan tidak akan cukup untuk menyelesaikan masalah jika tidak memiliki pengalaman, karena pengalaman akan selalu membawa pengetahuan baru. Tetapi akan sia sia, jika seseorang yang memiliki banyak pengalaman tetapi tidak memiliki akal untuk menalar

semua pengetahuan dan pengalaman yang ia miliki. Kombinasi dari pengetahuan, pengalaman, dan kemampuan menalar inilah yang membuat manusia menjadi cerdas dan dapat menyelesaikan permasalahan yang ia hadapi.

Berdasarkan konsep diataslah kecerdasan buatan dibuat. Agar mesin dapat bertindak seperti seorang manusia, maka mesin tersebut harus memiliki pengetahuan dan sejumlah pengalaman serta kemampuan menalar yang dapat mengubah pengetahuan dan pengalaman tersebut menjadi sebuah keputusan dalam menyelesaikan sebuah permasalahan.

Komputer awalnya diciptakan hanya untuk melakukan sebuah perhitungan saja. Jaman terus akhirnya berkembang hingga komputer kini diberdayakan manusia untuk membantu pekerjaannya kesehariannya. Maka dari itu dalam komputer diharapkan memiliki kemampuan yang hampir sama dengan manusia agar dapat mengerjakan segala sesuatu vang bisa dikerjakan oleh manusia Kecerdasan Buatan.

The art of creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)

The study of how to make computers do things at which, at the moment, people are better (Rich dan Knight, 1991)

A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes (Schalkoff, 1990)

The branch of computer science that is concerned with the automation of intelligent behavior (Luger dan Stubblefield, 1993)

Jadi apakah kecerdasan buatan itu? Kecerdasan buatan adalah salah satu bagian dari ilmu komputer yang membuat agar mesin dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia. Dengan demikian, untuk menciptakan sebuah aplikasi kecerdasan buatan terdapat dua bagian utama yang sangat dibutuhkan.



Gambar 1.1 Proses Kecerdasan Buatan

Ruang Lingkup Kecerdasan Buatan

Kecerdasan buatan merupakan teknologi yang fleksibel, dan dapat diterapkan di berbagai macam bidang ilmu. Kemampuan kecerdasan buatan menjadi sangat dibutuhkan di bidang ilmu lain, karena konsepnya tak lagi procedural melainkan meniru cara berpikir manusia. Tak heran kecerdasan buatan bisa di gunakan untuk bidang psikologi yang dikenal dengan cognition dan psycolinguistic. Namun yang paling sering dekat dengan kita ialah robotika, yakni kecerdasan buatan di dalam ilmu elektornika.

Semakin banyaknya ilmu yang menggunakan kecerdasan buatan, semakin sulit juga bagi manusia untuk mengkategorikannya, maka dari itu dibentuklah ruang lingkup kecerdasan buatan yang dapat mewakilinya (Turban dan Frenzel, 1992, pp21-26):

 Sistem Pakar. komputer digunakan untuk menyimpan pengetahuan para pakar. Dengan demikian komputer akan memiliki keahlian untuk menyelesaikan permasalahan dengan meniru keahlian yang dimiliki oleh pakar.

- Pengolahan Basa Alami. dengan pengolahan bahasa alami ini diharapkan user dapat berkomunikasi dengan komputer dengan menggunakan bahasa sehari-hari.
- Pengenalan Ucapan. Melalui pengenalan ucapan diharapkan manusia dapat berkomunikasi dengan komputer dengan menggunakan suara.
- 4. Robotika dan Sistem Sensor
- Computer Vision. Mencoba untuk dapat menginterpretasikan gambar atau objek-objek tampak melalui komputer.
- Intelligent Computer-aided Instruction.
 Komputer dapat digunakan sebagai tutor yang dapat melatih dan mengajar.
- 7. Game Playing.

Perbedaan Komputasi Kecerdasan Buatan dan Komputasi Konvensional

Komputasi Konvensional merupakan Komputer yang hanya digunakan untuk alat hitung. Sangatlah berbeda, kerja dan konsep dari kedua komputasi ini. Agar dapat memberikan gambaran, table berikut adalah

detail dari perbedaan komputasi kecerdasan buatan dan komputasi konvensional.

Dimensi	Komputasi Kecerdasan	Komputasi
	Buatan	Konvensional
Pemrosesan	Mengandung konsep-	Algoritmik
	konsep simbolik	
Sifat Input	Bisa tidak lengkap	Harus lengkap
Pencarian	Kebanyakan bersifat	Biasanya didasarkan
	heuristic	pada algoritma
Keterangan	Disediakan	Biasanya tidak
		disediakan
Fokus	Pengetahuan	Data dan Informasi
Struktur	Kontrol dipisahkan dari	Kontrol terintegrasi
	pengetahuan	dengan informasi
Kemampuan	Ya	Tidak
menalar		

Tabel 1.1 Perbedaan Kecerdasan Buatan dan Komputasi Konvensional

RapidMiner

Apa itu RapidMiner?

RapidMiner merupakan perangakat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan munggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya dikembangkan pada tahun mulai 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan mengunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML)Extensible Markup Language) yang mendefenisikan proses analitis keingginan pengguna untuk diterpkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analis secara otomatis. RapidMiner memiliki beberapa sifat sebagai berikut:

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukaran data.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari RapidMiner, antara lain:

- Banyaknya algoritma data mining, seperti decision treee dan self-organization map.
- Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation,

loading), data preprocessing, visualisasi, modelling dan evaluasi

- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI
- Mengintegrasikan proyek data mining Weka dan statistika R

Instalasi Software

System Requirement

Sebelum melakukan instalasi software RapidMiner, terdapat beberapa spesifikasi minimal dimiliki komputer harus vang vang pengguna. Spesifikasi minimal bergantung pada komputer dan sistem operasi yang akan diinstal. Berikut ini beberapa spesifikasi minimal dibutuhkan software yang RapidMiner:

1. Sistem Operasi

RapidMiner merupakan software yang multiplatform, sehingga software ini dapat dijalankan pada berbagai sistem operasi. Berikut ini beberapa jenis sistem operasi yang dapat diinstal RapidMiner:

- ✓ Microsoft Windows (x86-32) → Windows XP, Windows Server 2003, Windows Vista, Windows Server 2008, Windows 7
- ✓ Microsoft Windows (x64) → Windows XP untuk x64, Windows Server 2003 untuk x64, Windows Vista untuk x64, Windows Server 2008 untuk x64, Windows 7 untuk x64
- ✓ Unix sistem 32 atau 64 bit
- ✓ Linux sistem 32 atau 64 bit
- Apple Macintosh sistem 32 atau 64 bit
 Sebagai bahan pertimbangan, kami merekomendasikan untuk penggunaan sistem
 64 bit. Hal ini dikarenakan jumlah maksimum yang dapat digunakan oleh RapidMiner terbatas pada sistem operasi dengan sistem
 32, yaitu hanya sebasar 2GB.
- 2. Java Runtime Environment versi 6

Selain itu, penggunaan server RapidAnalytics dalam kombinasi dengan RapidMiner dapat memaksimalkan proses analisis pada RapidMiner, meskipun tugas analisis sudah banyak dapat dijalankan dengan RapidMiner desktop client. Dalam hal ini proses analisa dirancang dengan RapidMiner, kemudian dieksekusi oleh server RapidAnalytics.

Instalasi RapidMiner

Seperti yang yang telah dikemukakan sebelumnya bahwa RapidMiner merupakan software gratis yang bersifat terbuka (open source). Software ini dapat dijalankan pada sistem operai Windows, Linux, maupun Mac. RapidMiner dapat diunduh pada situs resminya, yaitu <u>www.rapid-i.com</u>. Pada bagian ini, akan dijelaskan bagaimana cara melakukan instalasi software RapidMiner versi 5.3 pada sistem operasi Microsoft Windows.

Untuk memulai instalasi software RapidMiner pada sistem operasi Microsoft Windows, jalankan file installer RapidMiner-5.3.000x32-install.exe, sehingga akan muncul tampilan wizard seperti pada Gambar 2.



Gambar 2.1 Form Awal Instalasi

Klik **Next >** untuk melanjutkan pada form persetujuan dan lisensi seperti pada Gambar 2.3



Gambar 2.2 Form Persetujuan Lisensi

Pilih **I Agree** untuk melanjutkan. Kemudian, wizard akan menampilkan form seperti pada gambar 2.4.

Choose Install Location		T
Choose the folder in which to install RapidMir	ner 5.	U
Please select the folder where RapidMiner 5 administrator rights you should select a direc your home directory. Installing into a writabi use the automatic update service of RapidMi	should be installed. If y tory into which you can e directory is especially ner <mark>5</mark> without having ac	ou are a user without a write, e.g. a directory in important if you want to dministrator rights.
Destination Folder		Browse
Space required: 187.7MB Space available: 63.8GB		

Gambar 2.3 Form Pemilihan Lokasi Instalasi

Pilih **Install** untuk melakukan proses instalasi. Kemudian wizard akan menampilkan progress dari proses tersebut, seperti yang ditunjukkan pada Gambar 2.5.

③ RapidMiner 5 Setup	
Installation Complete Setup was completed successfully.	
Completed	
Show details	
	Carrel
	Carca

Gambar 2.4 Form Proses Instalasi
Setelah proses selesai, pilih **Next** > untuk melanjutkan, maka wizard akan menampilkan informasi bahwa proses instalasi telah selesai dilakukan, seperti yang ditunjukkan pada Gambar 2.6.



Gambar 2.5 Form Instalasi selesai

Pilih Finish untuk mengakhiri proses instalasi.

Pengenalan Interface

RapidMiner menyediakan tampilan yang *user friendly* untuk memudahkan penggunanya ketika menjalankan aplikasi. Tampilan pada RapidMiner dikenal dengan istilah Perspective. Pada RapidMiner terdapat 3 Perspective, yaitu; Welcome Perspective, Design Perspective dan Result Perspective.

Welcome Perspective

Ketika membuka aplikasi Anda akan disambut dengan tampilan yang disebut dengan Welcome Perspective, seperti yang ditunjukkan pada Gambar 6. Pada bagian toolbar, terdapat toolbar **Perspectives** yang terdiri dari ikon-ikon untuk menampilkan perspective dari RapidMiner. Toolbar ini dapat dikonfigurasikan sesuai dengan kebutuhan Anda. Sedangkan **Views** menunjukkan pandangan (view) yang sedang Anda tampilkan.



Gambar 2.6 Tampilan Welcome Perspective

Jika komputer Anda terhubung dengan internet, maka pada bagian bawah Welcome Perspective akan menampilkan berita terbaru mengenai RapidMiner. Bagian ini dinamakan **News**. Pada bagian tengah halaman terlihat daftar **Last Processes** (Recent Processes), bagian ini menampilkan daftar proses analisis yang baru saja dilakukan. Hal ini akan memudahkan Anda jika ingin melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut. Bagian **Actions** menunjukkan daftar aksi yang dapat Anda lakukan setelah membuka RapidMine. Berikut ini rincian lengkap daftar aksi tersebut:

- New : Aksi ini berguna ntuk memulai proses analis baru. Untuk memulai proses analisis, pertama-tama Anda harus menentukan nama dan lokasi proses dan Data Repository. Setelah itu, Anda bisa mulai merancang sebuah analisis baru.
- 2. Open Recent Process : Aksi ini berguna untuk membuka proses yang baru saja ditutup. Selain aksi ini, Anda juga bisa membuka proses yang baru ditutup dengan mengklik dua kali salah satu daftar yang ada pada Recent Process. Kemudian tampilan Welcome Perspective akan otomotasi beralih ke Design Perspective.
- Open Process : Aksi ini untuk membuka Repository Browser yang berisi daftar proses. Anda juga bisa memilih proses untuk dibuka pada Design Perspective.
- 4. **Open Template** : Aksi ini menunjukkan pilihan lain yang sudah ditentukan oleh proses analisis.

5. Online Tutorial : Aksi digunakan untuk memulai tutorial secara online (terhubung internet). Tutorial yang dapat secara langsung digunakan dengan RapidMiner ini, memberikan perkanalan dan beberapa data mining. konsep Hal ini direkomendasikan untuk Anda yang sudah memiliki pengetahuan dasar mengenai data mining dan sudah akrab dengan operasi dasar RapidMiner.

RapidMiner dapat menampilkan beberapa view pada saat bersamaan. Seperti yang ditunjukkan pada Gambar 7, pada tampilan Welcome Perspective tedapat **Welcome view** dan **Log View**. Ukuran dari setiap view tersebut dapat diubah sesuai dengan kebutuhan Anda dengan Mengklik dan menarik garis batas diantara keduanya ke atas atau ke bawah.



Gambar 2.7 Welcome Perspective



Gambar 2.8 Header Tab

Anda bisa melakukan beberapa aksi terhadap view, dengan mengklik salah satu ikon yang tampak pada bagian view, seperti yang ditunjukkan pada gambar 2.8. Berikut ini beberapa aksi yang dapat Anda lakukan:

- Close : Aksi ini untuk menutup view yang ditampilkan pada perspective. Anda bisa menampilkan view kembali dengan mengklik menu view dan memilih view yang ingin ditampilkan.
- 2. *Maximize* : Aksi ini untuk memperbesar ukuran view pada perspective.
- 3. *Minimize* : Aksi ini untuk memperkecil ukuran view pada perspective.
- Detach : Aksi ini untuk melepaskan view dari perspective menjadi jendela terpisah, kemudian Anda juga dapat memindahkannya sesuai dengan keinginan Anda.

Design Perspective

Design Perspective merupakan lingkungan kerja RapidMiner. Dimana Design Perspective ini merupakan perspective utama dari RapidMiner yang digunakan sebagai area kerja untuk membuat dan mengelola proses analisis. Seperti yang ditunjukkan pada Gambar 2.10, perspective ini memiliki beberapa view dengan fungsinya masing-masing yang dapat mendukung Anda dalam melakukan proses analisis data mining. Anda bisa mengganti perspective dengan mengklik salah satu ikon dari tollbar perspective yang sebelumnya telah dijelaskan. Selain dengan cara tersebut, Anda juga bisa mengganti perspective dengan mengklik menu view, kemudian pilih perspective, lalu pilih perspective yang ingin Anda tampilkan.



Gambar 2.9 Tampilan Design Perspective

Sebagai lingkungan lingkungan kerja, Design Perspective memiliki beberapa view. Berikut ini beberapa view yang ditampilkan pada Design Perspective:

1. Operator View

Operator View merupakan view yang paling penting pada perspective ini. Semua operator atau langkah kerja dari RapidMiner disajikan dalam bentuk kelompok hierarki di Operator View ini sehingga operator-operator tersebut dapat digunakan pada proses analisis, seperti yang ditunjukkan pada Gambar 2.10. Hal ini akan memudahkan Anda dalam mencari dan menggunakan operator yang sesuai dengan kebutuhan Anda. Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut:

- Process Control : Operator ini terdiri dari operator perulangan dan percabangan yang dapat mengatur aliran proses.
- Utility : Operator bantuan, seperti operator macros, loggin, subproses, dan lain-lain.
- Repository Access : Kelompok ini terdiri dari operator-operator yang dapat digunakan untuk membaca atau menulis akses pada repository.
- Import : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk membaca data dan objek dari format tertentu seperti file, database, dan lain-lain.
- Export : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk menulis data dan objek menjadi format tertentu.

- Data Transformation : kelompok ini terdiri dari semua operator yang berguna untuk transformasi data dan meta data.
- Modeling : kolompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru.
- Evaluation : kelompok ini berisi operator yang dapat digunakan untuk menghitung kualitas pemodelan dan untuk data baru.



Gambar 2.10 Kelompok Operator dalam Bentuk Hierarki

2. Repository View

Repository View merupakan komponen utama dalam Design Perspective selain Operator View. View ini dapat Anda gunakan untuk mengelola dan menata proses Analisis Anda menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.

3. Process View

menunjukkan langkah-langkah View Process tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Anda dapat menambahkan langkah baru dengan beberapa cara. hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali. bekerja dengan RapidMiner ialah Pada dasarnya mendefinisikan proses analisis. yaitu dengan menunjukkan serangkaian langkah kerja tertentu. Dalam RapidMiner, komponen proses ini dinamakan RapidMiner sebagai operator. Operator pada didefinisikan sebagai beikut:

- * Deskripsi dari input yang diharapkan.
- * Deskripsi dari output yang disediakan.
- Tindakan yang dilakukan oleh operator pada input, yang akhirnya mengarah dengan penyediaan output.
- Sejumlah parameter yang dapat mengontrol action performed.

4. Parameter View

Beberapa operator dalam RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar. Namun terkadang parameter tidak mutlak dibutuhkan, meskipun eksekusi operator dapat dikendalikan dengan menunjukkan nilai parameter tertentu. Parameter view memiliki toolbar sendiri sama seperti view-view yang lain. Pada Gambar 2.12, Anda dapat melihat bahwa pada Parameter View ini terdapat beberapa ikon dan nama-nama operator terkini yang dikuti dengan aktual parameter.



