

Silahkan cari satu paper yang membahas tentang association rule dan buat ringkasan dan pembahasan tentang domain masalah yang dibahas dalam paper dan berikan pendapat anda (review paper). Format bebas.

Tulis dalam ms word dan kirimkan sebelum deadline

ASSOCIATION RULE ADVANCED DATABASE

IMPLEMENTASI ALGORITMA-ALGORITMA *ASSOCIATION RULES* SEBAGAI
BAGIAN DARI PENGEMBANGAN *DATA MINING ALGORITHMS COLLECTION*

RINGKASAN PEMBAHASAN :

Association Rules

Association rules digunakan untuk menemukan hubungan di antara data atau bagaimana suatu kelompok data mempengaruhi suatu keberadaan data yang lain. Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Sebuah *association rule* dengan *confidence* sama atau lebih besar dari *minimum confidence* γ dapat dikatakan sebagai *valid association rule*.

Algoritma-Algorithm Association Rules

1. Algoritma Apriori

Algoritma ini dicetuskan oleh Agrawal. Ide dasarnya adalah menghitung pola kemunculan *item* yang muncul dalam data transaksi dengan beberapa iterasi.

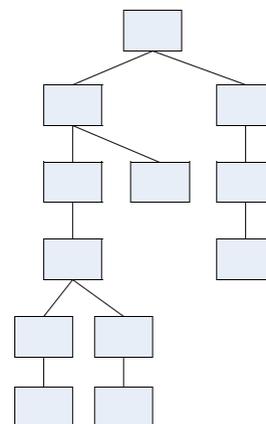
Selain algoritma Apriori yang diimplementasikan sendiri, dalam *data mining workbench* yang dibangun juga diintegrasikan implementasi algoritma Apriori oleh Christian Borgelt yang sering diacu dan digunakan oleh peneliti di bidang algoritma *data mining*.

2. Algoritma FP-Growth

Algoritma FP-Growth merepresentasikan transaksi dengan menggunakan struktur data *FP-Tree* Gambar 1. memperlihatkan proses penyusunan *FP-Tree* dari mulai representasi awal transaksi, pengurutan dengan hanya mempertahankan *frequent 1-itemset*, dan penyimpanannya di *FP-Tree*. Setelah *FP-Tree* terbentuk, langkah selanjutnya adalah memperoleh *frequent itemset* tanpa melakukan *candidates generation*.

TID	Itemset
01	f,a,c,d,g,i,m,p
02	a,b,c,f,l,m,o
03	b,f,h,j,o
04	b,c,k,s,p
05	a,f,c,e,l,p,m,n

TID	Itemset
01	f,c,a,m,p
02	f,c,a,b,m
03	f,b
04	c,b,p
05	f,c,a,m, p



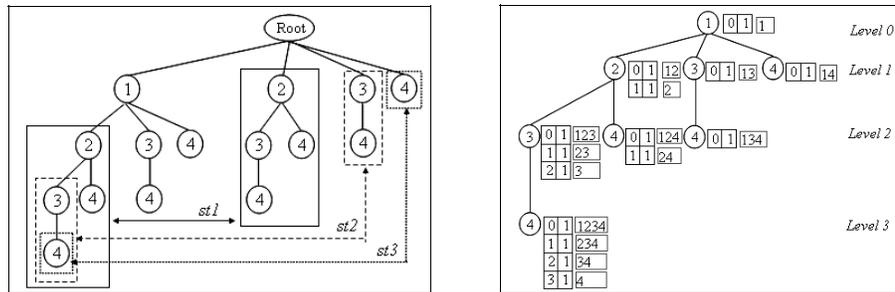
(a) Transaksi sebelum diurutkan (b) Transaksi setelah diurutkan

(c) *FP-Tree* yang dihasilkan

Gambar 1. Pengolahan Transaksi dalam Algoritma FP-Growth

3. Algoritma CT-Pro

Algoritma ini berakar dari FP-Growth dimana modifikasi yang dilakukan adalah pada struktur data yang digunakan. Struktur data yang digunakan adalah *Compressed FP-Tree (CFP-Tree)* dimana informasi dari sebuah *FP-Tree* diringkas dengan struktur yang lebih kecil, sehingga baik pembentukan *tree* maupun *frequent itemset mining* dapat dilakukan lebih cepat (lihat Gambar 2).



(a) *FP-Tree* (b) *Compressed FP-Tree* Gambar 2. *FP-Tree* dan *Compressed FP-Tree (CFP-Tree)*

ANALISA DOMAIN MASALAHNYA:

1. Algoritma Apriori

Untuk penerapan algoritma Apriori, secara umum dibutuhkan struktur data untuk menyimpan *candidate frequent itemset* untuk suatu iterasi ke k dan untuk menyimpan *frequent itemset* yang dihasilkan. Ketika membaca tiap *item* dari seluruh transaksi, selain mendapatkan *item-item* baru juga dilakukan perhitungan nilai *support item-item* yang sudah ditemukan, sehingga untuk mendapatkan *candidate 1-itemset* beserta nilai *support*-nya **cukup membutuhkan satu kali pembacaan data.**

2. Algoritma FP-Growth

Algoritma FP-Growth menggunakan struktur data *FP-Tree*. Informasi yang disimpan sebuah *node FP-Tree*: *Item*, *Index parent*, *Support*, dan *Next (Pointer)*. Ketika selesai membuat *FP-Tree*, kita tidak begitu saja bisa mendapatkan *frequent itemset* yang terdapat dalam *dataset*. Suatu kombinasi *itemset* bisa saja berada di beberapa *path* yang berbeda. Untuk mendapatkan suatu *pattern* dalam FP-Growth langkah yang lebih mudah adalah mencari arah dari ujung suatu *path*, kemudian kita mencari mulai dari *header* untuk *item* di ujung tersebut, barulah kemudian dibuat berdasarkan tiap *node* berisi *item* tersebut dicari arah *path node* ke atas. Hal ini tentu lebih cepat dari pada *up-down* karena *pointer* langsung yang dimiliki tiap *node* adalah *pointer* ke *parent*. *Path-path* yang dieksplorasi hanyalah *path-path* yang memiliki *node* yang sedang dicari. Jadi dalam struktur *FP-Tree* ada *link* dari suatu *item* ke *path-path* yang memiliki *item* tersebut, sehingga **ketika dibutuhkan pencarian *pattern-pattern* untuk suatu *item* tertentu, hanya mencari dari *path-path* tersebut saja.**

3. Algoritma CT-Pro

Algoritma CT-Pro telah diimplementasikan dalam bahasa C++ sehingga perubahan yang dilakukan lebih berfokus kepada integrasi ke dalam *workbench*. Modifikasi yang dilakukan di antaranya adalah: **mengubah struktur algoritma ke dalam bentuk objek, mengganti parameter nilai *minimum support* yang *absolute* agar bisa menerima input parameter *minimum support* dalam persentase (%), dan mengubah hasil *print out* algoritma agar bisa ditampilkan pada program.**

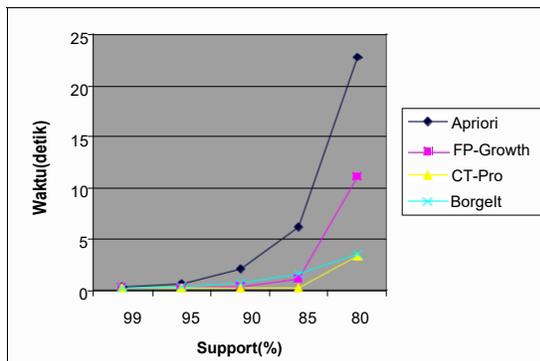
Generate Rules :

Pada tahap ini dilakukan pemeriksaan tiap *frequent itemset* dengan anggota *item* lebih dari satu. Karena *subset* dari *frequent itemset* juga *frequent*, maka langkah selanjutnya adalah mencari subset *itemset* tersebut pada kelompok *frequent itemset* berukuran lebih kecil satu dari *itemset* tersebut, jika *subset* yang terbentuk merupakan *subset* dari *itemset* maka syarat pertama *rule* terhadap *item* yang tidak ditemukan pada *subset* terpenuhi. Misalnya *item* yang tidak ditemukan adalah *item i*. Setelah dihitung nilai *confidence*-nya untuk melakukan cek apakah *rule* tersebut valid telah terbukti benar, maka *rule* {subset} $\mathcal{A} \ i$ adalah valid. Sebagai contoh, ketika memeriksa *frequent itemset* (1 2 3 4) dan (1 2 3). Karena (1 2 3) adalah *subset* dari (1 2 3 4) dan nilai *minimum confidence* dipenuhi, maka ditemukan *rule* baru yaitu $\{1,2,3\} \mathcal{A} \ 4$.