Gambar 2.11 Tampilan Parameter View

Huruf tebal berarti bahwa parameter mutlak harus didefinisikan oleh analis dan tidak memiliki nilai default. Sedangkan huruf miring berarti bahwa parameter diklasifikasikan sebagai parameter ahli dan seharusnya tidak harus diubah oleh pemula untuk analisis data. Poin pentingnya ialah beberapa parameter hanya ditunjukkan ketika parameter lain memiliki nilai tertentu.

5. Help & Comment View

Setiap kali Anda memilih operator pada Operator View atau Process View, maka jendela bantuan dalam Help View akan menunjukkan penjelasan mengenai operator ini. Penjelasn yang ditampilkan dalam Help View meliputi:

- Sebuah penjelasan singkat mengenai fungsi operator dalam satu atau beberapa kalimat.
- Sebuah penjelasan rinci mengenai fungsi operator.
- Daftar semua parameter termasuk deskripsi singkat dari parameter, nilai default (jika tersedia), petunjuk apakah parameter ini adalah parameter ahli serta indikasi parameter dependensi.

Sedangkan Comment View merupakan area bagi Anda untuk menuliskan komentar pada langkahlangkah proses tertentu. Untuk membuat komentar, Anda hanya perlu memilih operator dan menulis teks di atasnya dalam bidang komentar. Kemudian komentar tersebut disimpan bersama-sama dengan definisi proses Anda. Komentar ini dapat berguna untuk melacak langkah-langkah tertentu dalam rancangan nantinya.

6. Problem & Log View

Problem View merupakan komponen yang sangat berharga dan merupkan sumber bantuan bagi Anda selama merancang proses analisis. Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam Problem View, seperti yang ditunjukkan pada Gambar 2.13.



Gambar 2.12 Problem & Log View

Pada kolom Message, Anda akan menemukan ringkasan pendek dari masalah. Kolom Location berisi tempat di mana masalah muncul dalam bentuk nama Operator dan nama port input yang bersangkutan. Kolom Fixes memberikan gambaran dari kemungkinan solusi tersebut, baik secara langsung sebagai teks (jika hanya ada satu kemungkinan Solusi) atau sebagai indikasi dari berapa banyak kemungkinan yang berbeda untuk memecahkan masalah.

Cara Menggunakan Repositori

Repositori merupakan Tabel, database, koleksi teks, yang kita miliki untuk dapat digali datanya untuk mendapatkan informasi yang kita inginkan. Ini merupakan awal dari seluruh proses Data Mining. Maka dari itu adalah penting bagi kita untuk mengetahui cara menggunakan repository.

Sample Data Repository

RapidMiner menyediakan contoh database yang dapat digunakan, berikut cara menggunakan Sample Data Repository.



Gambar 2.13 Kumpulan Sample Data Repository

Pada bagian Repositori terdapat 3 buah lokasi repositori, yakni Samples, DB dan Local Repository.

Untuk mengambil Sample Data Repository, buka hirarki Samples, masuk ke folder Data. Sehingga seperti gambar berikut.



Gambar 2.14 Tampilan Design Perspective Awal

Lakukan Drag dan Drop salah satu Example Repository. Kita ambil contoh Golf. Tarik dan lepaskan repository ke dalam Main Process, sehingga seperti gambar berikut.



Gambar 2.15 Repository berada dalam Main Process



Gambar 2.16 Menghubungkan Output Repositori ke Result

Hubungkan output pada Database ke Result seperti Gambar diatas. Lalu klik ikon Play ▶. Gambar 2.17 adalah Sample data repository dari Golf. Coba lakukan untuk memasukkan Sample Repository yang lain.

🛛 🐺 Resi	ult Overview	🗙 🖊 📳 Exa	mpleSet (Retrie	eve Golf) 🏾	
Data Vie	w 🔘 Meta D	ata View 🔘 F	Plot View 🔘 A	dvanced Cha	rts 🔘 Annotatior
ExampleSe	t (14 example	es, 1 special at	tribute, 4 regula	ar attributes)	
Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

Gambar 2.17 Isi Sample Golf Data Repository

Import Repository

Dibanyak kesempatan lain, kita akan selalu menggunakan database yang kita miliki. RapidMiner menyediakan layanan pengguna agar dapat mengimport database miliknya. Namun, tidak seperti kebanvakan tools Data Mining Lain. RapidMiner memiliki kelebihan tersendiri yakni dapat langsung melakukan import file dengan ekstensi .xls atau .xlsx, yakni file dari Microsoft Excel, Program yang relatif sering digunakan oleh pengguna. Berikut adalah cara untuk melakukan import file Microsoft Excel.

Lihat pada bagian Repository. Klik pada ikon import seperti gambar 2.18. Seperti yang dapat kita lihat, ada beberapa ekstensi file yang dapat kita masukkan kedalam repository kita. CSV File, Excel Sheen File, Access Database Table File, Database Table, Binary File. Namun pada Dasarnya cara melakukan import pada semua file ini sama. Sebagai contoh, pilih Import Excel Sheet.



Gambar 2.18 Repository

Setelah itu, akan muncul window baru yakni Step 1 dari 5 Step Data import Wizard. Disini akan diarahkan oleh RapidMiner bagaimana langkah untuk melakukan import data.



Gambar 2.19 Step 1 of 5 Import Wizard

Cari file excel kalian dengan klik pada bagian Look in

Look In: Documents

yang dibutuhkan lalu Klik tombol Next 🗩 🕬.

Berikutnya pada Step 2 ialah, pilih Sheet yang akan dimasukkan. Pada dasarnya, Repository RapidMiner hanya menyediakan 1 repositori untuk 1 buah table.

-	Data import wizard - Step 2 of 5						
"At	This wizard guides you to import your data. Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into Rapidkliner. Furthermore, you can mark a range of cells to be loaded.						
Sheet1							
A	В	С	D	E			
OUTLOOK	TEMPERATU						
SUNNY							
SUNNY							
OVERCAST							
RAIN			false				
RAIN			false				
RAIN							
OVERCAST							
SUNNY							
SUNNY							
RAIN							
	Previous Next Prinish Scancel						

Gambar 2.20 Step 2 of 5 Import Wizard

Klik tombol Next 🖃 🕬. Berikutnya ialah memberikan anotasi. Jika data kita tidak memiliki nama attribute, tidak usah melakukan apa-apa pada step 3 ini.

Image: System 20 Data import wizard - Step 3 of 5 This wizard guides you to import your data. Step 3: In Rapid/Imre, each attribute can be annotated. The most important annotation of antiribute is in the most import the attributes. If your data does not contain attribute names, do not set this property. If thref annotations are contained in the rows of your data file, you can assign them here.							
A	В	С	D	E			
OUTLOOK	TEMPERATU	HUMIDITY	WINDY	PLAY			
SUNNY	85	85	false	Don't Play			
SUNNY	80	90	true	Don't Play PLAY			
OVERCAST	83	78	false	Play			
RAIN	70	96	false	Play			
RAIN	68	80	false	Don't Play			
RAIN	65	70	true	Play			
OVERCAST	64	65	true	Don't Play			
SUNNY	72	95	false	Play			
SUNNY	69	70	false	Play	L		
ΡΔΙΝΙ	75	80	foleo	Play	\mathbf{r}		
	annotation of names of the his property. can assign th A DUTLOOK SUNNY SUNNY OVERCAST RAIN RAIN OVERCAST SUNNY SUNNY SUNNY RAIN	anotation of an attribute is ames of the attributes. If y his property. If further anno can assign them here. A B JUTLOOK TEMPERATU SUNNY 80 JUTLOK TEMPERATU SUNNY 80 JUTLOK TEMPERATU SUNNY 80 JUTLOK 80 RAIN 65 JUTLOK 65 JUTLOK 65 JUTLOK 70 RAIN 65 JUTLOK 70 RAIN 65 JUTLOK 70 RAIN 70 R	annotation of an attributes is the name - at a tributes. If your data does this property. If further annotations are color an assign them here. A B C JUTLOOK TEMPERATI. HUMIDITY. SUNNY 85 85 SUNNY 80 90 OVERCAST 83 78 RAIN 70 96 RAIN 68 80 OVERCAST 64 65 SUNNY 72 95 SUNNY 75 90 <td>anotation of an attribute is its name - a row with this 2 manes of the attributes. If your data does not contain a his property. If further anotations are contained in the can assign them here. A B C D OUTLOOK TEMPERATL HUMIDITY WINDY SUNNY 85 85 false SUNNY 85 90 true OVERCAST 83 78 false RAIN 70 96 false RAIN 65 70 true OVERCAST 64 65 true SUNNY 72 95 false SUNNY 72 95 false SUNNY 72 97 false SUNNY 75 90 false</td> <td>anotation of an attribute is its name - a row with this annotation defines the tames of the attributes. If your data does not contain attribute ames, do not set the property. If further annotations are contained in the rows of your data file, you can assign them here.</td>	anotation of an attribute is its name - a row with this 2 manes of the attributes. If your data does not contain a his property. If further anotations are contained in the can assign them here. A B C D OUTLOOK TEMPERATL HUMIDITY WINDY SUNNY 85 85 false SUNNY 85 90 true OVERCAST 83 78 false RAIN 70 96 false RAIN 65 70 true OVERCAST 64 65 true SUNNY 72 95 false SUNNY 72 95 false SUNNY 72 97 false SUNNY 75 90 false	anotation of an attribute is its name - a row with this annotation defines the tames of the attributes. If your data does not contain attribute ames, do not set the property. If further annotations are contained in the rows of your data file, you can assign them here.		

Gambar 2.21 Step 3 of 5 Import Wizard

Klik tombol Next 🗩 . Step ke 4 adalah memberikan tipe data pada tabel kita. Sebenarnya Rapid Miner akan memberikan tipe data yang tepat secara otomatis.

Beload data	Duess value types	Preview uses only	first 100 rows. 🚊ale for	mat	
2	ø	Ø	Ø	Ø	
OUTLOOK	TENPERATURE	HUVIDITY	WINDY	PLAY	
polynominal +	integer +	Integer +	binominal +	binominzi =	
atribult +	attribute +	attribute +	attribute +	attribute +	
SUNNY	85	80	talse	Con 1 Play	
SUNW	80	90	tue	Dom't Play	
OVERCAST	83	70	tates	Pter .	
RAN	70	96	talse	Play	
RAN	60	80 talse		Don't Play	
0				D	
😳 0 enors.			1	🛛 Jancre errors 🔲 Show only gro	
Bow Column		Error Original value		Message	

Gambar 2.22 Step 4 of 5 Import Wizard

Namun, jika kita merasa tipe data yang diberikan RapidMiner tidak cocok, kita bisa mengubahnya.



Gambar 2.23 Tipe Data

Klik tombol Next 🕬 Step ke 5 adalah memasukkan database kita kedalam repository. Disarankan untuk memasukkannya kedalam Local Repository untuk memudahkan kita mencarinya. Jangan lupa untuk memberikan nama repository kita.



Gambar 2.24 Step 5 of 5 Import Wizard

Kemudian klik tombol finish Prinish

Hasil Import Repository akan terlihat pada bagian Repository seperti dalam gambar 2.25.



Gambar 2.25 Repository yang sudah diimport

Untuk melihat isi dari repository kita, hubungkan output pada repository kearah result seperti gambar 2.26.



Gambar 2.26 Menghubungkan Output Repositori pada Result

klik ikon Play 🖻. Dan berikutnya akan muncul isi dari tabel yang kalian miliki.

🛛 🐺 Resu	lt Overview 🖇	👔 🗐 Exan	npleSet (//Loc	al Repository/	data/golfdatası	et) 🕺 🔪
Data Vie	w 🔘 Meta Da	ita View 🔘 P	lot View 🔘 A	dvanced Chai	ts 🔘 Annotat	ions
ExampleSet	(14 examples	, 0 special att	ributes, 5 regu	ular attributes)		
Row No.	OUTLOOK	TEMPERAT	HUMIDITY	WINDY	PLAY	
1	SUNNY	85	85	false	Don't Play	
2	SUNNY	80	90	true	Don't Play	
3	OVERCAST	83	78	false	Play	
4	RAIN	70	96	false	Play	
5	RAIN	68	80	false	Don't Play	
6	RAIN	65	70	true	Play	
7	OVERCAST	64	65	true	Don't Play	
8	SUNNY	72	95	false	Play	
9	SUNNY	69	70	false	Play	
10	RAIN	75	80	false	Play	
11	SUNNY	75	70	true	Play	
12	OVERCAST	72	90	true	Play	
13	OVERCAST	81	75	false	Play	
14	RAIN	71	80	true	Don't Play	
14	TV WIN	1.1	00	uuc	Don'thay	

Gambar 2.27 Tabel Repository

Bagian Dua Data Mining

Pengenalan Data Mining

Pengenalan Decision Tree

Pengenalan Neural Network

Pengenalan Market Basket Analysis

Chapter 3

Data Mining

Mengenal Data Mining

Pengertian Data Mining

Sebelum kita mulai, ayo kita coba beberapa eksperimen sebagai berikut.

- Pilih angka antara 1 sampai 10
- Kalikan dengan angka 9
- Hasil dari perkalian tersebut jumlahkan masingmasing angkanya
- Kalikan hasil dengan 4
- Bagi dengan 3
- Kurangi dengan 2

Jawabannya adalah 2. Kebetulan? Sebagai seorang analis, pasti jawabannya adalah tidak.

Bagaimana dengan kejadian acak lainnya, seperti "lempar koin." Tentu jika temanmu menebak secara langsung dan hasil dari kejadian tersebut ternyata tepat seperti yang temanmu tebak, kau pasti akan mengatakan bahwa itu merupakan kebetulan.

Kita ambil satu contoh sederhana lagi. Terdapat kejadian seperti: Seseorang menjatuhkan sebuah gelas dari ketinggian tertentu. Detik pertama orang tersebut menjatuhkan gelasnya, kau pasti akan mengatakan dengan pasti bahwa gelas tersebut akan pecah, padahal hukum fisika belum menunjukkan proses penghancuran gelas tersebut ketika bersentuhan dengan tanah. Dan lagi, tebakanmu itu dikatakan bukanlah kebetulan. Jadi secara logika, bagaimana kau tahu dengan sangat tepat hasil dari kejadian tersebut? Bukankah kondisinya sama seperti kejadian "lempar koin" sebelumnya?

Jadi apakah yang kita lakukan dalam otak kita? Kita mempertimbangkan karakteristik-karakteristik dari kejadian ini. Pada kasus gelas yang jatuh, kita dengan cepat mengetahui karakteristik penting dari serangkaian kejadian tersebut, bahan gelas, ketinggian, tipe pijakan, dan lain-lain. Kemudian kita menjawab dengan cepat berdasarkan analogi, contohnya kita kita

membuat perbandingan dengan kejadian gelas atau cangkir atau piring yang jatuh sebelumnya. Berarti dua diperlukan hal yang adalah: pertama, kita membutuhkan data dari kejadian-kejadian sebelumnya, dan kedua, seberapa mirip kejadian yang di tempat dengan kejadian sebelumnya. Kita bisa membuat estimasi atau prediksi dengan mencari kejadian yang paling mirip dengan kejadian di tempat. Karena kita lebih sering melihat bahwa benda berbahan kaca dijatuhkan akan pecah, maka secara otomatis inilah yang menjadi prediksi kita.

Bagaimanapun, prosedur diatas tidak cocok untuk kejadian "lempar koin." Ini disebabkan terdapat lebih banyak faktor yang harus dipertimbangkan, ada yang sulit dan ada yang tidak bisa diukur. Belum lagi kita harus dapat memikirkan proses kejadian menuju hasil dengan baik, memikirkan analogi yang paling cocok dengan kejadian untuk melakukan prediksi. Ditambah "lempar koin" memiliki kondisi yang dapat berubah-ubah tiap kejadiannya dan berlangsung cepat, ini berarti perhitungan juga harus dilakukan secara cepat. Mustahil untuk seorang manusia? Benar. Tetapi tidak mustahil untuk metode data mining.

Data Mining adalah serangkaian proses untuk menggali nlai tambah dari suatu kumpulan data

berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006)

Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya. (Pramudiono, 2006)

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005)

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. (Larose, 2005)

Kata *Mining* merupakan kiasan dari bahasa inggris, mine. Jika mine berarti menambang sumber daya yang tersembunyi di dalam tanah, maka Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), machine learning, statistik dan basis Data.

Pengelompokan Teknik Data Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan popular adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur herhirarki

Association

Digunakan untuk mengenali kelakuan dari kejadiankejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya adalah Market Basket Analysis, yaitu salah sati metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan.

Clustering

Digunakan menganalisis pengelompokkan untuk berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokkan belum didefinisikan sehelum dijalankannya tool data mining. Biasanya menggunkan metode *neural network* atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool data mining.