Uji Coba Implementasi :

Kedua algoritma dilihat kinerjanya pada *dataset chess* yang dibuat dari posisi dalam permainan catur. *Dataset chess* terdiri dari **3196 transaksi**, tiap transaksi terdiri dari **37 item**, dengan *item* **terkecil 1 dan item terbesar 75**. Kedua algoritma diuji dengan nilai *support* dari 99 hingga 80 dengan skala 5, gambar 4 memperlihatkan hasilnya dalam satuan detik. Untuk hasil pengujian pada *dataset chess*, perbedaan waktu mulai terlihat pada *support* 85, dimana Apriori semakin lambat dibandingkan yang lain. Namun untuk FP-Growth walaupun pada awalnya mendekati CT-Pro dan bahkan sempat lebih cepat dari Apriori Christian Borgelt, terdapat nilai *support* dimana FP-Growth memiliki penambahan waktu yang signifikan. Terlihat bahwa terdapat nilai *support* dimana **algoritma FP-Growth sangat membebani memory**. Secara umum dapat **dikatakan bahwa algoritma CT-Pro dan Apriori Christian Borgelt lebih cepat dari FP-Growth dan Apriori**.

Support	Apriori	FP-Growth	CT-Pro	Apriori Borgelt
99	0.36	0.27	0.22	0.27
95	0.66	0.28	0.23	0.42
90	2.07	0.34	0.24	0.8
85	6.2	1.09	0.26	1.57
80	22.81	11.07	3.27	3.63



Gambar 4. Hasil Pengujian pada *Dataset Chess*

Kesimpulan Pendapat :

Pada penelitian ini telah diimplementasikan algoritma *association rules* yaitu :

1. Apriori,
2. FP-Growth,
3. CT-Pro,
4. dan Apriori Cristian Borgelt

sebagai bagian dari pengembangan *data mining workbench*.

Dari hasil pengujian pada *dataset chess*, **CT-Pro** paling cepat dibandingkan dengan algoritma yang lain. **Keunggulan CT-Pro** adalah dari penggunaan memori yang lebih hemat dan digunakannya struktur data *CFP-Tree* yang memungkinkan proses pencarian *frequent itemset* menjadi lebih cepat.

Pada *support* 80, algoritma FP-Growth lebih lambat jika dibandingkan dengan Apriori Christian Borgelt karena FP-Growth mengalami penambahan ukuran *FP-Tree* di memori.

SELESAI

Judul jurnal : Penerapan Metode Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung
Pengarang : Mohamad Fauzy , Kemas Rahmat Saleh W, Ibnu Asror
Fakultas Informatika Telkom University, Bandung

Jurnal ini membahas tentang perkiraan cuaca dengan bantuan metode association rule, unsur-unsur yang terdapat di dalam data yaitu suhu, rata kelembapan, rata kecepatan angin, arah angin dan curah hujan masing-masing dianggap sebuah item. Pendekatan yang dilakukan adalah mencari kesamaan-kesamaan antar item pada tiap transaksi. Kemudian dibentuk aturan asosiatif berdasarkan kesamaan-kesamaan tersebut.

Teori yang digunakan

1. Association Rule Mining
2. Support
3. Confidence
4. Algoritma Apriori
5. Lift Ratio

Analisis dan Perancangan Sistem

1. Data training dan data testing adalah data yang sudah melalui proses preprocessing secara manual.
2. Dengan menggunakan algoritma apriori akan dilakukan training terhadap itemset yang terdapat di dalam data klimatologi yaitu dengan membentuk kandidat kandidat itemset atau biasa disebut dengan k-itemset.
3. Setiap k-itemset yang terbentuk sebelumnya kemudian dilakukan perhitungan nilai support untuk menghitung jumlah transaksi yang memuat semua item di dalam k-itemset tersebut. K-itemset yang nilai supportnya lebih tinggi dari nilai minimum support kemudian dijadikan kandidat sebagai pembentukan aturan asosiasi (association rule).
4. Dilakukan perhitungan confidence dari tiap kitemset untuk menentukan apakah kandidat tersebut dapat dijadikan sebagai aturan asosiasi (association rule) atau tidak.
5. Rule yang terbentuk di evaluasi kekuatannya dengan cara uji lift ratio. Nilai lift ratio dari sebuah rule didapatkan melalui perbandingan confidence rule tersebut dengan benchmark confidence.
6. Selanjutnya dikeluarkan hasil akhir aturan asosiasi dengan format jika (k-itemset) maka hujan atau tidak hujan. Nilai k-itemset tergantung berdasarkan k-itemset yang terbentuk.

Tujuan pengujian

1. Menganalisa pengaruh minimum support terhadap jumlah frequent itemset yang dibangkitkan algoritma apriori.
2. Menganalisa pengaruh minimum confidence terhadap jumlah rule yang dihasilkan berdasarkan frequent itemset.
3. Menganalisa pengaruh minimum support dan minimum confidence terhadap akurasi dari sistem yang telah dibangun.

- Mengetahui kekuatan dari tiap rule yang dihasilkan.

Dataset

Dataset yang digunakan oleh sistem yang dibangun berupa file bertipe excel yang berisikan data cuaca yang diambil dari BMKG stasiun geofisika kelas 1 Bandung

Analisis hasil pengujian

- Analisis Hasil Pengaruh Minimum Support dan Minimum Confidence terhadap Frequent Itemset

mincof\minsup	10%	20%	30%	40%	50%
50%	73	40	19	17	7
60%	73	40	19	17	7
70%	73	40	19	17	7

- Analisis Hasil Pengaruh Minimum Support Minimum dan Confidence terhadap Rule

mincof\minsup	10%	20%	30%	40%	50%
50%	9	5	2	2	0
60%	3	1	0	0	0
70%	2	0	0	0	0

- Analisis Hasil Pengaruh Minimum Support Minimum dan Confidence terhadap akurasi

mincof\minsup	10%	20%	30%	40%	50%
50%	76.78	76.71	49.73	49.73	0
60%	42.61	26.71	0	0	0
70%	15.96	0	0	0	0

- Analisis Hasil Lift Ratio terhadap Hasil Rule

minimum support	minimum confidence	aturan asosiatif	confidence	lift ratio
10%	50%	Agak Panas(temperature) E(windDirection) - > tidak hujan	73.08	1.46
10%	50%	E(windDirection) - > tidak hujan	72.81	1.45
10%	50%	W(windDirection) -> hujan	62.9	1.26
10%	50%	Agak Basah(humidity) Agak Kencang(windSpeed) -> tidak hujan	53.41	1.06
10%	50%	Agak Kencang(windSpeed) -> tidak hujan	53.38	1.06

10%	50%	Agak Panas(temperature) Pelan(windSpeed) - > tidak hujan	51.94	1.03
10%	50%	Pelan(windSpeed) - > tidak hujan	50.88	1.01
10%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
10%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
20%	50%	W(windDirection) -> hujan	62.9	1.26
20%	50%	Agak Panas(temperature) Pelan(windSpeed) - > tidak hujan	51.94	1.03
20%	50%	Pelan(windSpeed) - > tidak hujan	50.88	1.01
20%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
20%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
30%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
30%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01
40%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
40%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01

Implementasi *Association Rules* dengan Algoritma Apriori pada Dataset Kemiskinan

Winda Aprianti¹, Khairul Anwar Hafizd², M. Redhy Rizani³

^{1,2,3}Jurusan Teknik Informatika, Politeknik Negeri Tanah Laut

¹winda@politala.ac.id

Abstrak

Usaha pengentasan kemiskinan terus dilakukan di Kabupaten Tanah Laut. Untuk membantu pemerintah dalam perumusan kebijakan pengentasan kemiskinan maka diperlukan pengetahuan mengenai indikator yang berkaitan dengan kemiskinan dan bagaimana indikator-indikator tersebut saling mempengaruhi. Tujuan penelitian ini adalah menerapkan *Association rules* dengan algoritma Apriori pada dataset kemiskinan untuk mencari pola hubungan antar indikator. Dataset yang terdiri dari 46 atribut merupakan data sekunder BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan tahun 2010-2014. Hasil penerapan *association rules* dengan algoritma apriori menggunakan *minimum support* 30% dan *minimum confidence* 80% menghasilkan 4614 *rules* hubungan antar indikator.