Decision Tree

Mengenal Decision Tree

diketahui bahwa selalu Seperti manusia menghadapi berbagai macam masalah dalam di sehari-hari. kehidupannya Masalah-masalah yang timbul dari berbagai macam bidang ini memiliki tingkat kesulitan dan kompleksitas yang sangat bervariasi, mulai dari masalah yang sangat sederhana dengan sedikit faktor-faktor terkait hingga masalah yang sangat rumit dengan banyak sekali faktor-faktor yang terkait, sehingga factor-faktor yang berkaitan dengan masalah tersebut perlu untuk diperhitungkan.

Seiring dengan perkembangan kemajuan pola pikir manusia, manusia mulai mengembangkan sebuah sistem yang dapat membantu manusia dalam menghadapi masalah-masalah yang timbul sehingga dapat menyelesaikannya dengan mudah.

keputusan atau yang lebih dikenal Pohon dengan istilah Decision Tree ini merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebutdengan memperhitungkan berbagai macam faktor yang berkaitan di dalam lingkup masalah tersebut.

Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah sehingga dengan memperhitungkan faktor-faktor tersebut dapat dihasilkan penyelesaian terbaik untuk masalah tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah.

Pohon keputusan dalam analisis pemecahan masalah pengambilan keputusan merupakan pemetaan alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon keputusan juga memperlihatkan faktor-faktor kemungkinan yang dapat mempengaruhi alternative-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut.

Secara umum, pohon keputusan adalah suatu gambaran permodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah kepada solusi yang dihasilkan. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Seiring dengan perkembangannya, pohon keputusan kini telah banyak dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang prediksi. (Han, J., & Kamber, M. (2006). Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.)



Gambar 4.1 Bentuk Decision Tree Secara Umum

Algoritma c4.5

Pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Seperti yang telah dijelaskan sebelumnya, klasifikasi merupakan Suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Metode ini popular karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Banyak algoritma yang dapat digunakan untuk membangun suatu decision tree, salah satunya ialah algoritma C45.

Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C.45 menggunakan rasio perolehan (gain ratio). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropi.

Konsep Entropy

Entropy(S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropi digunakan untuk mengukur ketidakaslian S.



Gambar 4.2 Grafik Entropi

Besarnya Entropy pada ruang sampel S didefinisikan dengan:

 $Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$

Dimana:

- S : ruang (data) sampel yang digunakan untuk pelatihan
- p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu
- p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

- Entropi(S) = 0, jika semua contoh pada S berada dalam kelas yang sama.
- Entropi(S) = 1, jika jumlah contoh positif dan negative dalam S adalah sama.
- 0 > Entropi(S) > 1, jika jumlah contoh positif dan negative dalam S tidak sama.

Konsep Gain

Gain (S,A) merupakan Perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

$$Gain(S, A) \equiv Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$$

Dimana:

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- |S_i| : Jumlah sampel pada pertisi ke --i
- |S| : Jumlah sampel dalam S
Untuk memudahkan penjelasan mengenai algoritma C4.5berikut ini disertakan contoh kasus yang dituangkan dalam Tabel 4.1:

No	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Tabel 4.1 Keputusan Bermain Tenis

Tabel 1 merupakan kasus yang akan dibuat pohon keputusan untuk menentukan main tenis atau tida. Data ini memiliki atribut-atribut yaitu, keadaan cuaca (outlook), temperatur, kelembaban (humidity) dan keadaan angin (windy).

Berikut merupakan cara membangun pohon keputusan dengan menggunakan algoritma:

- 1. Pilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut-atribut yang ada.
- 2. Buat cabang untuk masing-masing nilai
- 3. Bagi kasus dalam cabang
- 4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

NODE			JUMLAH KASUS	NO (S ₁)	YES (S ₂)	ENTROPY	GAIN
1	TOTAL		14	4	10	0.863120569	
	OUTLOOK						0.258521037
		CLOUDY	4	0	4	0	
		RAINY	5	1	4	0.721928095	
		SUNNY	5	3	2	0.970950594	
	TEMPERATURE						0.183850925
		COOL	4	0	4	0	
		HOT	4	2	2	1	
		MILD	6	2	4	0.918295834	
	HUMIDITY						0.370506501
		HIGH	7	4	3	0.985228136	
		NORMAL	7	0	7	0	
	WINDY						0.005977711
		FALSE	8	2	6	0.811278124	
		TRUE	6	4	2	0.918295834	

Tabel 4.2 Perhitungan Simpul 1



Dari hasil pada Tabel 4.2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah HUMIDITY yaitu sebesar 0.37. Dengan demikian HUMIDITY dapat menjadi node akar.

Ada 2 nilai atribut dari HUMIDITY yaitu HIGH dan NORMAL. Dari kedua nilai atribut tersebut, nilai atribut NORMAL sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya Yes, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut HIGH masih perlu dilakukan perhitungan lagi hingga semua kasus masuk dalam kelas seperti yang terlihat pada Gambar di sebelah kanan.

Kelebihan Pohon Keputusan

Dalam membuat keputusan dengan menggunankan pohon keputusan, metode ini memiliki kelebihan sebagai berikut:

- Daerah pengambilan keputusan lebih simpel dan spesifik.
- Eliminasi perhitungan-perhitungan tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
- Fleksibel untuk memilih fitur dari internal node yang berbeda. Sehingga dapat meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
- Dengan menggunakan pohon keputusan, penguji tidak perlu melakukan estimasi pada distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Karena metode ini menggunakan kcriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Kekurangan Pohon Keputusan

Pohon keputusan sangat membantu dalam pengambilan keputusan, namun pohon keputusan juga memiliki beberapa kekurangan, diantaranya:

- Kesulitan dalam mendesain pohon keputusan yang optimal.
- Hasil kualitas keputusan yang didapat sangat tergantung pada bagaimana pohon tersebut didesain. Sehingga jika pohon keputusan yang dibuat kurang optimal, maka akan berpengaruh pada kualitas dari keputusan yang didapat.
- Terjadi overlap terutama ketika kelas-kelas dan criteria yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
- Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.

Decision Tree pada RapidMiner

RapidMiner sebagai software pengolah data mining menyediakan tool untuk membuat decision tree. Hal ini tentu akan memudahkan kita membuat decision tree dengan menggunakan RapidMiner dibandingkan membuat decision tree secara manual yaitu dengan melakukan perhitungan menggunakan algoritma C4.5 yang telah dijelaskan sebelumnya.

Contoh Kasus:

Keputusan Bermain Tenis

Pada contoh kali ini, kita akan membuat keputusan bermain tenis atau tidak. Untuk memudahkan dalam menggunakan RapidMiner untuk membuat decision tree, kita gunakan data sederhana yang ada pada sub bab decision tree. Pertama-tama data pada tabel 2 dibuat lagi dalam format excel seperti yang terlihat pada Gambar 4.3.

	А	В	С	D	E	F
1	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY	
2	Sunny	Hot	High	No	Don't Play	
3	Sunny	Hot	High	Yes	Don't Play	
4	Cloudy	Hot	High	No	Play	
5	Rainy	Mild	High	No	Play	
6	Rainy	Cool	Normal	No	Play	
7	Rainy	Cool	Normal	Yes	Play	
8	Cloudy	Cool	Normal	Yes	Play	
9	Sunny	Mild	High	No	Don't Play	
10	Sunny	Cool	Normal	No	Play	
11	Rainy	Mild	Normal	No	Play	
12	Sunny	Mild	Normal	Yes	Play	
13	Cloudy	Mild	High	Yes	Play	
14	Cloudy	Hot	Normal	No	Play	
15	Rainy	Mild	High	Yes	Don't Play	
16						

Gambar 4.3 Tabel Keputusan dalam Format xls

Setelah data yang kita punya dibuat dalam bentuk tabel format xls, selanjutnya lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada Bab 2. Lalu cari table Microsoft Excel yang telah dibuat dan masukan kedalam Local Repository seperti yang terlihat pada Gambar 4.4.



Gambar 4.4 Lokasi Tabel pada Repository

Lakukan Drag dan Drop Tabel PlayGolf kedalam Process view. Sehingga Operator Database muncul dalam View Proses seperti pada Gambar 4.5. Pada view Process, tabel PlayGolf yang dimasukkan ke dalam proses akan dijadikan sebagai Operator Retrieve.



Gambar 4.5 Repository PlayGolf pada Main Process

Untuk membuat decision tree dengan menggunakan RapidMiner, kita membutuhkan operator Decision tree, operator ini terdapat pada View Operators. Untuk menggunakannya pilih Modelling pada View Operator, lalu pilih Classification and Regression, lalu pilih Tree Induction dan pilih Decision Tree.



Gambar 4.6 Daftar Operator pada View Operators

Setelah menemukan operator Decision Tree, seret (*drag*) operator tersebut lalu letakkan (*drop*) ke dalam view Process. Kemudian susun posisinya disamping operator Retrieve, seperti yang tampak pada Gambar 4.7.



Gambar 4.7 Posisi Operator Decision Tree

Selanjutnya, hubungkan operator Retrieve dengan operator Decision Tree dengan menarik garis dari tabel PlayGolf ke operator Decision Tree dan menarik garis lagi dari operator Decision Tree ke result di sisi kanan, seperti yang tampak pada Gambar 4.8. Operator Decision Tree berguna untuk memperdiksikan keputusan dari atribut-aribut yang dimasukkan ke dalam operator retrieve. Dengan mengubah tabel (atribut) yang dimasukkan menjadi sebuah pohon keputusan.

6° Process X ▼ XML X 4= + ⇒ + ↑ ■ Root >	<i>d</i> • 🗉 🛊 🛛 🕹
inp) Main Process Retrieve Play out DecisionTree tra mod era	() res (res

Gambar 4.8 Menghubungkan Tabel Playgolf dengan Operator Decision Tree

Pada operator Decision tree terdapat input training set (tra), port ini merupakan output dari operator retrieve. Output dari operator lain juga dapat digunakan oleh port ini. Port ini menghasilkan ExampleSet yang dapat diperoses menjadi decision tree. Selain itu pada operator ini juga terdapat output model (mod) dan example set (exa). Mod akan mengonversi atribut yang dimasukkan menjadi mpdel keputusan dalam bentuk decision tree. exa merupakan port yang menghasilkan output tanpa mengubah inputan yang masuk melalui port ini. Port ini biasa digunakan untuk menggunakan kembali sama ExampleSet di operator lebih lanjut atau untuk melihat ExampleSet dalam Hasil Workspace.

Langkah selanjutnya ialah mengatur parameter sesuai dengan kebutuhan kita. Setelah menghubungkan operator retrieve dengan operator decision tree, atur parameter decision tree seperti pada gambar 4.9.

🛛 🛃 Parameters 💈	Context 🗶				
🚨 🖷 🖷 🔛 🛛	🕺 📑 👻				
💡 De	cision Tree				
criterion	gain_ratio 💌	Â			
minimal size for split 2					
minimal leaf size 1					
minimal gain	0.1				
maximal depth	20				
confidence	0.25				
number of preprunin	.3				
no pre pruning					
no pruning		~			

Gambar 4.9 Parameter Decision Tree





- Criterion, berguna memilih kriteria untuk menetapkan atribut sebagai akar dari decision tree. kriteria yang dapat dipilih, antara lain
 - Gain ratio merupakan varian dari information_gain. Metode ini menghasilkan information gain untuk

setiap atribut yang memberikan nilai atribut yang seragam

- 2. Information_gain, dengan metode ini, semua entropi dihitung. Kemudian atribut dengan entropi minimum yang dipilih untuk dilakukan perpecahan pohon (split). Metode ini memiliki bias dalammemilih atribut dengan sejumlah besar nilai.
- Gini_index merupakan ukuran ketidakaslian dari suatu ExampleSet. Metode ini memisahkan pada atribut yang dipilih memberikan penurunan indeks gini rata-rata yang dihasilkan subset.
- Accuracy, metode ini memimilih beberapa atribut untuk memecah pohon (split) yang memaksimalkan akurasi dari keseluruhan pohon.
- Minimal size of split, Ukuran untuk membuat simpul-simpul pada decision tree. simpul dibagi berdasarkan ukuran yang lebih besar dari atau sama dengan parameter Minimal size of split. Ukuran simpul adalah jumlah contoh dalam subset nya

- Minimal leaf size, Pohon yang dihasilkan sedemikian rupa memiliki himpunan bagian simpul daun setidaknya sebanyak jumlah minimal leaf size.
- Minimal gain merupakan nilai gain minimal yang ditentukan untuk menghasilkan simpul pohon keputusan. Gain dari sebuah node dihitung sebelum dilakukan pemecahan. Node dipecah jika gain bernilai lebih besar dari Minimal Gain yang ditentukan. Nilai minimal gain yang terlalu tinggi akan mengurangi perpaecahan pohon dan menghasilkan pohon yang kecil. Sebuah nilai yang terlalu tinggi dapat mencegah pemecahan dan menghasilkan pohon dengan simpul tunggal.
- Maximal depth, Parameter ini digunakan untuk • membatasi ukuran Putusan Pohon. Proses ketika generasi pohon tidak berlanjut kedalaman pohon adalah sama dengan kedalaman maksimal. Jika nilainya diatur ke '-1', parameter kedalaman maksimal menempatkan tidak terikat pada kedalaman pohon, pohon kedalaman maksimum dihasilkan. Jika nilainya diatur ke '1 ' maka akan dihasilkan pohon dengan simpul tunggal.

- Confidence, Parameter ini menentukan tingkat kepercayaan yang digunakan untuk pesimis kesalahan perhitungan pemangkasan.
- number of prepruning alternatives. Parameter ini menyesuaikan jumlah node alternatif mencoba untuk membelah ketika split dicegah dengan prepruning pada simpul tertentu.
 - no prepruning, Secara default Pohon Keputusan yang dihasilkan dengan prepruning. Menetapkan parameter ini untuk menonaktifkan benar prepruning dan memberikan pohon tanpa prepruning apapun.
 - no pruning Secara default Pohon Keputusan yang dihasilkan dengan pemangkasan. Menetapkan parameter ini untuk menonaktifkan benar pemangkasan dan memberikan sebuah unpruned

Setelah parameter diatur, klik ikon Run pada toolbar, seperti pada gambar 40 untuk menampilkan hasilnya. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 4.11 Ikon Run

Setelah beberapa detik maka RapidMiner akan menampilkan hasil keputusan pada view Result. Jika kita pilih Graph view, maka akan ditampilkan hasilnya berbentuk pohon keputusan seperti pada gambar 4.12. Hasil pohon keputusan dapat disimpan dengan mengklik save image pada sisi kiri View Result.



Gambar 4.12 Hasil Berupa Graph Pohon Keputusan

Selain menampilkan hasil decision tree berupa graph atau tampilan pohon keputusan, RapidMiner juga menyediakan tool untuk menampilkan hasil berupa teks view dengan mengklik button Text View seperti yang tampak pada Gambar 4.13.

🛛 🛒 Result Overview 🕱 🎧 Tree (Decision Tree) 🚿
Graph View . Text View: Annotations
Tree
HUMIDITY = High
OUTLOOK = Cloudy: Play {Don't Play=0, Play=2}
OUTLOOK = Rainy
WINDY = No: Play {Don't Play=0, Play=1}
WINDY = Yes: Don't Play {Don't Play=1, Play=0}
OUTLOOK = Sunny: Don't Play {Don't Play=3, Play=0}
HUMIDITY = Normal: Play {Don't Play=0, Play=7}

Gambar 4.13 Hasil Berupa Penjelasan Teks

Contoh Kasus : Keputusan seseorang mempunyai potensi menderita hipertensi

Sebelumnya kita telah mengetahui bagaimana pohon membuat keputusan untuk menentukan bermain tenis dengan menggunakan operator decision tree. Pada pembahasan kali ini kita akan membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi atau tidak. Untuk menambah pengeatahuan kita mengenai kegunaan operator yang ada pada RapidMiner, oleh karena itu untuk membuat pohon keputusan kali ini kita menggunakan operator X-Validation, Apply Model dan Performance. Selain itu, kita juga tetap menggunakan operator decision tree dalam pembuatan pohon keputusan kali ini.

Cara yang digunakan dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit hipertensi, tidak jauh berbeda dengan cara membuat pohon keputusan yang sebelumnya, yaitu pertama-tama import data ke dalam repository RapidMiner, lalu lakukan drag dan drop data tersebut pada view process untuk mengubah data yang berisi atribut pohon keputusan menjadi operator retrieve. setelah itu, lakukan drag dan drop operator decision tree ke dalam view process dengan cara yang sama seperti penjelasan sebelumnya.