Katakunci: *Association Rules, Algoritma Apriori, Indikator Kemiskinan*

1 Pendahuluan

Kemiskinan masih menjadi permasalahan yang memerlukan perhatian lebih dari pemerintah. Kalimantan Selatan adalah salah satu Provinsi di Indonesia yang mengalami permasalahan dalam penanggulangan kemiskinan. Penurunan persentase kemiskinan di Kalimantan Selatan dari

tahun 2000 sampai dengan 2013 adalah sebesar 8.29%. Persentase kemiskinan di Kabupaten Tanah Laut sendiri telah menurun 5.71% dari tahun 2000-2013 [1].

Sekarang ini, kemiskinan tidak lagi dipandang sebagai permasalahan rendahnya pendapatan seseorang, melainkan disebabkan oleh beragamnya faktor penyebab kemiskinan [2, 3, 4]. Faktor pengangguran dan Produk Domestik Regional Bruto (PDRB) termasuk yang diteliti terhadap kemiskinan pada penelitian Wijayanto (2010), Rusdati dan Sebayang (2013), tetapi memberikan hasil yang berbeda.

Hal ini menunjukkan bahwa permasalahan kemiskinan sangat kompleks dan perbedaan karakteristik wilayah, serta teknik yang digunakan juga berpengaruh untuk mengetahui indikator penyebab kemiskinan dan perumusan kebijakan untuk pengentasan kemiskinan. Penelitian tentang kemiskinan sebelumnya menggunakan data numerik, tetapi dalam penelitian ini data kemiskinan ditransformasi menjadi data kategorikal. *Task* data mining yang digunakan untuk mencari pola hubungan antar atribut adalah *Association Rules* dengan Algoritma Apriori. Penggunaan Algoritma Apriori dinilai efektif untuk menemukan pola hubungan antara beberapa atribut dengan tingkat kecelakaan lalu lintas, hubungan lingkungan fisik dengan terjadinya kebakaran hutan, serta pola pergerakan harga saham antar sektor dan antar perusahaan [5, 6, 7]. Berdasarkan penelitian terdahulu, maka pada penelitian ini *Association Rules* dengan Algoritma Apriori diimplementasikan pada dataset kemiskinan.

2 Tinjauan Pustaka

2.1 Penelitian Terkait

Sitanggang (2013) menerapkan *association rules* dengan Algoritma Apriori pada dataset kebakaran hutan yang berisi data pada lingkungan fisik (tutupan lahan, sungai, jalan dan pusat kota), sosial-ekonomi (sumber pendapatan, populasi, dan jumlah sekolah), cuaca (hujan, angin kecepatan, dan temperatur layar), dan lahan gambut. Hasil penelitian mengungkapkan

324 aturan asosiasi multidimensi yang menunjukkan hubungan antara hotspot kejadian dan faktor lainnya. Asosiasi antara hotspot terjadinya dengan objek geografis lainnya ditemukan untuk *minimum support* 10% dan *minimum confidence* 80% [6]. Algoritma apriori juga digunakan untuk menentukan pola hubungan antara usia, jenis kecelakaan, waktu terjadinya kecelakaan, kepemilikan Surat Ijin Mengemudi (SIM), jenis kelamin, dan pekerjaan dengan tingkat kecelakaan kecelakaan lalu lintas di Jalan Raya Kabupaten Sleman [5]. Penggunaan variabel *fuzzy* sebelum penerapan Algoritma Apriori diterapkan pada dataset harga saham untuk mengetahui pola pergerakan harga saham antar perusahaan [7].

2.2 Association Rules

Association rules adalah salah satu *task* data mining deskriptif yang bertujuan untuk menemukan aturan asosiatif antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam *database*, yang disebut sebagai *frequent patterns* [8]. Pramudiono dalam [7] menyatakan bahwa penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi item dalam *database* dan *confidence* yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Support dan *confidence* dituliskan sebagai Persamaan 1 dan Persamaan 2 [8].

$$\text{support}(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$\text{confidence}(A \Rightarrow B) = P(B | A) \quad (2)$$

Jika *support itemset* dari *itemset I* memenuhi *minimum support threshold* yang sudah ditentukan maka *I* adalah *frequent k-itemset*. Secara umum *frequent k-itemset* dilambangkan dengan *L_k*. Berdasarkan Persamaan (2) diperoleh

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \quad (3)$$

2.3 Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma untuk melakukan pencarian *frequent itemset* dengan *association rules*. Algoritma Apriori menggunakan pendekatan *level-wise search*, dimana *k-itemset* digunakan untuk memperoleh *(k+1)-itemset*. Proses ini dilakukan hingga tidak ada lagi kombinasi yang dapat dibentuk[8]. Berdasarkan persamaan (1), (2), dan (3), serta pembahasan pada [9], diperoleh Persamaan (4), (5), dan (6).

Persamaan (4) digunakan untuk menghitung nilai *support* untuk sebuah *item*.

$$\text{Support } (A) = \frac{\text{Jumlah Transaksi yang mengandung } A}{\text{Total Transaksi}} \quad (4)$$

Persamaan (5) digunakan untuk menghitung nilai *support* dari 2 *item*.

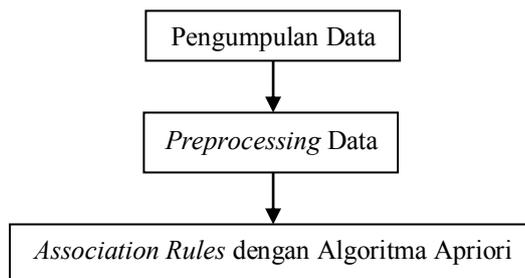
$$\text{Support } (A, B) = \frac{\text{Jumlah Transaksi yang mengandung } A \text{ dan } B}{\text{Total Transaksi}} \quad (5)$$

Persamaan (6) digunakan untuk menghitung nilai *confidence* dari *rule* $A \rightarrow B$.

$$\text{Confidence } (A \rightarrow B) = \frac{\text{Jumlah Transaksi yang mengandung } A \text{ dan } B}{\text{Jumlah Transaksi yang mengandung } A} \quad (6)$$

3 Metodologi Penelitian

Metodologi penelitian disajikan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

3.1 Pengumpulan Data

Dataset yang digunakan merupakan data sekunder dari BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan Tahun 2009-2014. Adapun dataset yang berjumlah 46 atribut ini terdiri dari:

1. Jumlah penduduk miskin (A1).
2. Pengeluaran rata-rata perkapita untuk kelompok makanan (A2).
3. Pengeluaran rata-rata perkapita untuk kelompok non makanan (A3).
4. Produk Domestik Regional Bruto (PDRB) yang terbagi menjadi Atas Dasar Harga Berlaku (ADHB) dan Atas Dasar Harga Konstan (ADHK) pada setiap sektornya sejumlah 14 sektor, yaitu PDRB Sektor Bangunan; PDRB Sektor Industri Pengolahan; PDRB Sektor Jasa-Jasa; PDRB Sektor Keuangan; PDRB Sektor Persewaan dan Jasa Perusahaan; PDRB Sektor Listrik dan Gas; PDRB Sektor Air Bersih; PDRB Sektor Pengangkutan; PDRB Sektor Komunikasi; PDRB Sektor Perdagangan, Hotel dan Restoran; PDRB Sektor Pertambangan dan Penggalian; PDRB Sektor Pertanian dan Tanaman Pangan; PDRB Sektor Perkebunan; PDRB Sektor Peternakan; PDRB Sektor Kehutanan; serta PDRB Sektor Perikanan (A4 sampai dengan A35).
5. Inflasi (A36).
6. Luas panen tanaman padi dan palawija yang terdiri dari 7 kelompok, yaitu padi; jagung; kedelai; kacang tanah; kacang hijau; ubi kayu; dan ubi jalar (A37 sampai dengan A43).
7. Jumlah industri manufaktur besar dan sedang (A44).
8. Jumlah pengangguran (A45).
9. Jumlah penduduk (A46).

Semua atribut pada dataset merupakan data numerik. Namun data inflasi dan luas panen tanaman berupa data bulanan, sedangkan data yang lain berupa data tahunan. Sehingga data inflasi dan luas panen tanaman diubah menjadi data tahunan dengan cara menghitung rata-rata dari masing-masing data tersebut.

3.2 Preprocessing Data

Tahapan *preprocessing* data dibagi menjadi 2 tahap, yaitu *cleaning* data dan transformasi data kategorikal.

Cleaning data untuk menghilangkan data yang tidak konsisten. Atribut ADHB (A6) dan ADHK (A22) Sektor Jasa-jasa untuk Tahun 2009-2010 dan Tahun 2011-2014 mempunyai pengumpulan data jasa-jasa yang berbeda sehingga kedua atribut harus dihilangkan dari dataset. Selain kedua atribut tersebut, atribut jumlah industri besar dan sedang (A44) juga perlu dihilangkan dari dataset karena data untuk tahun 2011-2014 tidak diketahui. Sehingga dataset sekarang berjumlah 43 atribut.

Berdasarkan kajian pada [7], maka proses transformasi data dimulai dengan melakukan perhitungan selisih dari data yang telah terurut dan mencari persentase data. Hasil persentase data kemudian dibagi menjadi 3 bagian yaitu kenaikan, tetap, dan penurunan. Untuk pergerakan yang tidak berubah diberikan kategori Tetap. Sedangkan untuk menentukan kategori kenaikan dan penurunan, data difuzzifikasi menggunakan fungsi keanggotaan segitiga tipe bahu untuk nilai persentase kenaikan dan penurunan. Setiap fungsi keanggotaan kenaikan dan penurunan dibagi menjadi 3 kategori, yaitu Rendah, Sedang, dan Tinggi. Hal ini menyebabkan setiap atribut numerik mempunyai kemungkinan berubah menjadi 7 data kategorikal, sehingga dataset yang awalnya berukuran 43×5 sekarang menjadi dataset berukuran 301×5 .

3.3 Association Rules dengan Algoritma Apriori

Setelah dataset ditransformasi menjadi atribut kategorikal, berikutnya diterapkan *association rules* dengan algoritma Apriori. Langkah-langkah pada tahap ini adalah sebagai berikut.

1. Menentukan *minimum support* dan *minimum confidence*.
2. Menghitung *k-itemset*, dengan *k* adalah jumlah kombinasi *itemset*.

3. Menentukan *frequent(k) itemset*, yaitu *itemset* yang memenuhi *minimum support*.
4. Menentukan *rules*.

4 Hasil Penelitian

Dataset hasil *preprocessing* yang menghasilkan 43 atribut numerik disajikan pada Tabel 1.

Tabel 1. Dataset Hasil *Preprocessing* Data

No	Nama Atribut	Atribut	Tahun					
			2009	2010	2011	2012	2013	2014
1	Penduduk Miskin	A1	14039	13654	14104	14698	15218	13856
2	Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan	A2	308434	345529	364140	424891	474028	474164
3	Pengeluaran Rata-Rata per Kapita untuk Kelompok Non Makanan	A3	264873	242825	326409	383895	370033	369209
...
41	Luas Panen Tanaman Ubi Jalar (Ha)	A43	151	174	180	203	205	228
42	Jumlah Pengangguran	A45	7992	5991	7962	6519	4589	4775
43	Jumlah Penduduk	A46	274526	297814	303190	308510	131725	319098

Dataset pada Tabel 1 kemudian ditransformasi menjadi data kategorikal. Apabila hasil selisih data menunjukkan kenaikan maka digunakan fungsi keanggotaan *fuzzy* yang disajikan pada Persamaan 4, 5, dan 6.

Fungsi keanggotaan untuk kategori Naik Rendah (NR):

$$\mu_{NR}(x) = \begin{cases} 1 & , x \leq 33.33 \\ \frac{40-x}{6.67} & , 33.33 < x < 40 \\ 0 & , x \geq 40 \end{cases} \quad (4)$$

Fungsi keanggotaan untuk kategori Naik Sedang (NS):

$$\mu_{NS}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-33.33}{16.67} & , 33.33 \leq x < 50 \\ \frac{66.67-x}{16.67} & , 50 < x < 66.67 \\ 0 & , x \geq 66.67 \end{cases} \quad (5)$$

Fungsi keanggotaan untuk kategori Naik Tinggi (NT):

$$\mu_{NT}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-60}{6.67} & , 60 < x < 66.67 \\ 1 & , x \geq 66.67 \end{cases} \quad (6)$$

Sedangkan apabila hasil selisih data menunjukkan penurunan maka digunakan Persamaan 7, 8, dan 9.

Fungsi keanggotaan untuk kategori Turun Rendah (TR):

$$\mu_{TR}(x) = \begin{cases} 1 & , x \geq -33.33 \\ \frac{x+40}{6.67} & , -40 < x < -33.33 \\ 0 & , x \leq -40 \end{cases} \quad (7)$$

Fungsi keanggotaan untuk kategori Turun Sedang (TS):

$$\mu_{TS}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{-x-33.33}{16.67} & , -50 \leq x < -33.33 \\ \frac{x+66.67}{16.67} & , -66.67 < x < -50 \\ 0 & , x \leq -66.67 \end{cases} \quad (8)$$

Fungsi keanggotaan untuk kategori Turun Tinggi (TT):

$$\mu_{TT}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{x+60}{6.67} & , -66.67 < x < -60 \\ 1 & , x \leq -66.67 \end{cases} \quad (9)$$

Hasil transformasi data disajikan pada Tabel 2.

Tabel 2. Dataset Hasil Transformasi Data

No	Atribut	Tahun					
		2009	2010	2011	2012	2013	2014
1	A1	TR_A1	NR_A1	NR_A1	NR_A1	TR_A1	TR_A1
2	A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2
3	A3	TR_A3	NR_A3	NR_A3	TR_A3	TR_A3	TR_A3
...
41	A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43
42	A45	TR_A45	NR_A45	TR_A45	TR_A45	NR_A45	TR_A45
43	A46	NR_A46	NR_A46	NR_A46	TS_A46	NT_A46	NR_A46

Pada Tabel 2, dataset telah berupa atribut kategori: Turun Rendah (TR), Turun Sedang (TS), Turun Tinggi (TT), Tetap (T), Naik Rendah (NR), Naik Sedang (NS), dan Naik Tinggi (NT). Dataset ini kemudian dibentuk menjadi *itemset* yang disajikan pada Tabel 3.

Tabel 3. *Itemset*

No	Item	waktu (t)					Jumlah
		t1	t2	t3	t4	t5	
1	NR_A1	0	1	1	1	0	3
2	NS_A1	0	0	0	0	0	0
3	NT_A1	0	0	0	0	0	0
...
299	TR_A46	0	0	0	0	0	0
300	TS_A46	0	0	0	1	0	1
301	TT_A46	0	0	0	0	0	0

Dengan *minimum support* = 30% dan *minimum confidence* = 80%, penerapan *association rules* dengan algoritma Apriori menghasilkan 1-*itemsets*, 2-*itemsets*, 3-*itemsets*, 4-*itemsets*, 5-*itemsets*, 6-*itemsets*, 7-*itemsets*, dan 8-*itemsets* seperti yang disajikan pada Gambar 2.

```

Associator output
Apriori
=====
Minimum support: 0.95 (5 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8
Size of set of large itemsets L(2): 28
Size of set of large itemsets L(3): 56
Size of set of large itemsets L(4): 70
Size of set of large itemsets L(5): 56
Size of set of large itemsets L(6): 28
Size of set of large itemsets L(7): 8
Size of set of large itemsets L(8): 1

Best rules found:

1. A4=NR_A4 5 ==> A2=NR_A2 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. A2=NR_A2 5 ==> A4=NR_A4 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

```

Gambar 2. Hasil Penerapan *Association Rules* dengan Algoritma Apriori Menggunakan Weka

Setelah mendapatkan *itemsets*, maka selanjutnya diperoleh 4614 *rules* yang menunjukkan hubungan antar indikator (Tabel 4).