	А	В	С	D	E
1	Usia	Berat	Kelamin	Hipertensi	
2	Muda	Overweight	Pria	Ya	
3	Muda	Underweight	Pria	Tidak	
4	Muda	Average	Wanita	Tidak	
5	Tua	Overweight	Pria	Tidak	
6	Tua	Overweight	Pria	Ya	
7	Muda	Underweight	Pria	Tidak	
8	Tua	Overweight	Wanita	Ya	
9	Tua	Average	Pria	Tidak	
10					

Gambar 4.14 Tabel SakitHipertensi dalam format xls



Gambar 4.15 Lokasi Tabel pada Repository



Gambar 4.16 Tabel SakitHipertensi pada Main Process

Untuk membuat pohon keputsan kali ini kita menggunakan operator X-Validation. Operator ini melakukan validasi silang untuk memperkirakan kinerja statistik operator pembelajaran (biasanya pada set data yang tak terlihat). Operator ini juga digunakan untuk memperkirakan seberapa akurat suatu model yang akan tampil dalam praktek. Operator X-Validasi merupakan operator bersarang yang memiliki dua subproses: training subprocess (subproses percobaan) testing subprocess (subproses dan pengujian). Subproses percobaan digunakan untuk melatih sebuah model. Model yang terlatih kemudian diterapkan dalam subproses pengujian.

Biasanya proses belajar mengoptimalkan parameter model untuk membuat model sesuai dengan data percobaan. Jika kita kemudian mengambil sampel independen dari data pengujian, umumnya model tersebut tidak cocok dengan data percobaan maupun data pengujian. Hal ini disebut dengan istilah 'over-pas', dan sangat mungkin terjadi ketika ukuran set data training kecil, atau ketika jumlah parameter dalam model besar. Sehingga validasi silang merupakan cara untuk memprediksi kesesuaian model untuk satu set pengujian hipotesis ketika set pengujian eksplisit tidak tersedia.

Untuk menemukan operator X-Validation, pilih Evaluation pada View Operator, lalu pilih Validation, lalu pilih X-Validation .Setelah menemukan operator X-Validation, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.17 Hirarki Operator X-Validation



Gambar 4.18 Operator Validation

Operator X-Validation memiliki port input yaitu, training example set (tra) sebagai port input memperkirakan ExampleSet untuk melatih sebuah model (training data set). ExampleSet yang sama akan digunakan selama subproses pengujian untuk menguji model.

Selain itu, operator ini juga memiliki port output sebagai berikut:

- model (mod), Pelatihan subprocess harus mengembalikan sebuah model yang dilatih pada input ExampleSet. Harap dicatat bahwa model yang dibangun ExampleSet disampaikan melalui port ini.
- training example set (tra), The ExampleSet yang sebagai masukan pada port input diberikan pelatihan dilewatkan tanpa mengubah ke output melalui port ini. Port ini biasa digunakan untuk menggunakan kembali ExampleSet sama di operator lebih lanjut atau untuk melihat ExampleSet dalam Workspace Result.
- averagable (ave), subproses pengujian harus mengembalikan Vector Kinerja. Hal ini biasanya dengan menerapkan model dihasilkan dan mengukur kinerjanya. Dua port tersebut diberikan tetapi hanya dapat digunakan jika diperlukan. Harap dicatat bahwa kinerja statistik dihitung dengan skema estimasi hanya perkiraan (bukan perhitungan yang tepat) dari kinerja yang akan dicapai dengan model yang dibangun pada set data vang disampaikan secara lengkap.

🖉 📴 Parameters 🕅	Context 🐹		
2 5 5 7 5	t 📭 🕶		
% Validation	n (X-Validation)		
🖌 average performar	nces only		
leave one out			
number of validations	10		
sampling type	stratified sampling 💌		
use local random s	seed		
55			
A A A A A A A A A A	^		
Compatiblity level	5.3.000		

Gambar 4.19 Parameter X-Validation

Operator X-Validation juga memiliki parameter yang perlu diatur, diantaranya:

- average performances only (boolean), ini merupakan parameter ahli yang menunjukkan jika vector kinerja harus dirata-ratakan atau semua jenis dari hasil rata-rata.
- leave one out (boolean) Seperti namanya, leave one out validasi silang melibatkan penggunaan satu contoh dari ExampleSet asli sebagai data pengujian (dalam pengujian subproses), dan contoh-contoh yang tersisa sebagai data pelatihan (dalam pelatihan subprocess). Namun hal ini biasanya sangat mahal untuk ExampleSets besar dari sudut

pandang komputasi karena proses pelatihan diulang sejumlah besar kali (jumlah waktu contoh). Jika diatur dengan benar, parameter number of validations dapat diabaikan.

- number of validations (integer), parameter ini menentukan jumlah subset ExampleSet yang harus dibagi (setiap subset memiliki jumlah yang sama dari contoh). Juga jumlah yang sama dari iterasi yang akan berlangsung. Setiap iterasi melibatkan pelatihan model dan pengujian model. Jika ini ditetapkan sama dengan jumlah contoh dalam ExampleSet, Hal ini akan setara dengan operator X-Validasi dengan parameter leave one out set true.
- sampling type (selection), Operator X-Validasi dapat menggunakan beberapa jenis sampling untuk membangun subset. Sampel yang tersedia, diantaranya:
 - linear_sampling, Linear sampling hanya membagi ExampleSet ke partisi tanpa mengubah urutan contoh yaitu subset dengan contoh-contoh berturut-turut diciptakan.
 - shuffled_sampling, Shuffled Sampling membangun subset acak ExampleSet. Contoh dipilih secara acak untuk membuat subset.
 - 3. stratified_sampling, Stratified Sampling membangun subset acak dan memastikan

bahwa distribusi kelas dalam himpunan adalah sama seperti dalam ExampleSet seluruh.

- use local random seed (boolean), Parameter ini menunjukkan jika local random seed harus digunakan untuk mengacak contoh subset. Dengan menggunakan nilai yang sama dengan local random seed maka akan menghasilkan subset yang sama. Mengubah nilai parameter ini mengubah cara contoh menjadi acak, sehingga subset akan memiliki satu set yang berbeda dari contoh. Parameter ini hanya tersedia jika Shuffled atau Stratified sampling dipilih. Hal ini tidak tersedia untuk pengambilan sampel Linear karena tidak membutuhkan pengacakan, contoh yang dipilih secara berurutan
 - Iocal random seed (integer), Parameter ini hanya tersedia jika parameteruse local random seed dipilih. parameter ini menentukan local random seed

Seperti yang telah disebutkan sebelumnya bahwa dalam membuat pohon keputusan pada contoh ini, kita menggunakan operator Apply Model. Operator ini menerapkan suatu model terlatih pada sebuah ExampleSet. Sebuah model pertama kali dilatih di sebuah ExampleSet, informasi yang berkaitan dengan ExampleSet dipelajari oleh model. Maka model tersebut dapat diterapkan pada ExampleSet yang lain dan biasanya untuk prediksi. Semua parameter yang diperlukan disimpan dalam objek model. Ini adalah wajib bahwa kedua ExampleSets harus persis nomor yang sama, order, jenis dan peran atribut. Jika sifat meta data dari ExampleSets tidak konsisten, hal itu dapat menyebabkan kesalahan serius.

Untuk menemukan operator Apply Model, pilih Modeling pada View Operator, lalu pilih Model Application, lalu pilih Confidence dan pilih Apply Model .Setelah menemukan operator Apply Model, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.20 Hirarki Operator Apply

Operator ini memiliki port input yaitu, **model** (**mod**) port ini mengharapkan model. Port ini harus memastikan bahwa nomor, order, jenis dan peran atribut dari ExampleSet pada model yang dilatih konsisten dengan ExampleSet pada port input data unlabeled. **unlabelled data (unl)** port ini mengharapkan suatu ExampleSet. Ini harus memastikan bahwa nomor, order, jenis dan peran atribut ExampleSet ini konsisten dengan ExampleSet pada model yang dikirim ke port input model dilatih.

Operator ini juga memiliki port output, diantaranya, **labeled Data (lab)**, Model yang diberikan dalam input diterapkan pada ExampleSet yang diberikan dan ExampleSet terbaru disampaikan dari port ini. Beberapa informasi akan ditambahkan ke input ExampleSet sebelum dikirimkan melalui port output. Dan **model (mod)**, Model yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui port ini.



Gambar 4.21 Operator Apply Model

Seperti yang terlihat pada gambar 4.22, Operator Apply Model hanya memiliki dua parameter yaitu, application parameters (menu) parameter ini merupakan parameter ahli yang berguna memodelkan parameter untuk aplikasi (biasanya tidak diperlukan). Dan create view (boolean) Jika model diterapkan pada port input mendukung Views, Hal ini mungkin untuk bukannya mengubah data yang membuat View mendasarinya. Transformasi yang akan biasanva dilakukan langsung di data kemudian akan dihitung setiap kali nilai diminta dan hasilnya dikembalikan mengubah Beberapa tanpa data. model tidak mendukung Views.



Gambar 4.22 Parameter Apply Model

Dalam membuat pohon keputusan untuk menentukan apakah seseorang berpotensi sakit menggunakan Hipertensi, kita juga operator Performance. Operator ini digunakan untuk evaluasi kinerja. Operator ini memberikan daftar nilai kriteria kinerja. Kriteria kinerja secara otomatis ditentukan agar sesuai dengan jenis tugas belajar. Berbeda dengan operator lain, operator ini dapat digunakan untuk semua jenis tugas belajar. Secara otomatis menentukan jenis tugas belajar dan menghitung kriteria yang paling umum untuk jenis tersebut.

Untuk menemukan operator Performance, pilih Evaluation pada View Operator, lalu pilih Performance and Measurement, lalu pilih Performance. Setelah menemukan operator Performance, seret (drag) operator tersebut lalu letakkan (drop) ke dalam view Process.



Gambar 4.23 Hirarki Operator Performance

Operator Performance memiliki port input vaitu, labelled data (lab), Port ini mengharapkan mengharapkan ExampleSet berlabel. Apply Model merupakan contoh yang baik dari operator yang menyediakan berlabel. data Pastikan bahwa ExampleSet memiliki atribut label dan atribut prediksi. performance (per) Ini adalah parameter opsional yang membutuhkan Performance Vector.

Selain itu, Operator ini juga memiliki port output yaitu, **performance (per)**, port ini memberikan Peformance Vector (kita menyebutnya outputperformance-vektor untuk saat ini). Peformance Vector adalah daftar nilai kinerja kriteria. **example set (exa)**, ExampleSet yang diberikan sebagai masukan dilewatkan tanpa mengubah ke output melalui port ini.



Gambar 4.24 Operator Performance

Operator ini hanya memiliki satu parameter yaitu, **use example weights** (boolean) Parameter ini memungkinkan contoh bobot contoh yang akan digunakan untuk perhitungan kinerja jika memungkinkan. Parameter ini memiliki tidak memiliki efek jika atribut tidak memiliki peran bobot.



Gambar 4.25 Parameter Performance

Selanjutnya, susun dan hubungkan port-port dari operator decision tree, operator Apply Model dan operator Performance seperti yang terlihat pada Gambar 55.



Gambar 4.26 Susunan Operator Decision Tree, Apply Model, Performance

Kemudian hubungkan operator retrieve (tabel SakitHipertensi) dengan operator validation dengan menarik garis pada port input dan output yang terdapat pada operator tersebut, seperti yang tampak pada Gambar 56.



Gambar 4.27 Susunan Operator Retrieve dengan Operator Validation

Setelah parameter dari masing-masing operator diatur, dan posisi operator disusun dengan benar, klik Run, lalu tunggu beberapa detik hingga RapidMiner akan menampilkan hasil Keputusan decision tree berupa graph pohon. seperti yang tampak pada Gambar 4.28.



Gambar 4.28 Tampilan Decision Tree

Chapter 5

Neural Network

Apa itu Neural Network?

Dapat dikatakan bahwa neural netwok dapat mempelajari pemetaan input data ke output data. Neural network merupakan model komputasi yang terinspirasi oleh prinsip-prinsip mengenai bagaimana cara otak manusia bekerja. Mereka dapat mempelajarinya dari data, mereka mampu mengeneralisasi dengan baik, dan mereka tahan dengan kebisingan.

Biasanya jaringan saraf digunakan untuk masalah-masalah seperti klasifikasi (classification), prediksi (prediction), pengenalan pola (pattern recognition), pendekatan (approximation), dan asosiasi (association). Mereka hanya perlu belajar dari beberapa data sampel, dan setelah mereka telah mempelajarinya, mereka dapat bekerja dengan input data yang tidak diketahui, atau bahkan input data yang bising maupun tidak lengkap.

Secara umum Neural Network (NN) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia. NN ini merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut.

Secara sederhana NN adalah sebuah alat pemodelan data statistik non-linear. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pada data. Secara mendasar. pola-pola sistem pembelajaran merupakan proses penambahan pengetahuan pada NN yang sifatnya kontinuitas sehingga pada saat digunakan pengetahuan tersebut maksimal akan dieksploitasikan dalam secara mengenali suatu objek. Neuron adalah bagian dasar dari pemrosesan suatu Neural Network. Dibawah ini merupakan bentuk dasar dari suatu neuron.

Bentuk Neural Network

Setiap neural network terdiri dari unit pengolahan dasar yang saling berhubungan, yang disebut Neuron. Network belajar dengan memodifikasi bobot hubungan antara neuron selama proses pelatihan. Bentuk dasar arsitektur suatu Neural Network adalah sebagai berikut:



Gambar 5.1 Arsitektur Dasar Neural Network

Secara umum, terdapat tiga jenis Neural Network yang sering digunakanberdasarkan jenis network-nya, yaitu:

- 1. Single-Layer Neural Network
- 2. Multilayer Perceptron Neural Network
- 3. Recurrent Neural Networks

Single-Layer Neural Network

Neural Network jenis ini memiliki koneksi pada inputnya secara langsung ke jaringan output.



Gambar 5.2 Single-layer Neural Network

Jenis Neural Network ini sangatlah terbatas, hanya digunakan pada kasus-kasus yang sederhana.

Multilayer Perceptron Neural Network

Jenis Neural Network ini memiliki layer yang dinamakan "hidden", ditengah layer input dan output. Hidden ini bersifat variable, dapat digunakan lebih dari satu hidden layer.


Gambar 5.3 Multilayer Perceptron Neural Network

Gambar di atas menunjukkan sebuah jaringan saraf sederhana yang dibuat dengan easyNeurons. Jenis jaringan ini disebut Multi Layer Perception dan itu merupakan salah satu jaringan yang paling umum digunakan.

Recurrent Neural Network

Neural network jenis ini memiliki ciri, yaitu adanya koneksi umpan balik dari output ke input.



Gambar 5.4 Recurrent Network

Kelemahan dari jenis ini adalah Time Delay akibat proses umpan balik dari output ke titik input.

Proses Pembelajaran pada Neural Network

Proses pembelajaran merupakan suatu metoda untuk proses pengenalan suatu objek yang sifatnya kontinuitas yang selalu direspon secara berbeda dari setiap proses pembelajaran tersebut. Tujuan dari pembelajaran ini sebenarnya untuk memperkecil tingkat suatu error dalam pengenalan suatu objek.

Secara mendasar, neural network memiliki sistem pembelajaran yang terdiri atas beberapa jenis berikut:

- 1. Supervised Learning
- 2. Unsupervised Learning

Supervised Learning

Sistem pembelajaran pada metoda Supervised learning adalah system pembelajaran yang mana, setiap pengetahuan yang akan diberikan kepada sistem, pada awalnya diberikan suatu acuan untuk memetakan suatu masukan menjadi suatu keluaran yang diinginkan. Proses pembelajaran ini akan terus dilakukan selama kondisi error atau kondisi yang diinginkan belum tercapai. Adapun setiap perolehan error akan dikalkulasikan untuk setiap pemrosesan hingga data atau nilai yang diinginkan telah tercapai.

Unsupervised Learning

Sistem pembelajaran pada neural network, yang mana sistem ini memberikan sepenuhnya pada hasil komputasi dari setiap pemrosesan, sehingga pada sistem ini tidak membutuhkan adanya acuan awal agar perolehan nilai dapat dicapai. Meskipun secara mendasar, proses ini tetap mengkalkulasikan setiap langkah pada setiap kesalahannya dengan mengkalkulasikan setiap nilai weight yang didapat.

Siapa yang menggunakan Neural Network?

Beberapa aplikasi yang khas adalah gambar (image), sidik jari dan pengenalan wajah (fingerprint face recognition), prediksi and saham (stock prediction), prediksi untuk taruhan (sport bets prediction), klasifikasi pola dan pengakuan (pattern classification and recognition), pengawasan dan (monitoring pengendalian and control). Mereka digunakan dalam industri, kedokteran (diagnosa), aplikasi militer (seperti radar pada pengenalan citra),

keuangan dan robotika. Akhir-akhir ini mereka sangat popular di industri game karena berkat mekanisme belajar yang dilakukan, mereka dapat memberikan kontrol adaptif dan pembelajaran untuk karakter yang dikendalikan computer.