Tabel 4. *Rules* yang diperoleh

No	Rule
1	A4=NR_A4 ==> A2=NR_A2
2	A2=NR_A2 ==> A4=NR_A4
3	A5=NR_A5 ==> A2=NR_A2
...	...
4612	A5=NR_A5 A23=NR_A23 A25=NR_A25 A33=NR_A33 A43=NR_A43 ==> A29=NR_A29
4613	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A43=NR_A43 ==> A33=NR_A33
4614	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A33=NR_A33 ==> A43=NR_A43

Pada Tabel 4 untuk *rule* 1 diperoleh pola hubungan antara indikator 4, yaitu ADHB Sektor Bangunan dengan indikator 2, yaitu Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan yang menunjukkan bahwa apabila ADHB Sektor Bangunan naik rendah maka Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan juga akan mengalami naik rendah. Pola hubungan yang diperoleh dari *rule* 1 ini dapat membantu pemerintah dalam pengambilan kebijakan yang berkaitan dengan ADHB Sektor Bangunan dan Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan dengan memperhatikan pola hubungan kedua atribut ini. Begitu juga dengan *rule* 2 sampai dengan *rule* 4614.

5 Kesimpulan

Berdasarkan hasil penerapan *association rules* dengan algoritma Apriori pada dataset kemiskinan menggunakan *minimum support* 30% dan *minimum confidence* 80% diperoleh 4614 *rules* hubungan antar indikator. Namun, banyaknya jumlah *rules* yang dihasilkan mengakibatkan banyaknya pengetahuan mengenai pola hubungan antar indikator. Pengetahuan yang terlalu luas ini berakibat pada kesulitan pengambilan keputusan untuk menentukan pola hubungan antar indikator mana yang lebih berpengaruh terhadap kemiskinan.

Guna menghasilkan pola hubungan indikator yang lebih spesifik maka akan diterapkan *Principal Component Analysis* sebagai salah satu *preprocessing* data pada penerapan Algoritma Apriori untuk penelitian selanjutnya.

6 Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Kemenristekdikti atas bantuan biaya pada skema Penelitian Dosen Pemula Tahun Anggaran 2017.

7 Pustaka

- [1] Badan Pusat Statistik (BPS), diakses dari <http://www.bps.go.id> pada tanggal 16 Maret 2016 jam 10.00 WITA.
- [2] Nurwati, N., 2008, “Kemiskinan: Model Pengukuran, Permasalahan, dan Alternatif Kebijakan”, *Jurnal Kependudukan Padjadjaran*, Vol. 10, No. 1, Januari 2008: 1-11.
- [3] Rusdarti dan Sebayang, K.L., 2013, “Faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Jawa Tengah”, *Jurnal Economia*, Volume 9, Nomor 1, April 2013.
- [4] Wijayanto, R. D., 2010, Analisis Pengaruh PDRB, Pendidikan, dan Pengangguran terhadap Kemiskinan di Kabupaten/Kota Jawa Tengah Tahun 2005 – 2008, Skripsi, Fakultas Ekonomi Universitas Diponegoro, Semarang.
- [5] Hakim, L. dan Fauzy, A., 2015, “Penentuan Pola Hubungan Kecelakaan Lalu Lintas Menggunakan Metode *Association Rules* dengan Algoritma *Apriori*”, *University Research Colloquium 2015*, ISSN 2407-9189.
- [6] Sitanggang, S. I., 2013, “Spatial Multidimensional Association Rules Mining in Forest Fire Data”, *Journal of Data Analysis and Information Processing*, 2013, 1, 90-96.
- [7] Arafah, A.A. dan Mukhlash, I., 2015, “The Application of Fuzzy Association Rule on Co-Movement Analyze of Indonesian Stock Price”, *Procedia Computer Science* 59 (2015) 235-243.
- [8] Han, J., Kamber, M., dan Pei, J., 2011, *Data Mining: Concepts and Techniques Third Edition*, Morgan Kaufmann, USA.
- [9] Kusriani dan Lutfi, E.T., 2009, *Algoritma Data Mining*, Andi, Yogyakarta.

Association Rules

- ❖ Adalah proses mendeteksi kumpulan atribut-atribut yang muncul bersamaan (co-occur) dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.
- ❖ Contoh : 90% orang yang berbelanja di suatu supermarket yang membeli roti juga membeli selai, dan 60% dari semua orang yang berbelanja membeli keduanya.
- ❖ *Association rule mining adalah analisa* dari kebiasaan belanja konsumen dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan konsumen dalam keranjang belanjanya (Yang, 2003)
- ❖ Contoh *Association rule* misalnya: "70% dari orang-orang yang membeli mie, juice dan saus akan membeli juga roti tawar".
- ❖ Dengan kemajuan teknologi, data penjualan dapat disimpan dalam jumlah besar yang disebut dengan "*basket data*."
- ❖ *Aturan asosiasi yang didefinisikan* pada basket data tersebut, dapat digunakan untuk menganalisa data dalam rangka :
 - ❖ keperluan desain katalog promosi,
 - ❖ proses pembuatan keputusan bisnis,
 - ❖ segmentasi konsumen dan target pemasaran.

Contoh aplikasi

- ❖ Marketing and Sales Promotion
 - Misal :
 - Ketergantungan {bagels, ... } → {Potato Chips}
 - Potato Chips sebagai consequent → dapat digunakan untuk menentukan apa yang dilakukan untuk meningkatkan penjualan
 - Bagels in the antecedent → dapat digunakan untuk melihat produk mana yang akan terkena dampak jika toko tersebut tidak lagi menjual bagels.
- ❖ Supermarket Shelf Management
- ❖ Inventory Management

Apriori Algoritma

- ❖ Algoritma Apriori pertama kali dikenalkan oleh Agrewal, Imielinski dan Swami.

- ❖ Algoritma Apriori merupakan salah satu algoritma yang digunakan di dalam memecahkan persoalan *association rule mining*.
- ❖ Yang mengolah suatu database transaksi dengan setiap transaksi adalah suatu himpunan item-item. Kemudian mencari seluruh kaidah apriori yang memenuhi kendala minimum support dan minimum confidence yang diberikan user.
- ❖ Algoritma Apriori dapat digunakan untuk menemukan tren bisnis dengan menganalisa transaksi konsumen.
- ❖ Contoh: 30% dari transaksi yang memuat bir juga memuat popok 5% yang artinya 30% merupakan confidence dan 5% merupakan support dari kaidah ini.

Pseudo-code apriori algoritma

C_k: Kandidat itemset dari ukuran *k*;

L_k : Frequent itemset dari ukuran *k*.

L₁ = {frequent items};

for (*k* = 1; *L_k* != 0; *k*++) do begin

C_{k+1} = {kandidat dibangun dari

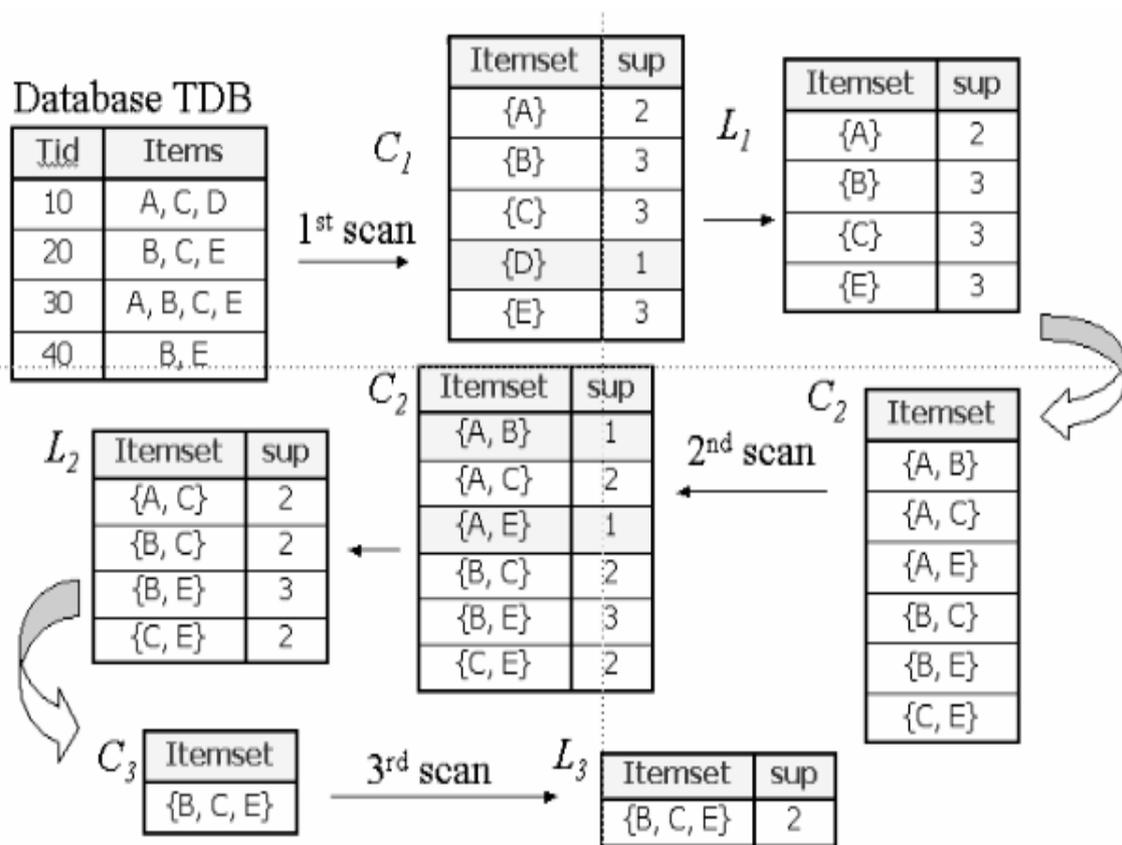
for each transaksi *t* yang dimuat dalam database do naikkan hitungan dari seluruh kandidat dalam *C_{k+1}* yang dimuat dalam *t*

L_{k+1} = {kandidat dalam *C_{k+1}* dengan *min_support*}

end

return .*k L_k*;

ilustrasi



- ❖ Berikut ini Contoh dari 4 Transaksi belanja konsumen, akan dicari hubungan asosiasi antar item dengan minimal support 50%

No	Itemset
1	A.Kopi, C.Gula, D.Bir
2	B.Teh, C.Gula, E.Roti
3	A.Kopi, B.Teh, C.Gula, E.Roti
4	B.Teh, E.Roti

- ❖ Langkah 1: $L1 = \{\text{large 1-itemset}\}$

Jumlah transaksi = 4.

Min support = 50% artinya 2 dari 4 transaksi.

Langkah 1: $L1 = \{\text{large 1-itemset}\}$

Itemset	Support
A	50%
B	75%
C	75%
D	25%
E	75%

- ❖ Langkah 2 : Mencari kandidat itemset untuk L2

- 2.1 Gabungkan itemset pada L1 (algoritma apriori gen)

{ A B, A C, A D, A E, B C, B D, B E, C D, C E, D E }

- 2.2 Hapus yang tidak ada dalam itemset

Itemset { B D, D E } dihapus karena tidak ada dalam itemset

Itemset	Support
A B	25%
A C	50%
A D	25%
A E	25%
B C	50%
B E	75%
C D	25%
C E	50%

- ❖ Langkah 3 : Hitung Support dari tiap kandidat itemset

❖ Langkah 4 : L2 {Large 2-itemset}

Itemset	Support
AB	25 %
AC	50 %
AD	25 %
AE	25 %
BC	50 %
BE	75 %
CD	25 %
CE	50 %

Itemset	Support
AC	50 %
BC	50 %
BE	75 %
CE	50 %

❖ Langkah 5 : Ulangi langkah 2-4

- 5.1 Gabungkan itemset pada L2 dan L2 :
- 5.2 Hapus yang tidak ada dalam itemset : {ACE}

❖ Langkah 6 : Hitung Support dari setiap kandidat itemset L3

❖ Langkah 7 : L3 {Large 3-Itemset} {BCE}

❖ Langkah 8 : STOP karena sudah tidak ada lagi kandidat untuk 4-itemset

Langkah 5

Langkah 6

Itemset	Support
ABC	25 %
ABE	25 %
BCE	50 %

Itemset	Hasil Gabungan (3 itemset)
AC+BC	ACB
AC+BE	ACB, ACE, ABE
AC+CE	ACE
BC+BE	BCE
BC+CE	BCE
BE+CE	BCE

Hasil Akhir :

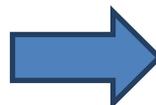
L1		L2		L3	
A	50%	A C	50%	B C E	50%
B	75%	B C	50%		
C	75%	B E	75%		
D	25%	C E	50%		
E	75%				

Untuk mencari aturan asosiasi

diperlukan juga min-confidence :

Misal min-conf : 75%.

Aturan yang mungkin terbentuk :



Aturan ($X \rightarrow Y$)	Sup($X \cup Y$)	Sup(X)	Confidence
B C \rightarrow E	50%	50%	100%
B E \rightarrow C	50%	75%	66.67%
C E \rightarrow B	50%	50%	100%
A \rightarrow C	50%	50%	100%
C \rightarrow A	50%	75%	66.67%
B \rightarrow C	50%	75%	66.67%
C \rightarrow B	50%	75%	66.67%
B \rightarrow E	75%	75%	100%
E \rightarrow B	75%	75%	100%
C \rightarrow E	50%	75%	66.67%
E \rightarrow C	50%	75%	66.67%

Association Rule Mining

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (frequent pattern mining) [5].

Secara umum association rule mempunyai bentuk : LHS => RHS dimana LHS dan RHS tersebut adalah himpunan item; jika setiap item-item dalam LHS terdapat dalam transaksi maka item-item dalam RHS juga terdapat dalam transaksi.

Aturan asosiasi biasanya dinyatakan dalam bentuk[5]:

$$\{A,B\} \Rightarrow \{C\} \text{ (support = 10\%, confidence = 50\%)}$$

Support

Support dari suatu association rule adalah presentasi kombinasi item tersebut dalam database, dimana jika mempunyai item A dan item B maka support adalah proporsi dari transaksi dalam database yang mengandung A dan B. Rumus untuk menghitung nilai support dari dua item tersebut adalah sebagai berikut[5][6]:

$$\text{Support} (A, B) = P(A \cap B)$$

$$\text{Support} (A, B) = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi}}$$

2.3 Confidence

Confidence dari association rule adalah ukuran ketepatan suatu rule, yaitu presentasi transaksi dalam database yang mengandung A dan mengandung B. Dengan adanya confidence kita dapat mengukur kuatnya hubungan antar-item dalam association rule. Rumus untuk menghitung nilai confidence dari dua item tersebut adalah sebagai berikut[5][6]:

$$\text{Confidence} = P(B | A)$$

$$\text{Confidence} = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi mengandung A}}$$

Algoritma Apriori

Algoritma apriori adalah langkah untuk proses menemukan frequent-itemset dengan melakukan iterasi pada data. Dimana itemset adalah himpunan item-item yang berada di dalam himpunan yang diolah oleh sistem, sedangkan frequent-itemset menunjukkan itemset yang memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan (ϕ). Pada iterasi ke- k , semua itemset yang ditemukan yang memiliki k item disebut k -itemset. Setiap iterasi terdiri dari dua tahap yaitu pembangkitan kandidat dan pembangkitan rule.

Pada tahap pembangkitan kandidat (candidate generation) dimana himpunan semua frequent($k - 1$)-itemset yang digunakan pada pass ke- $(k-1)$ digunakan untuk membangkitkan kandidat itemset

. Prosedur pembangkitan kandidat menjamin bahwa adalah superset dari himpunan. Kemudian di-scan dalam tahap perhitungan support (support counting). Pada akhir pass diperiksa untuk menentukan kandidat mana yang muncul, menghasilkan . Perhitungan support berakhir ketika atau kosong.

Pada tahap membangkitkan rule, akan dibangkitkan lebih dahulu candidate rule. Candidate rule berisi semua kemungkinan rule yang memiliki support > minimum support karena inputan candidate rule adalah frequent-itemset. Kemudian candidate rule akan di-join dengan table F untuk menemukan support antecedent. Confidence rule dihitung dengan cara membandingkan support rule dengan support antecedent rule. Hanya rule yang mempunyai confidence > minimum confidence yang disimpan dalam table rule (table R)[6].