Kegunaan Neural Networks

- 1. Pengenalan karakter optikal (Optical character recognition)
- 2. Pengenalan citra (Image recognition)
- 3. Pengenalan sidik jari (Fingerprint recognition)
- 4. Prediksi saham (Stock prediction)
- 5. Prediksi taruhan (Sport bets prediction)
- 6. Kontrol computer untuk karakter game (Computer controlled game characters)
- 7. Model statistical (Statistical modeling)
- 8. Data mining

Neural Network pada RapidMiner

Kita mulai dengan menggunakan data sederhana dalam tabel GE.xls. Data tersebut juga bisa kita dapatkan dengan melakukan pengunduhan melalui salah satu add-ins Microsoft Excel yang bernama DownloaderXL, dimana data mengenai harga saham yang terjadi dalam rentang waktu tertentu telah dicatat pada sebuah web hosting.

Contoh Kasus:

Perkiraan harga saham dengan menggunakan metoda Neural Network.

M	licrosoft Exc	el - GE						
	File Edit	View Inser	t Format	Tools Dat	a Window	Monte Carlo	xl bit Help Adobe PDF	
2111	200	alar	A ABC 451			n nu 14		0004
: 14			£ ✓ ⊪2	6 40 4	🛛 • 🔺 🕛	1 + (* +)		.00% • (
A	rial	- 10	- B 1	<u>u</u> ∣≣⊧	콜 콜 🔜	\$ % ,	號 🔐 🏥 🗰 🖽 • 🔇	- <u>A</u> -
: 00	aluzerYI 🔻 🗍 Br	dk∩untes¥L ▼	Downloade		ntecYl • D	edictorYI + F	Racktesting 1 - Classifier VI -	PortfolioYI
: 01	O10	ancorossia a	EF 47544	00440404	iocoste i D	concorne 11	gachicostingne i <u>C</u> iassino ne i	Cordione
-	610	• p	< 55.17541	00140191	F	-	0	
-	A	8	6	U	E	F	6	н
1	Date	Open	High	Low	Close	Volume	StochasticOscillator(8,5)	
2	05/08/2009	14.18	14.35	13.95	14.33	129813025	ን 90.00	
3	05/11/2009	14.18	14.22	13.90	13.99	94827338	87.69	
4	05/12/2009	14.04	14.11	13.08	13.49	119778085	83.76	
5	05/13/2009	13.03	13.07	12.57	12.73	108138237	69.63	(
6	05/14/2009	12.59	13.00	12.33	12.86	91507642	57.62	
7	05/15/2009	12.89	13.07	12.52	12.68	84562005	39.26	
8	05/18/2009	12.93	13.31	12.90	13.28	77949206	32.04	
9	05/19/2009	13.45	13.77	13.30	13.51	93399522	32.21	
10	05/20/2009	13.84	14.18	13.55	13.58	115136364	42.82	
11	05/21/2009	13.32	13.34	12.86	13.05	106967634	45,49	
12	05/22/2009	13.18	13.22	12.84	12.92	54782537	48.77	
13	05/26/2009	12.84	13.42	12.67	13.20	97279770	48.79	
14	05/27/2009	13.28	13.31	12.68	12.81	108987013	40.92	
15	05/28/2009	13.05	13.10	12.64	13.01	83621254	32.51	
16	05/29/2009	13.14	13.34	12.05	13.29	86101902	37.25	
17	06/01/2009	13.63	13.80	13.32	13.67	87599560	47.17	
18	06/02/2009	13.51	13.66	13.36	13.61	67039102	55.18	
19	06/03/2009	13.46	13.51	13.11	13.31	69318595	65.10	

Gambar 5.5 Tabel GE.xls dalam Microsoft Excel

Buatlah *file* baru pada Microsoft Excel berdasarkan tabel harga saham. Berikan nama Header: Date, Open,

High, Low, Close, Volume, Stochastic Oscilator. Isilah sel seperti gambar [berapa]. Simpan dengan nama GE.xls

Lakukan pemilihan *repository* GE_TEMP_XLSDATA dengan melakukan *drag and drop* yang ditempatkan pada *panel main process* seperti gambar 5.6.



Gambar 5.6 Import Repository

Lakukan pemilihan operator *Neural Network* seperti gambar 5.7. Kemudian *drag and drop ke Main Process* seperti sebelumnya



Gambar 5.7 Operator Neural network

Lakukan pembuatan hubungan antara *repository* dan *operator*, kemudian antara *operator* dengan hasil *output*.



Gambar 5.8 Menghubungkan Seluruh Operator ke Result

klik ikon Play . Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 5.9 Ouput Neural Network

Gambar 5.9 merupakan grafik berbentuk *node* yang saling terhubung seperti layaknya sebuah jaringan syaraf dari hasil rules yang telah kita dapatkan

Chapter 6

Market Basket Analysis

Memahami Market Basket Analysis

Retail atau Eceran salah satu cara pemasaran produk meliputi semua aktivitas yang melibatkan penjualan barang secara langsung ke konsumen akhir, konsumen akhir membeli kumpulan produk dengan jumlah yang berbeda di waktu yang berbeda. Namun penjualan secara ritel hari ini bukanlah apa-apa jika insdustrinya tidak mampu berkompetisi dengan baik. Lanskap yang kompleks dan cepat berubah, persaingan yang ketat, dan pelanggan yang semakin menuntut mendorong *retailer* harus memikirkan kembali bagaimana mereka beroperasi. Kemampuan untuk memahami pola pikir konsumen adalah hal yang sangat penting bagi *retailer*.

Teknologi telah membantu *retailer* dengan memungkinkan untuk menyimpan data konsumen dengan volume yang sangat besar dan biaya yang sangat wajar. *Retailer* kini dapat memiliki miliyaran informasi tentang informasi pelanggan mereka. Informasi ini dapat menjawab pertanyaan-pertanyaan penting termasuk: Kapan pelanggan akan membeli? Bagaimana pembayaran dilakukan? Berapa banyak dan apa item tertentu yang dibeli? Apa hubungan antara barang yang dibeli?

Tidak ada keraguan bahwa data point-of-sales (POS) ini yang (ketika digunakan secara efektif) diberdayakan pengecer untuk lebih memahami bisnis mereka dan meningkatkan pengambilan keputusan. Pengecer proaktif menggunakan informasi ini untuk memberikan penawaran yang ditargetkan yang sesuai dengan harapan konsumen dan kemudian memberikan dampak penghasilan positif. Namun pada dasarnya, bagaimanakan *retailer* menggunakan miliyaran informasi ini? Jawabannya adalah menghubungkan produk-produk yang ada.

Sering kali, sebagai konsumen, kita cenderung mengabaikan bagaimana barang secara fisik diatur dalam sebuah toko *retail* atau supermarket. Apa yang mungkin terlihat (bagi kita) hanyalah seperti sebuah 'distribusi acak', namun sebenarnya hal tersebut merupakan pengaturan barang yang direncanakan secara cermat. Pada intinya, toko *retail* menilai pola pembelian pelanggan dan mengatur produk-produk yang akan dibeli secara sesuai. Sehingga menyebabkan pelanggan melakukan kegiatan pembelian beberapa produk sekaligus tanpa disadarinya.

Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan inilah yang dikenal sebagai *Market Basket Analysis* (MBA). Seperti namanya, *Market Basket Analysis* pada dasarnya melibatkan penggunaan data transaksional konsumen untuk mempelajari pola pembelian dan menjelajahi kemungkinan (probabilitas dan) *crossselling*. Tujuan dari MBA adalah untuk memanfaatkan data penjualan efektif untuk meningkatkan taktik pemasaran dan penjualan di tingkat toko.

Contoh yang paling umum dari Market Basket Analysis adalah "Beer dan Diapers". Contoh ini merupakan kasus dari salah satu toko retail besar yang ada di US, Wal-Mart. Seorang manajer toko menemukan hubungan yang kuat antara salah satu merek popok bayi (diapers) dan salah satu merek beer pembeli. pada beberapa Analisa pembelian mengungkapkan bahwa kegiatan pembelian dilakukan oleh laki-laki dewasa pada hari jumat malam terutama sekitar jam enam dan tujuh sore. Setelah beberapa observasi, supermarket mengetahui bahwa:

- Karena bungkus dari popok bayi sangat besar, para istri, dimana dalam banyak kasus adalah seorang ibu rumah tangga, akan menyuruh suaminya untuk membelinya.
- Pada akhir dari minggu, para suami dan ayah akan menghabiskan minggunya dengan membeli beberapa beer.

Jadi, apa yang akan dilakukan supermarket dari pengetahuan ini?

- Mereka menempatkan *premium beer* tepat disebelah *diapers*
- Hasilnya adalah para ayah akan membeli diapers dan yang biasanya membeli beer biasa sekarang

membeli *premium beer* seperti yang sudah diperkirakan.

 Secara signifikan, para pria yang biasanya tidak membeli bir sebelum mulai berbelanja akan membelinya karena itu begitu mudah dilihat dan diambil - hanya sebelah popok (cross-sell)

Istilah Market Basket Analysis sendiri datang dari kejadian yang sudah sangat umum terjadi di dalam swalayan, vakni ketika para konsumen pasar memasukkan semua barang yang merak beli ke dalam keranjang (basket) yang umumnya telah disediakan oleh pihak swalayan itu sendiri. Informasi mengenai produk-produk yang biasanya dibeli secara bersamasama oleh para konsumen dapat memberikan "wawasan" tersendiri bagi para pengelola toko atau swalayan untuk menaikkan laba bisnisnya (Albion Research, 2007).

Metodologi Association Rules

Metodologi Association Rules, atau Analisis Asosiasi adalah sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar. Salah satu penerapan Metode Association rules adalah pada Market Basket Analysis. Association rule adalah sebuah ekspresi implikasi dari bentuk $X \rightarrow Y$, dimana X dan Y adalah itemset yang saling terpisah (disjoint), dengan kata lain $X \cap Y = \emptyset$. Dalam menentukan Association Rule, terdapat suatu interestingness measure (ukuran ketertarikan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu. Ada dua ukuran yaitu:

1. Support: Bagian transaksi yang mengandung kedua X dan Y.

 $Support(A) = rac{Jumlah transaksi mengandung A}{Total Transaksi}$

Atau jika terdapat dua buah item dalam X, nilai support diperoleh dari rumus berikut:

 $Support(A \cap B) = \frac{Jumlah \ transaksi \ mengandung \ A \ dan \ B}{Total \ Transaksi}$

2. Confidence: Seberapa sering item dalam Y muncul di transaksi yang mengandung X.

 $Cofidence = P(B|A) = \frac{Jumlah \ transaksi \ mengandung \ A \ dan \ B}{Jumlah \ transaksi \ mengandung \ A}$

Kudua ukuran ini nantinya berguna dalam menentukan interesting association rules, yaitu untuk dibandungkan dengan batasan (threshold) yang ditentukan oleh user. Batasan tersebut umumnya bernama *minimum support* dan *minimum confidence*.

Mengapa menggunakan Support dan Confidence? Support adalah ukuran yang penting karena jika aturan memiliki support yang kecil, maka kejadian bisa saja hanyalah sebuah kebetulan. Aturan Support yang rendah juga cenderung tidak menarik dari mungkin perspektif bisnis karena tidak akan memberikan keuntungan saat mempromosikan barangbarang yang jarang dibeli pelanggan bersamaan. Untuk alasan ini. dukungan sering digunakan untuk menghilangkan ketidak-menarikan ini. Confidence, adalah ukuran kehandalan dari kesimpulan yang dibuat oleh aturan. Semakin besar Confidence, semakin besar kemungkinan untuk Y hadir dalam transaksi yang mengandung X. Confidence juga memberikan probabilitas bersyarat dari Y yang diberikan ke X.

Contoh Association Rules

Untuk lebih memahami Association Rules, mari kita telusuri contoh berikut. Sebuah toko retail telah melakukan transaksi dengan pembeli seperti yang tertulis pada tabel.

Kode Transaksi	Produk yang terjual
001	Pena, Roti, Mentega
002	Roti, Mentega, Telur
003	Buncis, Telur, Susu

Tabel 6.1 Tabel Transaksi

004	Roti, Mentega
005	Roti, Mentega, Kecap, Telur, Susu

Tahap pertama adalah mencari nilai dari Support sesuai dengan rumus yang telah disebutkan sebelumnya. Misalnya, Untuk transaksi yang memuat {roti, mentega} ada 4, maka nilai supportnya adalah 80%. Lalu jumlah transaksi yang memuat {Roti, Mentega, Susu} ada 2, maka nilai supportnya adalah 40%. Sedangkan transaksi yang memuat {buncis} hanya 1, maka nilai supportnya adalah 20%. Jika kita tentukan bahwa *minimum support*nya adalah 30%, maka rule yang memenuhi adalah sebagai berikut:

Kombinasi Produk	Nilai Support
{roti}	80%
{mentega}	80%
{telur}	60%
{susu}	60%
{roti, mentega}	80%
{mentega, telur, susu}	40%
{roti, mentega, telur, susu}	40%

Tabel 6.2 Kombinasi Produk dan Nilai Support

Setelah semua pola kombinasi dan nilai dari Supportnya ditemukan, barulah dicari Association Rules yang memenuhi syarat minimum untuk confidence. Bila ditentukan syarat minimum untuk confidence sebesar 50% maka Association Rules yang dapat dipakai adalah:

Association Rules	Support	Confidence
{roti} → {mentega}	80%	100%
{roti} → {telur}	40%	50%
{roti} → {susu}	40%	50%
{roti} → {mentega, telur}	40%	50%
{roti} → {mentega, susu}	40%	50%
{roti} → {telur, susu}	40%	50%
{roti} → {mentega, telur, susu}	40%	50%
{mentega, telur} \rightarrow {roti}	40%	100%
{roti, mentega, susu} \rightarrow {telur}	40%	100%
{roti, telur, susu} → {mentega}	40%	100%
{mentega, telur, susu} \rightarrow {roti}	40%	100%

Tabel 6.3 Association Rules dan Nilai Confidence

Assosiation Rule akan dipilih sesuai kebijakan manajer toko, semakin tinggi support dan confidence semakin baik hasilnya. Misalkan kita ambil contoh yaitu {mentega, telur} \rightarrow {roti} yang memiliki nilai Support 80% dan Confidence 100%, artinya adalah: "Seorang konsumen yang membeli mentega dan telur memiliki kemungkinan 100% untuk juga membeli roti. Aturan ini

cukup signifikan karena mewakili 40% dari catatan selama ini."

Frequent Itemset Generation dan Rule Generation

Frequent Itemset Generation

Tujuannya adalah untuk menemukan semua itemset yang memenuhi *minimum support*. Item set ini sering disebut dengan frequent. Namun Masalah utama pencarian Frequent Itemset adalah banyaknya jumlah kombinasi itemset yang harus diperiksa apakah memenuhi minimum support atau tidak. Salah satu cara untuk mengatasinya adalah dengan mengurangi jumlah kandidat itemset yang harus diperiksa.

Apriori adalah salah satu pendekatan yang sering digunakan pada Frequent Itemset Mining. Prinsip Apriori adalah jika sebuah itemset infrequent, maka itemset yang infrequent tidak perlu lagi diexplore supersetnya sehingga jumlah kandidat yang harus diperiksa menjadi berkurang. Kira kira ilustrasinya seperti ini:



Gambar 6.1 Frequent Item Set tanpa Apriori

Pada gambar 36, pencarian Frequent Itemset dilakukan tanpa menggunakan prinsip Apriori. Dengan menggunakan prinsip Apriori, pencarian Frequent Itemset akan menjadi seperti di bawah ini:



Gambar 6.2 Frequent Item Set dengan Apriori

Dapat dilihat bahwa dengan menggunakan Apriori, jumlah kandidat yang harus diperiksa cukup banyak berkurang.

Rule Generation

Tujuannya adalah untuk mengekstrak semua aturan yang memiliki high-confidence dari itemsets yang ditemukan dari langkah sebelumnya. Aturan ini disebut Strong Rules.

Market Basket Analysis pada RapidMiner

Sekali lagi, pencarian Rule pada Association Rules merupakan sebuah proses yang luar biasa panjang. Manusia tidak akan mampu untuk melakukan pengitungan dengan berates-ratus (belum data kombinasi dari seluruh item yang ada). Maka dari itu, untuk mencari seluruh Rules yang ada, RapidMiner menyediakan tools telah untuk mempermudah pengguna. Untuk memahami cara menggunakan tools ini, ikuti manual berikut secara seksama.

Contoh Kasus :

Transaksi Penjualan Sederhana.

Kita mulai dengan menggunakan data sederhana yang kita miliki yang terdapat pada sub bab pengenalan Market Basket Analysis, Tabel 5.1.