Lift Ratio

Lift ratio adalah suatu ukuran untuk mengetahui kekuatan aturan asosiasi (association rule) yang telah terbentuk. Nilai lift ratio biasanya digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid. Untuk menghitung lift ratio digunakan rumus sebagai berikut:

$$\text{Lift Ratio} = \frac{\text{Confidence} (A, B)}{\text{Benchmark Confidence} (A, B)}$$

Untuk mendapatkan nilai benchmark confidence sendiri dapat dihitung menggunakan rumus sebagai berikut:

$$\text{Benchmark Confidence} = \frac{Nc}{N}$$

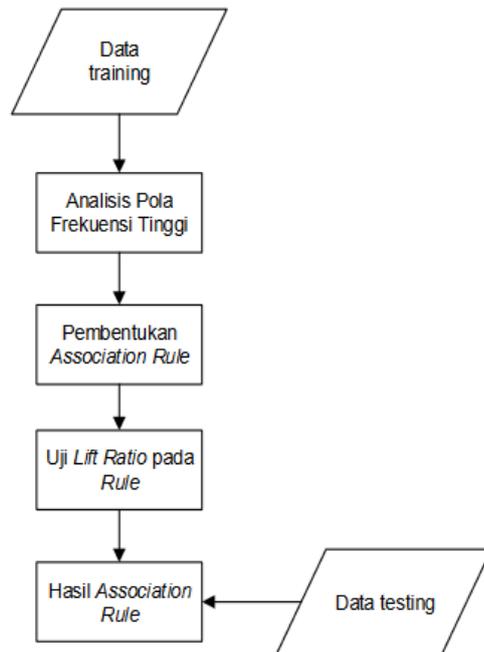
Keterangan:

- Nc = jumlah transaksi dengan item yang menjadi consequent

- N = jumlah transaksi basis data

Analisis dan Perancangan Sistem

Gambar 1. Perancangan Umum Sistem



Skenario keseluruhan sistem :

1. Data training dan data testing adalah data yang sudah melalui proses preprocessing secara manual.
2. Dengan menggunakan algoritma apriori akan dilakukan training terhadap itemset yang terdapat di dalam data klimatologi yaitu dengan membentuk kandidat-kandidat itemset atau biasa disebut dengan k-itemset.
3. Setiap k-itemset yang terbentuk sebelumnya kemudian dilakukan perhitungan nilai support untuk menghitung jumlah transaksi yang memuat semua item di dalam k-itemset tersebut. K-itemset yang nilai supportnya lebih tinggi dari nilai minimum support kemudian dijadikan kandidat sebagai pembentukan aturan asosiasi (association rule).
4. Dilakukan perhitungan *confidence* dari tiap k-itemset untuk menentukan apakah kandidat tersebut dapat dijadikan sebagai aturan asosiasi (association rule) atau tidak.
5. Rule yang terbentuk di evaluasi kekuatannya dengan cara uji lift ratio. Nilai lift ratio dari sebuah rule didapatkan melalui perbandingan *confidence* rule tersebut dengan *benchmark confidence*.
6. Selanjutnya dikeluarkan hasil akhir aturan asosiasi dengan format jika (k-itemset) maka hujan atau tidak hujan. Nilai k-itemset tergantung berdasarkan k-itemset yang terbentuk.

4. Pengujian

Tujuan Pengujian

Tujuan dari pengujian yang dilakukan adalah sebagai berikut ini:

1. Menganalisa pengaruh *minimum support* terhadap jumlah *frequent itemset* yang dibangkitkan algoritma apriori.
2. Menganalisa pengaruh *minimum confidence* terhadap jumlah *rule* yang dihasilkan berdasarkan *frequent itemset*.
3. Menganalisa pengaruh *minimum support* dan *minimum confidence* terhadap akurasi dari sistem yang telah dibangun.
4. Mengetahui kekuatan dari tiap *rule* yang dihasilkan.

Dataset

Dataset yang digunakan oleh sistem yang dibangun berupa *file* bertipe excel yang berisikan data cuaca yang diambil dari BMKG stasiun geofisika kelas 1 Bandung. Penginputan data ke *file* excel dilakukan secara manual karena data yang di dapat dari BMKG merupakan data fisik. Lalu data akan masuk ke proses *data preprocessing* yang akan menghasilkan dataset yang akan dipakai oleh sistem.

5. Analisis Hasil Pengujian

Analisis Hasil Pengaruh *Minimum Support* dan *Minimum Confidence* terhadap *Frequent Itemset*

Tabel 1. *frequent itemset*

mincof\minsup	10%	20%	30%	40%	50%
50%	73	40	19	17	7
60%	73	40	19	17	7
70%	73	40	19	17	7

Berdasarkan tabel 1, dapat dilihat jumlah *frequent itemset* terbanyak dibangkitkan oleh *minimum support* terkecil dalam pengujian ini yaitu 10%, sedangkan *frequent itemset* paling sedikit dibangkitkan oleh *minimum support* sebesar 50%. Hal ini dikarenakan nilai *minimum support* adalah nilai acuan dalam perhitungan sebuah *itemset*. Hanya *itemset* yang nilai supportnya sama atau lebih besar dari nilai *minimum support* saja yang dijadikan sebagai *frequent itemset*. Sedangkan *minimum confidence* tidak memiliki pengaruh dalam pembangkitan *frequent itemset* karena tidak terjadi perhitungan *confidence* pada tahap ini. Maka dapat disimpulkan bahwa semakin tinggi *minimum support* yang digunakan maka akan semakin sedikit jumlah *frequent itemset* yang dibangkitkan.

Analisis Hasil Pengaruh *Minimum Support* dan *Minimum Confidence* terhadap *Rule*

Tabel 2. Jumlah *rule*

mincof\minsup	10%	20%	30%	40%	50%
50%	9	5	2	2	0
60%	3	1	0	0	0
70%	2	0	0	0	0

Berdasarkan tabel 2, dapat dilihat jumlah aturan asosiatif yang terbentuk terbanyak dihasilkan oleh *minimum support* 10% dan *minimum confidence* 50% yaitu sebanyak 9 *rule*. Pada setiap penambahan nilai *minimum support* dan *minimum confidence* yang diujikan terlihat *rule* yang terbentuk semakin berkurang. Penurunan ini dipengaruhi oleh jumlah *frequent itemset* yang dibangkitkan pada pengujian sebelumnya. Pada *minimum support* 50% dapat dilihat tidak terbentuk satu pun aturan asosiatif, hal ini diakibatkan nilai *confidence* dari *frequent itemset* yang dibangkitkan tidak ada yang memenuhi syarat *minimum confidence* yang diujikan. Oleh karena itu, dapat dilihat bahwa *minimum confidence* berpengaruh pada pembentukan *rule*. Pada pengujian ini dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka *rule* yang dihasilkan akan semakin sedikit.

Analisis Hasil Pengaruh *Minimum Support* dan *Minimum Confidence* terhadap akurasi

Tabel 3. Akurasi data training

mincof\minsup	10%	20%	30%	40%	50%
50%	76.78	76.71	49.73	49.73	0
60%	42.67	26.71	0	0	0
70%	15.96	0	0	0	0

Berdasarkan tabel 3 di atas, akurasi tertinggi diperoleh sebesar 76.78% dengan *minimum support* 10% dan *minimum confidence* 50%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya

tidak dapat dihitung besarnya akurasi karena tidak ada rule yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data training diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data training. Dalam pengujian akurasi data training dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 76.71%.

Tabel 4. Akurasi data testing

minconf\minsup	10%	20%	30%	40%	50%
50%	75.89	75.89	50.41	50.41	0
60%	30.69	24.11	0	0	0
70%	6.58	0	0	0	0

Berdasarkan tabel 4 di atas, akurasi tertinggi diperoleh sebesar 75.89% terdapat pada dua minimum support berbeda yaitu *minimum support* 10% dan *minimum support* 20% dengan *minimum confidence* 50%. Akurasi berikutnya juga diperoleh pada dua *minimum support* yang berbeda yaitu *minimum support* 30% dan *minimum support* 40% dengan *minimum confidence* 50% sebesar 50.41%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya tidak dapat dihitung besarnya akurasi karena tidak ada rule yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data testing diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data testing. Dalam pengujian akurasi data testing dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 75.89%.