	Α	В	С	D	E	F	G	н	
1	TID	PENA	ROTI	MENTEGA	TELUR	BUNCIS	SUSU	KECAP	
2	001	1	1	1	0	0	0	0	
3	002	0	1	1	1	0	0	0	
4	003	0	0	0	1	1	1	0	
5	004	0	1	1	0	0	0	0	
6	005	0	1	1	1	0	1	1	
7									
									1

Gambar 6.3 Tabel Penjualan Sederhana

Buatlah Table baru pada Microsoft Excel berdasarkan tabel 5.1. Berikan nama Header: TID (Transaction ID), PENA, ROTI, MENTEGA, TELUR, BUNCIS, SUSU, KECAP. Isilah cell seperti gambar 5.3. Simpan dengan nama TransaksiMakanan.xls.

Lakukan *Importing Data* kedalam Repositori, seperti yang sudah dijelaskan pada **Bab 2**. Browse table Microsoft Excel yang telah dibuat, dan masukan kedalam Local Repository, seperti gambar disamping.



Gambar 6.4 Repositori

Lakukan Drag dan Drop Tabel TransaksiMakanan tadi kedalam Process. Sehingga Operator Database muncul dalam Main Proses seperti gambar 5.5.



Gambar 6.5 Database dalam Main Process

Untuk melakukan Market Basket Analysis, kita membutuhkan setidaknya tiga buah operator, antara lain Association Rule, FP-Growth, dan Numerical to Binomial.



Gambar 6.6 Operator Create Association Rules

Assocuation rules dilakukan dengan menganalisis data pada frequent if/then patterns

menggunakan kriteria support dan confidence untuk mengidentifikasikan suatu relasi antar item. *Frequent if/then pattern* digali menggunakan operator FP-Growth. Operator Create Association Rules menggunakan frequent itemsets ini dan menghasilkan association rules.



Gambar 6.7 Operator FP-Growth

Frequent itemsets merupakan kelompok item yang sering muncul bersama-sama dalam data. Operator *FP-Growth* mengkalkulasikan semua frequent itemset dari input yang diberikan menggunakan struktur data FP-tree. Adalah wajib bahwa semua atribut dari masukan merupakan bilangan binominal (true/false).



Gambar 6.8 Operator Numerical to Binominal

Operator *Numerical to Binominal* diperlukan untuk mengubah nilai atribut yang berada pada table TransaksiMakanan menjadi binominal.

Selanjutnya lakukan Pencarian Filter untuk memudahkan kita menemukan operator yang dibutuhkan, lakukan seperti pada gambar berikut.



Gambar 6.9 Pencarian Operator Numerical to Binominal

Untuk Mencari Operator Numerical to Binominal, lakukan pencarian seperti gambar disamping. Operator ini terdapat pada hirarki: Data Transformation \rightarrow Type Conversion



Gambar 6.10 Pencarian Association Rules

Susunlah ketiga operator tersebut menjadi seperti gambar 5.11.



Gambar 6.11 Menghubungan Database TransaksiMakanan pada Operator Numerical to Binomial

Hubungkan Tabel TransaksiMakanan yang kita miliki dengan operator Numerical to Binominal. Proses ini akan membuat nilai dari Tabel Transaksi makan mejadi *Binominal Attributes*.

🛛 🛃 Parameters	🛛 🕘 Context 🔀								
🚨 💀 💀 🦻	🕵 🖶 -								
🍕 Nume	rical to Binominal								
attribute filter type	all 🔻								
invert selection									
include special	include special attributes								
min	0.0								
max	0.0								

Gambar 6.12 Parameter Numerical to Binomial

Data yang kita miliki merupakan data sederhana. Kita hanya memperhitungkan 1 buah penjualan produk pada setiap transaksinya. Maka nilai yang terbaik untuk menjadi *false* adalah ketika tidak ada produk tertentu yang terjual dalam suatu transaksi, jadi kita sini nilai *min* dan *max* menjadi 0, Sehingga yang bernilai *false* adalah ketika sebuah produk tidak terdapat pada sebuah transaksi.

Hubungkan operator *Numerical to Binominal* dengan operator FP-Growth pada example output.



Gambar 6.13 Menghubungkan Operator Numerical to Binomial dengan Operator FP-Growth

Terdapat dua buah output untuk *Numerical to Binominal*, yaitu example dan original.

- Example, *numeric attributes* dikonversikan menjadi *binominal attributes* melalui output ini.
- Original, *numeric attributes* dilewatkan tanpa konversi. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

Lewatkan output pada *example*.

Isilah Parameter FP-Growth seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum support* senilai 30% atau 0.3.

🛛 🛃 Parameters 🕱	🔵 Context 🔀
🏅 🕫 🖻 🦻 🕵	- 🛼 🛨
🎘 FP-	Growth
🖌 find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.3
max items	-1
must contain	

Gambar 6.14 Parameter FP-Growth

Kemudian hubungkan operator *FP-Growth* dengan operator *Association Rules*.



Gambar 6.15 Menghubungkan Operator FP-Growth dengan Operator Create Association Rules

Terdapat dua buah output pada operator FP-Growth, yakni *example* dan *frequent*.

• *Example*, input yang diberikan dilewatkan tanpa adanya perubahan. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

• *Frequent,* frequent itemset dikirimkan melalui output ini.

Lewatkan output pada frequent.

Kemudian isilah Parameter Association Rules seperti gambar berikut. Sesuai dengan contoh pada sub bab seselumnya, isilah *minimum confidence* senilai 50% atau 0.5.



Gambar 6.16 Parameter Association Rules

Setelah itu hubungkan Association Rules pada result. Sehingga seluruhnya membentuk seperti gambar 5.17. Ialu klik ikon Play . Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan.



Gambar 6.17 Susunan Operator Association Rules

Setelah beberapa detik, akan muncul sebuah tab Association Rules yang baru, yang isinya adalah sebuah table berisi seluruh itemset yang memenuhi parameter FP-Growth dan Association Rules. Totalnya terdapat 152 rules yang ditemukan.

No.	Premises	Conclusion	Support	Confid	LaPI	Gain	p-s	Lift	Convi
131	TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
132	ROTI, SUSU	MENTEGA, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
133	MENTEGA, SUSU	ROTI, TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
134	ROTI, MENTEGA, SUSU	TELUR, KECAP	0.200	1	1	-0.200	0.160	5	00
135	ROTI, TELUR, SUSU	MENTEGA, KECAP	0.200	1	1	-0.200	0.160	5	00
136	MENTEGA, TELUR, SUSU	ROTI, KECAP	0.200	1	1	-0.200	0.160	5	00
137	ROTI, MENTEGA, TELUR, SUSU	KECAP	0.200	1	1	-0.200	0.160	5	00
138	KECAP	ROTI, MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
139	ROTI, KECAP	MENTEGA, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	∞
140	MENTEGA, KECAP	ROTI, TELUR, SUSU	0.200	1	1	-0.200	0.160	5	00
141	ROTI, MENTEGA, KECAP	TELUR, SUSU	0.200	1	1	-0.200	0.120	2.500	∞
142	TELUR, KECAP	ROTI, MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
143	ROTI, TELUR, KECAP	MENTEGA, SUSU	0.200	1	1	-0.200	0.160	5	00
144	MENTEGA, TELUR, KECAP	ROTI, SUSU	0.200	1	1	-0.200	0.160	5	00
145	ROTI, MENTEGA, TELUR, KECAP	SUSU	0.200	1	1	-0.200	0.120	2.500	00
146	SUSU, KECAP	ROTI, MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
147	ROTI, SUSU, KECAP	MENTEGA, TELUR	0.200	1	1	-0.200	0.120	2.500	00
148	MENTEGA, SUSU, KECAP	ROTI, TELUR	0.200	1	1	-0.200	0.120	2.500	00
149	ROTI, MENTEGA, SUSU, KECAP	TELUR	0.200	1	1	-0.200	0.080	1.667	00
150	TELUR, SUSU, KECAP	ROTI, MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
151	ROTI, TELUR, SUSU, KECAP	MENTEGA	0.200	1	1	-0.200	0.040	1.250	00
152	MENTEGA, TELUR, SUSU, KECAP	ROTI	0.200	1	1	-0.200	0.040	1.250	00

Gambar 6.18 Hasil Association Rules Pertama

Tentunya ini akan menyulitkan kita untuk mengambil kesimpulan karena jumlah rules yang terlalu banyak. Maka dari itu yang harus kita lakukan adalah mengubah nilai *minimum support* dan *minimum confidence*.

Klik ikon Edit 🔟 untuk kembali pada *model view.* Lalu klik Operator FP-Growth.



Gambar 6.19 Operator FP-Growth

Kemudian lihat bagian parameter. Ubah nilai minimum support menjadi 95%, seperti yang sudah dijelaskan pada sub bab Association Rules, semakin tinggi nilai support maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum supportnya terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

🦳 🛃 Parameters 🕱	🔵 Context 🔀
🚨 🗟 🖻 🦻 🕵	: 📭 👻
🎘 FP-	Growth
🖌 find min number of	itemsets
min number of items	100
max number of retries	15
positive value	
min support	0.95
max items	-1
must contain	

Gambar 6.20 Mengubah Parameter FP-Growth

Sekarang kita beralih pada Operator Create Association Rules.



Gambar 6.21 Operator Create Association Rules

Ubah nilai minimum confidence menjadi 95% atau 0.95, semakin tinggi nilai confidence maka semakin dapat dipercaya rules yang dihasilkan. Namun perhitungkan juga hasilnya nanti. Terkadang jika nilai minimum confidence terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi.

criterion	confidence 🔻
min confidence	0.95
gain theta	2.0
laplace k	1.0

Gambar 6.22 Mengubah Parameter Association Rules

Klik ikon Play 🖻 untuk menampilkan hasil yang baru.

No	Premises	Conclusion	Support	Confid	LaPla	Gain	D-S	Lift	Convi
1	ROTI	MENTEGA	0.800	1	1	-0.800	0.160	1.250	00
2	MENTEGA	ROTI	0.800	1	1	-0.800	0.160	1.250	00
3	SUSU	TELUR	0.400	1	1	-0.400	0.160	1.667	00
4	ROTI, TELUR	MENTEGA	0.400	1	1	-0.400	0.080	1.250	00
5	MENTEGA, TELUR	ROTI	0.400	1	1	-0.400	0.080	1.250	00

Gambar 6.23 Hasil Association Rules Kedua

Maka sekarang yang dihasilkan menjadi lima buah rules. Kita bisa mengambil salah satu dari rules ini untuk dijadikan sebuah pegangan dalam strategi penjualan retail. Tentunya yang memiliki nilai support dan confidence yang tinggi.

Untuk melihat dalam bentuk grafik. kita dapat memilih opsi Graph View. O Table View O Graph View O Annotations



Gambar 6.24 Hasil dalam bentuk Graph View

Glossarium

- *Algoritma* Kumpulan perintah untuk menyelesaikan suatu masalah.
- Apriori Algoritma untuk frequent itemset mining dan association rule dalam database transaksional. DIhasilkan dengan menidentifikasi setiap buah item, dan memperluasnya menjadi kombinasi kumpulan item yang lebih besar asalkan himpunan item muncul cukup sering dalam database.
- Association Rules Sebuah metodologi untuk mencari relasi (asosiasi) istimewa/menarik yang tersembunyi dalam himpunan data (atau data set) yang besar.
- **Binominal Attributes** Atribut dengan tipe Binominal (true dan false).
- Confidence (Market Basket Analysis) Seberapa sering item dalam Y muncul di transaksi yang mengandung X.
- Decision tree Struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan

simpul daun merepresentasikan kelas atau distribusi kelas.

- Disjoint Himpunan terpisah, tidak ada elemen yang berhubungan diantara kedua himpunan yang bersangkutan
- *Flowchart* Sebuah diagram dengan simbol-simbol grafis yang menyatakan aliran algoritma.
- Frequent Itemset Itemset yang memunyai support >= minimum support yang diberikan oleh user dalam Market Basket Analysis.
- Market Basket Analysis Teknik untuk menemukan hubungan dari produk-produk yang dibeli secara bersamaan.
- MBA Lihat Market Basket Analysis.
- Minimum SupportNilaiSupportTerkecildalamMarketBasket Analysis yang dapat di toleransi.
- Minimum Confidence Nilai Confidence terkecil dalam Market Basket Analysis yang dapat di toleransi.
- Neural Network Jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan jaringan syaraf manusia.
- **Numeric Attributes** Atribut dengan tipe Numerical (1-9).
- *Operator* suatu tanda atau simbol yang dipakai untuk menyatakan suatu operasi atau manipulasi nilai.
- ParameterNilai yang mengikuti acuan keterangan
atau informasi yang dapat menjelaskan
batas-batas tertentu dari suatu suatu sistem persamaan.

Pruning Teknik dalam machine learning yang mengurangi ukuran pohon keputusan dengan menghapus bagian dari pohon yang memberikan sedikit kekuatan untuk mengklasifikasikan kasus.

- RapidMinerSebuah tool yang digunakan untuk
melakukan analisis terhadap data mining,
text mining dan analisis prediksi.
- Repositori Kumpulan paket yang siap untuk diambil dan digunakan sesuai dengan kebutuhan pengguna.
- Simpul akar Simpul tanpa ayah yang berada pada tingkat tertinggi.
- Simpul daun Semua simpul yang berada pada tingkat terendah.
- *Simpul internal* Semua simpul dari pohon yang memiliki anak tetapi bukan daun.
- Support(Market Basket Analysis) Bagian transaksi
yang mengandung kedua X dan Y.
- Teori graf Cabang kajian yang mempelajari sifat-sifat graf.
- Validasi Tindakan yang membuktikan bahwa suatu proses/metode dapat memberikan hasil yang konsisten sesuai dengan spesifikasi yang telah ditetapkan.

Daftar Pustaka

Akhtar, Fareed dan Caroline Hahne. 2012. *RapidMiner 5 Operator Reference*, [online], (www.rapid-i.com, diakses tanggal 30 Januari 2013).

Amiruddin, dkk. Penerapan Association Rule Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan Untuk Menemukan Pola Sertifikasi Guru. Institut Teknologi Surabaya. Surabaya.

Basuki, Achmad dan Iwan Syarif. *Decision Tree*, [online], (http://lecturer.eepis-its.edu/~entin/Data%20Mining/ Minggu%205%20Decision%20Tree.pdf, diakses tanggal 05 Februari 2013).

Khusnawi. 2007. *Pengantar Solusi Data Mining*. Yogyakarta.

Kusumadewi, Sri. 2003. Artificial Intelligence: Teknik dan Aplikasinya.

Mitchel, Tom M. 1997. *Machine Learning*. New York: McGraw-Hill.

Prasetyo, Bowo. 2011. *Mengenal RapidMiner*, [online], (www.slideshare.net/bowoprasetyo/RapidMiner, diakses tanggal 31 Januari 2013).

Prasetyo, Kokoh Philips. 2006. *APriori*, [online] (http://philips.wordpress.com/2006/06/07/apriori, diakses tanggal 03 Februari 2013)

-----. 2006. Association Rule Mining, [online]. (http://philips.wordpress.com/2006/05/10/associationrule-mining, diakses tanggal 03 Februari 2013).

Rafaida, Ropi. *Decision Tree (Pohon Keputusan)*, [online],

(http://file.upi.edu/Direktori/FPEB/PRODI._MANAJEME N_FPEB/197302052005012-

ROFI_ROFAIDA/MATERI_KULIAH/DECISION_TREE.pdf, diakses tanggal 05 februari 2013).

Ross, Peter. 2000. Data Mining [online]. (http:// www.soc.napier.ac.uk/~peter/vldb/dm/dm.html, diakses tanggal 07 Februari 2013)

Wahono, Romi satria. *Data Mining:Proses Data Mining*, [online], (http://romisatriawahono.net/lecture/dm/ romi-dm-02-proses-june2012.pptx, diakses tanggal 31 Januari 2013).

2012. *RapidMiner 5.0 Manual English*, (online), (www.rapid-i.com, diakses tanggal 30 Januari 2013).

3 tips for Setting up Association Rules using RapidMiner, [online]. (http://www.simafore.com/blog /bid/110113/3-tips-for-setting-up-a-Market-Basket-Analysis-using-RapidMiner, diakses tanggal 08 Maret 2013).

Association Analysis: Basic Concepts and Algorithms, [online]. (http://www-users.cs.umn.edu/ ~kumar/dmbook/ch6.pdf, diakses tanggal 08 April 2013)

Decision Tree (Pohon Keputusan), [online], (http://www.google.co.id/url?sa=f&rct=j&url=http://no vrina.staff.gunadarma.ac.id/Downloads/files/21783/Alg oritma%2BC4.pdf&q=algoritma+c4&ei=6h9gUcbJFIqrrA

fT7IGQAw&usg=AFQjCNG7HbyNPOqa63ZoPexX76TrIIJ7g, diakses tanggal 05 februari 2013).

Landasan Teori Market Basket Analysis, [online]. (http://library.binus.ac.id/eColls/eThesis/Bab2/2010-1-00498-MTIF%20Bab%202.pdf, diakses tanggal 08 April 2013)

Understanding the Concept of Market Basket Analysis, [online]. (http://www.thesmartcube.com/ insights/blog/brand-strategy/understanding-theconcept-of-market-basket-analysis, diakses tanggal 08 Maret 2013)

RapidMiner Resources. (http://RapidMinerresources .com/uploads/videos/tomott/RapidMiner5-Vid1.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%201.flv, diakses tanggal 02 Februari 2013)

-----. (http://RapidMinerresources .com/uploads/videos/neural%20networks%202.flv, diakses tanggal 02 Februari 2013) rapid miner merupakan *software tool open source* untuk *data Mining*. Rapid miner dioperasikan pada sebuah lingkungan untuk *machine learning, data mining, text mining* dan *predictive analytics*. Sebelum masuk ke tahap penggolahan data, yang harus disiapkan adalah

- Data yang akan di olah
- Tools rapid miner

Untuk pengolahan data dengan algoritma apriori langkah pertama yang harus dilakukan adalah penetuan atribut, atribut yang di gunakan atau yang akan di pilih harus ada hubungan dan kaitan satu sama lain. Pada tutorial kali ini data yang akan digunakan adalah data penjualan pada sebuah toko furniture dan elektornik dengan 346 record dan 16 atribut tapi yang akan digunakan hanya 3 atribut saja.

Setelah penentuan atribut dilakukan, langkah selanjutnya yaitu tahap preprocessing data, pada tahap ini akan dilakukan beberapa hal, yaitu cleaning data dan transformas data.

Pada tahap cleaning dilakukan pembersihan data transaksi penjualan furniture dan elektronik. Tahap *cleaning* ini dilakukan dengan cara membuang data yang kosong atau data yang tidak sempurna kemudian membuang field-field yang tidak dibutuhkan agar tidak memperlambat dalam proses asosiasi dana mempercepat mendapatkan hasil pola gabungan.

Tahap selanjutnya adalah transformasi data, tahap ini akan dilakukan agar mempermudah saat memasukkan data ke dalam tools rapid miner dengan membuat inisial atau symbol symbol untuk data. Seperti contoh berikut:

Tabel Inisial item pembelian data transaksi penjualan

No	Items Pembelian
1	MC1 (mesin cuci 1 Tempat)
2	MC2 (mesin cuci 2 Tempat)
3	SP (speaker aktif)
4	RP (rak piring)
5	PFK (palung fadhil kaca)
6	LED24 (TV uk.24 dengan semua merk)
7	LED 32 (TV uk.32 dengan semua merk)
8	LED 43 (TV uk.43 dengan semua merk)
9	PR (parabola)
10	DG (digital)
11	LP2 (lemari pakaian 2 pintu)
12	LP3 (lemari pakaian 3 pintu)

13	LH (lemari hias)
14	LM (lemari mini)
15	MB (meja belajar)
16	KR (kursi)

Setelah dilakukan transformasi data, tahap selanjutnya adalah melakukan tranformasi data penjualan kedalam bentuk tabular. Berikut merupakan hasil transformasi data ke dalam bentuk tabular dapat dilihat pada Tabel:

cam	MC1	MC2	RP	PFK	LED24	LED32		JM
C1	0	0	0	0	0	0		0
C2	0	0	1	0	0	0		0
C3	1	0	0	0	0	0		0
C4	0	1	0	0	0	0		0
C5	0	0	0	0	0	0		0
C6	0	0	0	0	0	0		0
C7	0	0	0	0	0	0		0
C346	0	0	0	0	1	0	••••	0

Tabel Tabular atribut *item* pembelian data transaksi penjualan

Keterangan :

- a) Customers merupakan nomor faktur setiap pembeli
- b) MC1, MC2, RP, PFK, LED24 Dll merupakan barang-barang atau produk yang dijual oleh PT.Citra Mustika Pandawa cabang Kerinci.
- c) 0 merupakan tanda bahwa barang tersebut tiak dibeli oleh pembeli
- d) 1 merupakan tanda bahwa barang tersebut dibeli oleh pembeli

Setelah melalui tahap *preprocessing* dan *transformasi* data, selanjutnya adalah tahap asosiasi dengan menggunakan algoritma apriori untuk menentukan pola pembelian pelanggan. Hasil ini diukur dengan menggunakan nilai *Support* dan *Confidence*. Percobaan perhitungan ini menggunakan *tools* Rapid Miner dengan 346 *record* data. Dalam tahap imlementasi dengan *rapidminer* ini ada 3 operator yang digunakan dalam *tools*, yaitu: *Read Exel, Numeric to Binominal* dan *W-Apriori*. Operator pertama

adalah *Read Exel*, operator ini berfungsi untuk tempat data dan akan diinputkan data yang sudah diolah. Karena data yang diolah berupa *exel* maka operator yang digunakan adalah *read Exel*. Operator kedua adalah *Numeric to Binominal*, operator ini berfungsi untuk merubah data yang tadinya *numeric* menjadi nominal dua nilai pada operator read Exel karena semua atribut dari masukan wajib merupakan bilangan binominal yaitu nilai true/false. Operator ketiga adalah *W-Apriori*, operator ini berfungsi untuk perhitungan algoritma apriori. Desain dari ketiga operator ini dapat dilihat pada gambar berikut:



Setelah ketiaga operator terhubung dan dijalankan maka akan keluar hasil untuk item yang sering muncul dan saling berhubungan, pada data penjualan elektronik dan furniture ini hasil yang didapat nilai minimal *support* 4% dan minimal *confidance* 90%. Hasil yang diperoleh dari rapidminer ada 2 *rule* yaitu:

- 1. Jika membeli parabola maka akan membeli digital
- 2. Jika membeli LED32 dan parabola maka akan memebeli digital



Demikian tutorial pengolahan data dengan algoritma apriori dengan tools rapidminer yang sangat sederhana ini, mohon maaf apabila banyak terdapat kesalahan, terima kasih.

REFERENSI

[1].

https://www.researchgate.net/publication/336422159 Penerapan Algoritma Apriori Terhadap Data P enjualan Di Swalayan Koperasi Bappenas Jakarta Pusat

[2]. https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/1496f

Open lembar kerja baru shobat dengan cara double click NEW PROSES pada tools yang telah disediakan oleh RapidMiner. Seperti tampilan berikut ini :

Contra prices Republican 5.3.0	C & Darks PC 100		Marchine Marchine Marchine
In the poper has been	line .		
19889	······································		
Develop 0.1.0 0 0 0	2 Prosent - 2 He - 2 H - H - 2 B Prosent -	a - 0.1	
8 Schepenten konsen fü 9 Schepetrich 9 Schepetric	-		 ngen ()
■ Augusteres = 2 = 0 ■ d = 0 0 0 = = 52 = 52 = 7			Enter opergenerates Topores Process

Ok shobat, langkah pertama sudah shobat lalui sekarang kita kelangkah berikutnya, terlebih dahulu import data yang ingin shobat kelompokkan, disini saya menggunakan data tugas akhir mahasiswa untuk dikelompookkan berdasarkan bidang kompetensi. Oke lanjut, silahkan lakukan seperti tampilan dibawah ini : dan untuk pengetahuan shobat, kita melakukan import data ada beberapa tipe atau format file yang digunakan seperti CSV, Excell Sheet, XML disini shobat sesuaikan dengan format file data shobat karena data saya menggunakan format Excell makadari itu saya pilih import data menggunakan format Excell Sheet.

Composition - Applifica 11.0	Hit ben At The American State		Sacily and a
$\begin{array}{c c} \label{eq:constraints} \left\{ \begin{array}{c} \mbox{tr} \end{tabular} \\ \mbox{tr} \end{tabular} \end{tabular} \\ \mbox{tr} \end{tabular} \end{tabular} \end{tabular} \end{tabular} \end{tabular} \end{tabular} \end{tabular} tabular$	HT	Ø + 0 ₽ 8 Ø Ø	Binnel Breasen Breasen Breasen Synthetic Synth
	Armanna () (a =) 3 more some tage	Pres Longitor	The rod operator which is the odde most operator
0			0

Ok, kalau sudah seperti pada tampilan diatas maka silahkan shobat arahkan pointer shobat kearah tempat dimana data uji coba shobat simpan.

🤹 men prozen i Hapibline 133	U 9 Date K					(a)#1.8
the fat from lots per	Color report mineral - Days	2 - 1				1
# landes 1 1 1 1 1	7 Bathas	the year's import year state. select the file that should be imported	ŝ			O Long 1
E Proces Caller (21)	a Doumants					
E Republic Acres (5)	Butmate	Y delana	dist.	Test .	Lastmoter	2 Paces
 Topol (2) 		in the formulation		file folder Borned (Mos Ercer	No.28.275 No. 49:28.270	
Ausselmen						
	Data of color optimized and					
	Exertipeutineet Lite, at	el .			*	1
		127			Mires X Dear	Ne not operator which The outer work operator
0			_			

Kemudian silahkan shobat tekan tombol next hingga sampai pada step terakhir seperti tampilan berikut ini, kemudian jangan lupa shobat beri nama pada data cluster shobat seperti yang saya lakukan disini saya beri nama "DATA_UJI_COBA" kemudian tekan tombol finish.

(in the formal land part	Convertient Sector	1
Construction Operation IP Image: Control of Contro	Security approximation of the security of	
Approximate (1) (1) (1) A = (2) (2) (2) (2) Suppose (1) (1) Suppose (1) Suppose (1) Suppose (1) Suppo	gara (dat_at_)/mit	nine spirporties Ing Conset 1
	Luden Austhepathepath, JA, OBA	The root operator which The root operator which The rooter wood operator

Jika shobat sudah menekan tombol finis maka dari itu akan otomatis akan mengarahkan kita pada data yang sudah kita import tadi seperti tampilan berikut ini jika shobat menemukan tampilan seperti berikut ini itu artinya shobat berhasil melakukan import data pada RapidMiner dan jika belum silahkan dicoba lagi.

	-	Mire 1330	B Date PC	-						or such and
Dia Sal	(Crosses)	(entite (plane)							_	
		19 0	20	b 11	8 5	12 m				
			and the later	-		-				(Rentes VV + B.)
	- Carro			Advanced like	-					
A fee on Constitution Constitutions Constitution										
Danger3	e ret ears	No. 7 special s	detiums, To	quie efficie	16			transfoller (100-1001) all		a la ca
Avette		anglater.	decain.	and the second	aphines.	1001.00	solution. M.			A 19 DOUT TO A STATE OF
	100,001	2014	4.017	1054	0.891					8 -8 -8 - 1
	100,004	0014	18.750	1.742	0.871	10.000	0.079			O D DISCHARGE COMM
-	100,007	1014	1.010	0.758	0.001	10.177	0.788			0474, 8.8, 0084 mm
	1002104	2014	107	187	0.791	10.773	0.0071			
	1062071	2014	0.007	1875	0.733	6.798	4/98			
	1002012	2014	1	440	8.778	12.00.0	0.007			
	1062043	2014	4454	4400	0.794	16.756	4744			
	1002014	2014	4447	1/102	1.801	4.796	4.007			
	1003914	2014	4400	0.708	0.754	4.878	1007			
	100317	2014	8.750	4/78	1111	4.727	0.7%			
11	1002118	2014	4.602	1.008	0.684	4.790	4/98			
0	1002048	2014	1403	178	0.122	1.000	1479			
0	1003027	2014	8.7%0	4.77%	0.647	16790	4798			
54	1003029	2014	1447	1.00	1.605	10.054	0.0021			
es	1062100	2014	10.0117	4/102	0.032	10.000	8.000			
55	1062936	2014	4.405	4.007	0.122	4.197	1007			
67	100300	2014	4402	1.702	1419	4.175	4798			Q.
10.64										Manager and a line line
20	8									
									0	
							0.11 mil			And States
Dec 18, 2		IN COMPANY	passing a	e						

Oke shobat, proses load data sudah selesai sekarang kita lanjut ke proses pengelompokan atau clustering di RapidMiner:

Ikuti langkah berikut ini untuk hasil yang lebih optimal

- a. Langkah pertama yaitu kita memili retrieve sebagai tempat penyimpanan data cluster
- b. Langkah kedua pilih K-Means untuk proses clustering
- c. Langkah ketiga pilih **cluster distance performance** untuk tampilan output cluster Berikut tampilannya beserta koneksi

Sebelum melakukan koneksi lakukan terlebih dahulu load data pada database shobat yg telah shobat simpan sebelumnya. Atau dapat juga shobat tarik database shobat atau bahasa kasarnya seret kedalam lembar kerja shobat seperti tampilan berikut ini:



Kemudian silahkan shobat include point2 yang telah saya sebutkan diatas tadi. Seperti pada tampilan berikut ini:

	Constanting of the second seco	kaelt	-
	14889 A4 4889 999		
	6 ment 0 ≥ 0 m + 0 - 2 # from 1	della da Branna I	
Constant and the second s		Kunandi Tasar. Maratelari Maratelari A Trades and another	
	Constant over Constan	Pres Loader () to per this series () to per	-

Dismping kanan atas ada input jumlah cluster (K) = 5, itu bisa shobat tentukan sendiri. Dan pastikan tidak ada error pada saat pengkoneksian antara database dengan model. Jika shobat telah sukses sama seperti gambar diatas makadari itu silahkan shobat tekan perinta RUN atau tekan tombol F11 pada keyboard shobat.

	10		Patrian	•			I torute	d Jan at Rep (Restored Del)	NAMES OF A CONTRACTOR OF A CON	Breates a local
2.046	Intel Contra	Date from ()	Pid Net C	Advanced Cha	n Ower	store				Card - + Sterent
transfe	647,788 alum	yes, l'ineter	10040.01	spin drives					'des/18e/18011800 at	+ # B.08
day 1		shafe	- 100	-	ange	Arritan.	1000.00	1000.00		a manufacture of the second seco
	18	(hole), J	1942/041	10.001	11874	0.001	1	1		All a property local
ł.		shader_2	1002004	4.794	1.702	18893	100	1.879		Bella, una dolla como
2		inster,2	-962467	180	11708	10.001	1071	16.786		· · · · · · · · · · · · · · · · · · ·
4	4	-thefter J	100306	4447	1.875	4.766	875	1401		
6		distant, 2	1002911	10007	1.875	4.787	1.765	14.750		
4		state, J	100210		1410	4798	1.014	1.80		
F		dealer J	100310	1403	1475	4.784	0.708	1754		
		dates (100204	1847	1.760	1.001	1.76	1.812		
		them, 2	100211	140	1.74	10.130	10.0	187		
1		Outer, J	100817	1786	1.128	1100	8.727	11794		
-		1400.7	-96/918	1411	1.008	0.004	1.752	4 PS		
u	1	note J	100,048	1463	1.798	11.122	100	1479		
		Added 1	-94/827	4.794	1.44	4.647	8.758	4.754		
10.		nate 2	-	1000	1.00	1.000	1004	1821		
H		states, 2	-962418	140	1.76	410	1.01	1.000		
10		new J	100,000	1000	1.01	4.00	1.12	1.00		
-		100								The basic strates of 12 or 1
a										
	-									

Tampilan diatas merupakan cluster masing-masing anggota dan dibawah ini merupakan hasil grafik pengelompokan. Silahkan shobat perlajari dan kembangkan sendiri.



Tugas 06

Nama	: Juminovario
NIM	: 202420018
Kelas	: MTI 23 Reguler A
MK	: Advanced Database

Pada tutorial ini akan membahas bagaimana cara penggolahan data dengan algoritma apriori dengan tools rapid miner, sebelum masuk ke pembahasan inti kita akan membahas terlebih dahulu apa itu algoritma apriori dan apa itu tools rapid miner.

Apa itu algoritma apriori? Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis. Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan suatu kombinasi item. Tujuan dari algoritma apriopri yaitu untuk menentukan frequent itemsets yang dijalankan pada beberapa atau sekumpulan data serta mencari pola yang sering muncul pada setiap transaksi. Sedeangkan rapid miner merupakan software tool open source untuk data Mining. Rapid miner dioperasikan pada sebuah lingkungan untuk machine learning, data mining, text mining dan predictive analytics.

Sebelum masuk ke tahap penggolahan data, yang harus disiapkan adalah

• Data yang akan di olah

• Tools rapid miner

Untuk pengolahan data dengan algoritma apriori langkah pertama yang harus dilakukan adalah penetuan atribut, atribut yang di gunakan atau yang akan di pilih harus ada hubungan dan kaitan satu sama lain. Pada tutorial kali ini data yang akan digunakan adalah data penjualan pada sebuah toko furniture dan elektornik dengan 346 record dan 16 atribut tapi yang akan digunakan hanya 3 atribut saja.

Setelah penentuan atribut dilakukan, langkah selanjutnya yaitu tahap preprocessing data, pada tahap ini akan dilakukan beberapa hal, yaitu cleaning data dan transformas data.

Pada tahap cleaning dilakukan pembersihan data transaksi penjualan furniture dan elektronik. Tahap cleaning ini dilakukan dengan cara membuang data yang kosong atau data yang tidak sempurna kemudian membuang field-field yang tidak dibutuhkan agar tidak memperlambat dalam proses asosiasi dana mempercepat mendapatkan hasil pola gabungan.

Tahap selanjutnya adalah transformasi data, tahap ini akan dilakukan agar mempermudah saat memasukkan data ke dalam tools rapid miner dengan membuat inisial atau symbol symbol untuk data.

Seperti contoh berikut:

Tabel Inisial item pembelian data transaksi penjualan

No.	Items Pembelian
1	MC1 (mesin cuci 1 Tempat)
2	MC2 (mesin cuci 2 Tempat)
3	SP (speaker aktif)
4	RP (rak piring)
5	PFK (palung fadhil kaca)
6	LED24 (TV uk.24 dengan semua merk)
7	LED 32 (TV uk.32 dengan semua merk)
8	LED 43 (TV uk.43 dengan semua merk)

9	PR (parabola)
10	DG (digital)
11	LP2 (lemari pakaian 2 pintu)
12	LP3 (lemari pakaian 3 pintu)
13	LH (lemari hias)
14	LM (lemari mini)
15	MB (meja belajar)
16	KR (kursi)

Setelah dilakukan transformasi data, tahap selanjutnya adalah melakukan tranformasi data penjualan kedalam bentuk tabular. Berikut merupakan hasil transformasi data ke dalam bentuk tabular dapat dilihat pada Tabel:

cam	MC1	MC2	RP	PFK	LED24	LED32		JM
C1	0	0	0	0	0	0		0
C2	0	0	0	0	0	0	•••••	0
C3	1	0	0	0	0	0	•••••	0
C4	0	1	0	0	0	0	•••••	0
C5	0	0	0	0	0	0	•••••	0
C6	0	0	0	0	0	0	•••••	0
C7	0	0	0	0	0	0	•••••	0
•••••			••••		••••		••••	••••
C346	0	0	0	0	1	0	•••••	0

Tabel Tabular atribut item pembelian data transaksi penjualan

Keterangan

a) Customers merupakan nomor faktur setiap pembeli

- b) MC1, MC2, RP, PFK, LED24 Dll merupakan barang-barang atau produk yang dijual oleh PT.Citra Mustika Pandawa cabang Kerinci.
- c) 0 merupakan tanda bahwa barang tersebut tiak dibeli oleh pembeli
- d) 1 merupakan tanda bahwa barang tersebut dibeli oleh pembeli

Setelah melalui tahap preprocessing dan transformasi data, selanjutnya adalah tahap asosiasi dengan menggunakan algoritma apriori untuk menentukan pola pembelian pelanggan. Hasil ini diukur dengan menggunakan nilai Support dan Confidence. Percobaan perhitungan ini menggunakan tools Rapid Miner dengan 346 record data. Dalam tahap imlementasi dengan rapidminer ini ada 3 operator yang digunakan dalam tools, yaitu: Read Exel, Numeric to Binominal dan W-Apriori. Operator pertama adalah Read Exel, operator ini berfungsi untuk tempat data dan akan diinputkan data yang sudah diolah. Karena data yang diolah berupa exel maka operator yang digunakan adalah read Exel. Operator kedua adalah Numeric to Binominal, operator ini berfungsi untuk merubah data yang tadinya numeric menjadi nominal dua nilai pada operator read Exel karena semua atribut dari masukan wajib merupakan bilangan binominal yaitu nilai true/false. Operator ketiga adalah W-Apriori, operator ini berfungsi untuk perhitungan algoritma apriori. Desain dari ketiga operator ini dapat dilihat pada gambar berikut:



Setelah ketiga operator terhubung dan dijalankan maka akan keluar hasil untuk item yang sering muncul dan saling berhubungan, pada data penjualan elektronik dan furniture ini hasil yang didapat nilai minimal support 4% dan minimal confidance 90%. Hasil yang diperoleh dari rapidminer ada 2 rule yaitu:

- 1. Jika membeli parabola maka akan membeli digital
- 2. Jika membeli LED32 dan parabola maka akan memebeli digital

```
Best rules found:
1. PR=true 19 ==> DG =true 19 conf:(1)
2. LED32=true PR=true 14 ==> DG =true 14 conf:(1)
```

Demikian tutorial pengolahan data dengan algoritma apriori dengan tools rapidminer yang sangat sederhana ini, mohon maaf apabila banyak terdapat kesalahan, terima kasih.

REFERENSI

- <u>https://www.researchgate.net/publication/336422159_Penerapan_Algoritma_Apriori_Ter</u> hadap Data Penjualan Di Swalayan Koperasi Bappenas Jakarta Pusat
- 2. https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/1496fdi

Nama : Mustakim

NIM : 202420028

Tugas : 6

Cari beberapa tutorial yang membahas pengolahan data menggunakan metode prediksi pada tool rapidminer. Silahkan buat ringkasan tutorialnya kembali dengan menggunakan dataset yang kamu buat sendiri. Tuliskan dalam format ms word dan sertakan semua sumber rujukan tutorial yang anda gunakan.

Jawaban:

Sumber: https://www.youtube.com/watch?v=snNeTNMkSPY

1. Buka Program/Aplikasi rapidminer



2. Setelah itu klik new



3. Memasukkan data yang akan kita analisis dengan mengetik read dan pilih type file yang akan kita masukkan (Jenis file data)

n Tutorial Klasifikasi Da	ata Mining dengan Rani	idminer			- 0 ×
Poveniew 32.25 0 B	Process X FXML X				
	🔍 + 🗠 + 🕆 🖉 Process 🕨			Ø • \$* 💷 🗟 A •	To a a a a
	Main Process				Process
	~ D				logverbosity init *
Coperators 31: 50 co 8				1	logfie
@ - read 0 > 1					
Import (20) Import (20) Read (11) Read-Road Read All Read Solution Read					🚯 4 hidden expert parameters
8.8-0.0.0					R Comment X
8 Samples (norm) 8 (B DB					O Help X
8 49 NewLocalRepository (site serve)					· · · ·
	A Problems × BLog ×				Aread CSV
	hio problems found				Synopsis
	Mess	929	Fines	Location	This operator can read cav files.
				~ 🗈	備 (小) IND 10:09 PM 口 10/29/2020

Setelah itu tekan dan drag jenis datanya ke lembar kerja/main proses lalu klik import configurasi untuk memasukkan data

Tutorial Klasifikasi D	ata Mining dengan Panidminer	cess*> - RapidMiner@Darma		× 1
	ata Mining dengan Kapidininer			
Powerview 14 55 co. 10	Process X XXL X			Westman with a li
	🗇 🕶 👻 👚 Process 🕨		a • 🎝 🔢 🖉 🗍 •	To D D D D
	Main Process			A Read CEV
				in measicav
	and D			> Import Configuration
Coperators H 20 4 0	Dead CSV	- P	-	cav file
e read		-4		column sep.
E Data (18)				
Read Excel				(C) use docues
Read Excel with Format				quotes char
Read SAS				skip comments
Read All.				parse numbers
Read 300FF				
Read SPSS	0			decimal char
Repositories X 💥 🗆 🗓				A hidden expert parameters
8.4-0.0.0				Comment 16
8 🜍 Samples (nore) 8 🕼 OB				Help X
8 W NewLocaRepository (site iterne)				
	A Problems X DLog X			A Read CSV
	St One potential problem			Synopele
	Message	Fixes	Location	This operator can read cov
	The mandatory parameter "cev file" is undefined.	Set mandatory parameter "csv file".	# Read CSV	Sles.
	m			10:11 PM
			~ 🗅	10/29/2020

4. Setelah dklik import configurasu maka akan muncul eperti gambar dan tinggal kita pilih datanya..

Tutorial Klasifika	asi Data M	ining dengan Rapidmine	Data Import wizard - Step 1 of 4		×	- 0
Operators O	Look ja Daam Beokma	n the set of the file that should be imported to the set of the file that should be imported to the set of the file that should be imported to the set of	e6. Tie Name Sica 3 HB	v Tise File Folder CSV File		Parameters Read CSV Read CSV Fitmport Configuration courtin sep. use quotes quotes char. skip comments parse numbers decimal char. B hidden expert parameters Comment R
 IS DB IP NewLocalRepository and 	File Name	elimiter separated files			*	Read CSV Synopsis This operator can read cov Res.
💶 💿 📼 🌅	•		<u>v</u>	الانتفاع المحد	^ D /	えか)IND 10:13 PM 口 10/29/2020

Kemudia klik next maka akan muncul konfiguran tentang datanya, apakah sdh terpisah atau belum pd setiap data, kemudian klik next lg dan setelah selesai semua konfigurasi tingga kita klik finish

Tutorial Klasifikasi D	ata Mining dengan Rapidminer			
📑 🦦 📰 🖬 🕬 🔊	• • II II 🛐 🛱 🖲			
Coperators	Process K Dime K K		<i>4</i> 7 • \$7 □ ∞ <i>4</i> • (== (==	Parameters Parameters Read CSV Read CS
B Samples (norma) B B DB B MP NewLocalRepository (non-damma)				Comment X
	A Problems # a Log #			% Split Validation
	2 No problems found Message	Fixes	Location	Synopsis A SimpleValidation randomly splits up the example set into training and test set and

5. Kemudian ketik validation untuk memvalidasi datanya

Kemudian drag validasinya ke main proses dan hubungkan data nya, seperti pada gambar

Tutorial Klasifikasi Da	ta Mining dengan Rapidminer	- 0 - 1
Overview 100 00 00 Overview 100 00 Overview 100 00 Overview 100 00 Overview 100 Ove	Process N N NAL X Process N Process N N N N N N N N N N N N N N N N N N	Parameters
 IB OB IP NewLocaRepository (sets doma) 	Problems M. Tallan W.	Read CSV
	4 potential problems Kessage Fixes Location Kadatory input missing at port Validation training Fixes Location Kadatory input missing at port Validation training Fixed operator generating Exampledet. Validation training Fixed operator generating Exampledet. Validation training	Synopsis This operator can read cav files.
🔲 o 💷 👩 🚍		信 ① IND 10:18 PM 口 10/29/2020

6. Kemudian double klik validation maka akan muncul 2 box yaitu training dan testing. Pad box training akan kita masukkan algoritma yang kan kita gunakan

Tutorial Klasifikasi Da	ata Mining dengan Rapidminer	+> - RapidMiner@Darma		- 8 ×
C 🗃 🖬 🖬 🕬 🔊 🕫	🕨 🕨 🔲 🛐 🛒 🕕			
Overnow X 100 Operators X 100 Ope	Process X 2 XM, X 4 - 10 - 2 Process + % Validation + Torening 0	(mit mit) (mit mit) (mit mit) (mit)	0 - \$	Parameters 2 2 2 0 10 Validation (Split Validation) split relative * split ratio 0.7 sampling type shuffed*
				1 hidden expert parameter Compatbijty level Comment # Phop #
	Problems X Lop X One potential problem Messape Mandatory input missing all polt Validation averagable 1.	Fixes El Inset operator generating PerformanceV	Location % Validation averagable 1	Split Validation Synopsis A SimpleValidation randomly splits up the example set into braning and test set and
= o 🗆 👩 🚍		· • ·	^ ⊡ /	えか) IND 10:23 PM 口 10/29/2020

Pada testing kita tambahkan apply model dan performance

THE PARTY AND	ta Mining dengan Deni	dminor	n e-Larma		- 8
	ta Mining dengan Rapi	aminer			
Overview X 32 o E	Process X XML X	Unidation .			Parameters 31.32
m 10 0	Contraction in the	varianten v		**************************************	20000
	training max		Testas Reply Model		% Validation (Split Valid
		(not not)		formance are	split relative
Operators X 52 4 B		a male and	0	% 5 m	split ratio 0.7
Import (1)				_	sampling type shuffed
Results (1)					
Export (1)					
Evaluation (17)					
Performance Measurement (1 G G Classification and Regress)					
Performance (Binomin					
Padromanca (Rantata					🟦 1 hidden expert param
Repositories X H = B					Compatibility level
Samples (new)					Comment 3
DB NewLocalRepository (sets derma)				_	Piep R
	L				
	Problems X B Log X				To Split Validati
	Messa	01	Fines	Location	Synopsis
					splits up the example set
torial Klasifikasi Da	ta Mining dengan Rapi	dminer	v@Darma		- 0
🐚 🗔 🖾 🕬 🔊 🕫					
Overview X 35 4 0	Process X 2 XML X				
60 (C)	💠 🛥 🗄 🛥 🔮 🖉 Process 🕨			a • \$* 🗉 🖻 🕭 •	No Parameters 31 14
	Nun Process				Walitation (Solit Valid
					The amount of the same
	in D	The second se		6.00	spit relative
Operators - H. 50 -0. 83	Read CSV	An instant of the second second		4	split relative
Operators 31: 50 4: 60	Read CSV	tra mad			split relative
Operators 10.00 (1)	Read CSV	tra mat B re		C as	split relative split ratio 0.7 sampling type shuffed
Coperators # 50 41 60 perfor @ > 10 Persuts (1) @ Resuts (1) @ Resuts (2) @ Resut		4 m nut 36 m		C 195	spit relative spit ratio 0.7 sampling type shuffed
Cperators 00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Prend CSV	100 million 100 mi		4 m 0 m 0 m	spit relative spit ratio 0.7 sampling type shuffed
Ceerators C C C C C C C C C C C C C C C C C C C	Pred CSV	and the second s		4 m	split relative split ratio 0.7 sampling type shuffed
Coperators C C C C C C C C C C C C C C C C C C C	Pred CSV			 € m 4 m	split relative split ratio 0.7 sampling type shuffed
Coperators and the second seco	Pred CSV			€ 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	spilt relative spilt ratio 0.7 sampling type shuffed
Creators Construction of Results (1) Results (1) Result	Pred CSV			€ == € ==	spilt relative spilt ratio 0.7 sampling type shuffed
Creators	Pred CSV			4 m	spilt retailse spilt ratio 0.7 sampling type shuffed
Creators Construction of the second s				C ==	spilt retailor spilt ratio ampling type shuffed
Creators of the second				4 m	spilt retailor spilt ratio ampling type shuffed
Creators perfor perfor Performance Evaluation (1) Results (1) Results (1) Results (1) Results (1) Classification and Represent Performance Beasurement (1) Performance Beasurement (1) Performance Beasurement Performance Beasurement P	Problems X D Log X				spilt retailor spilt ratio sampling type shuffed

Synopsis

へ ID (深 中) IND 10:26 PM 10/29/2020

A SimpleValidation randoms splits up the example set int training and test set and

🔹 💿 💿 📷 🔤 kemudian kan tampil hasilnya, dimana kita dapat melihat akurasinya

Message

Tutorial Klas	sifikasi Data Mining d	engan Rapidminer	eve process - Rapid Mine a 1	larna	а.) -	Repositories X 22 4 10
accuracy	Multiclass Classification Performant Antibiotece Distributer	ice () Annotations			B 9 + 10	8 B DB
	accuracy: 86.67%					
	W	true yes	true no	class precision		
	pred yes pred no	20	6	95.24%		
	class recall	85.95%	85.71%		- 8	
Character Vice B	🕎 tugas 6.docx - Microsoft Word					Cartan Manter M. S. and D.
Nov 25, 2015 9 28 57 AM Nov 25, 2015 9 28 57 AM		wax.xmLws.WebServiceException	Failed to access the WSDL at:		9	
	💌 🚍 💌		~		9	へ ID: 🧖 (小) IND 10:27 PM 10/29/2020

Kemudian klo mau melihat pohon tree ID3 nya yang dihasilkan tinggal klik maka akan tampil

a Tutorial Klasifikasi Data Mi	seev process	* = RapidMiner@Darma		- 0 - ×
Tutorial Klasnikasi Data Mi	ning dengan Rapidininei			
Result Oveniew 😤 📆 PerformanceVector	(Performance) 😤 👔 ExampleSet (Read CSV) 😤 👰	Tree (ID 3) X		Repositories 11 50 ca 10
Graph View Text View Annotations			G 4 -	8.0-0.0.0
Zoom				8 B DB
99				MP NewLocaRepository (sets dama)
Mode	Legree			
	= high = loss	* medium		
	Implic	Capitor Capitor		
Tree · sign	- impression are put- solar bautan	- left - middle		
Node Labels	14 NO 945 NO 1	• Nove bet		
Edge Labels	× — — — —			
Save Image.	A A		i	
Help - topic	Jack House H	here here we here we h		
= ingrease politicat tool	an .			
no (pss	bas	Þ		
		~		
¥ 4	Y YES			
C BLog M 22 a B				Bystem Monitor X 32 - 68
Nov 25, 2015 9.28 57 AM INFO: Checking for updates.	and state income and any Mish Paralitative and the Paralitation	and the MATTE of	1	
All	e Service Tweek. It feiled with:	est de marcal	_	Res 1.2 Cb
🔲 🗖 💁 🗇 🕒				へ ID // ↓ ND 10:29 PM 10/29/2020 □