Analisis Hasil Lift Ratio terhadap Hasil Rule

Tabel 5. Lift ratio tiap rule

minimum support	minimum confidence	aturan asosiatif	confidence	lift ratio
10%	50%	Agak Panas(temperature) E(windDirection) -> tidak hujan	73.08	1.46
10%	50%	E(windDirection) -> tidak hujan	72.81	1.45
10%	50%	W(windDirection) -> hujan	62.9	1.26
10%	50%	Agak Basah(humidity) Agak Kencang(windSpeed) -> tidak hujan	53.41	1.06
10%	50%	Agak Kencang(windSpeed) -> tidak hujan	53.38	1.06
10%	50%	Agak Panas(temperature) Pelan(windSpeed) -> tidak hujan	51.94	1.03
10%	50%	Pelan(windSpeed) -> tidak hujan	50.88	1.01
10%	50%	Agak Basah(humidity) -> tidak hujan	50.66	1.01
10%	50%	Agak Panas(temperature) -> tidak hujan	50.62	1.01

20%	50%	<i>W(windDirection)</i> -> hujan	62.9	1.26
20%	50%	<i>Agak Panas(temperature) Pelan(windSpeed) -</i> -> tidak hujan	51.94	1.03
20%	50%	<i>Pelan(windSpeed) -</i> -> tidak hujan	50.88	1.01
20%	50%	<i>Agak Basah(humidity) -></i> tidak hujan	50.66	1.01
20%	50%	<i>Agak Panas(temperature)</i> -> tidak hujan	50.62	1.01
30%	50%	<i>Agak Basah(humidity) -> tidak hujan</i>	50.66	1.01
30%	50%	<i>Agak Panas(temperature)</i> -> tidak hujan	50.62	1.01
40%	50%	<i>Agak Basah(humidity) -></i> tidak hujan	50.66	1.01
40%	50%	<i>Agak Panas(temperature)</i> -> tidak hujan	50.62	1.01

Berdasarkan tabel 5 diatas *lift ratio* untuk semua *rule* yang berhasil dibentuk memiliki nilai lebih besar dari 1 (*lift ratio* >1). Hal ini menunjukkan bahwa semua *rule* tersebut bersifat kuat dan valid untuk digunakan sebagai acuan dalam memprediksi Hujan.

KESIMPULAN

Berdasarkan pengujian yang dilakukan diatas, dapat disimpulkan bahwa :

1. Metode *association rule mining* dengan menggunakan algoritma apriori dapat diterapkan pada sistem simulasi prediksi hujan.
2. Semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka semakin sedikit jumlah *frequent itemset* dan *rule* yang terbentuk serta akurasi semakin berkurang.
3. Semua rule yang dihasilkan pada penelitian ini memiliki nilai *lift ratio* lebih dari 1.00 sehingga dapat digunakan sebagai acuan dalam memprediksi hujan.

SARAN

Saran yang diperlukan untuk pengembangan sistem lebih lanjut adalah sebagai berikut :

1. Kategori yang dilakukan pada penelitian ini adalah hujan dan tidak hujan. Untuk kedepannya dapat dilakukan pengkategorian yang lebih luas sehingga dapat diketahui intensitas hujan yang turun.
2. Untuk selanjutnya dapat menggunakan dataset yang jauh lebih besar dari sekarang, misalnya data yang diolah sebanyak 10 tahun keatas.
3. Diharapkan untuk penelitian selanjutnya dapat menerapkan algoritma pembangkitan *frequent itemset* yang berbeda, seperti *FP-Growth*.

Pencarian Pola Asosiasi Keluhan Pasien Menggunakan Teknik Association Rule Mining

ABSTRAK

Perkembangan dan pertumbuhan data di bidang kesehatan semakin meningkat dan bertambah, baik dari kualitas maupun kuantitas, dilihat dari sisi kualitas, perkembangan data ini mengalami perubahan dari bentuk dokumen tulis menjadi dokumen digital atau yang biasanya kita sebut dengan file. Isu yang muncul adalah apakah informasi yang bisa diambil atau didapatkan dari sekian banyak data medis yang tersedia hanya berupa informasi – informasi pada umumnya, sedangkan dari suatu basis data yang tersedia seringkali memuat beberapa variabel sekaligus, bahkan apabila diteliti lebih jauh lagi, basis data yang berbeda bisa jadi memuat beberapa variabel yang sama, dari isu tersebut maka diperlukan suatu metode untuk bisa menggali lebih dalam informasi – informasi yang belum diketahui. Berkaitan dengan data medis serta data mining, maka penelitian kali ini akan membahas tentang implementasi atau kegunaan dari data mining pada data kunjungan pasien dengan cara menerapkan association rule mining untuk mendapatkan pola – pola asosiasi dari basis data kunjungan pasien yang tersedia menggunakan algoritma apriori dan algoritma *FP-Growth*. Baik algoritma apriori dan algoritma *FP-Growth* menghasilkan output yang sama. Perbedaan hasil uji coba terletak pada jumlah rule asosiasi yang ditemukan, dengan menggunakan algoritma apriori ditemukan 3 buah rule asosiasi, sedangkan ketika digunakan algoritma *FP-Growth* ditemukan 2 buah rule asosiasi, hal ini terjadi pada saat uji coba yang dilakukan menggunakan confidence sebesar 80%.

1. Latar Belakang

Perkembangan dan pertumbuhan data di bidang kesehatan semakin meningkat dan bertambah, baik dari kualitas maupun kuantitas, dilihat dari sisi kualitas, perkembangan data ini mengalami perubahan dari bentuk dokumen tulis menjadi dokumen digital atau yang biasanya kita sebut dengan *file*, baik file dokumen, gambar, maupun video, sedangkan apabila dilihat dari sisi kuantitas pertumbuhan data di bidang kesehatan mengalami pertambahan jumlah dari waktu ke waktu [2][4]. Sebagai contoh, data yang paling sering kita temui adalah data kunjungan pasien, data rekam medis pasien, data transaksi obat, data penyakit, sampai dengan data citra medis (hasil *USG*, hasil *CT-Scan*, hasil *Rontgen*), pada umumnya tumpukan data ini disimpan dan digunakan oleh pihak terkait untuk berbagai keperluan sesuai kebutuhan, data rekam medis biasanya digunakan sebagai sarana kontrol terhadap pasien, data transaksi obat untuk sumber informasi persediaan obat serta distribusinya, data penyakit untuk mengetahui jenis – jenis penyakit serta cara penanggulangannya. Isu yang muncul adalah apakah informasi yang bisa diambil atau didapatkan dari sekian banyak data medis yang tersedia hanya berupa informasi – informasi pada umumnya, sedangkan dari suatu basis data yang tersedia sering kali memuat beberapa variabel sekaligus, bahkan apabila lebih teliti lagi, basis data yang berbeda bisa jadi memuat beberapa variabel yang sama, dari isu tersebut maka diperlukan suatu metode untuk bisa menggali lebih dalam informasi – informasi yang belum diketahui [1][5][7]. *Data Mining* merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine*

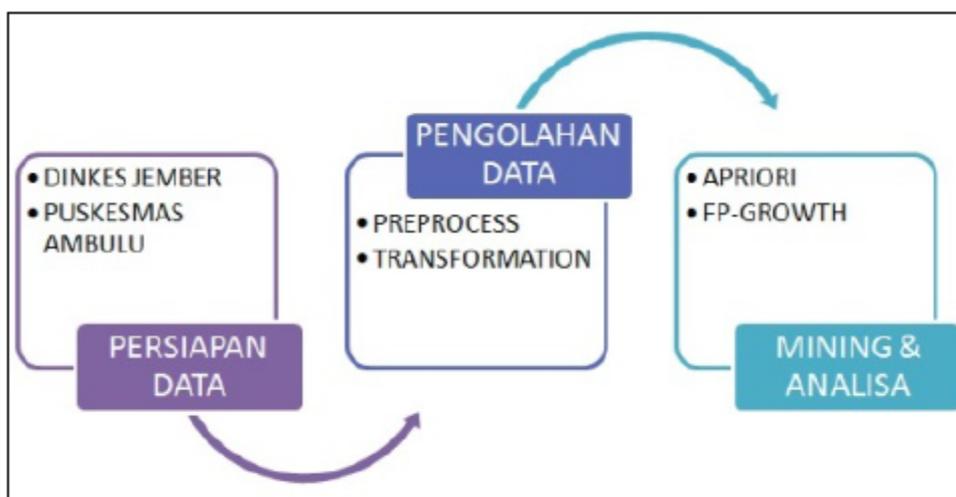
learning, menjadikan *data mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna [4]. Fungsi *data mining* yang sering digunakan adalah untuk klasifikasi, klasterisasi, estimasi, prediksi, serta penemuan pola asosiasi. *Association rule mining* merupakan salah satu teknik dalam *data mining* yang berguna untuk menemukan pola asosiasi tersembunyi dalam suatu basis data, pola yang dimaksud disini adalah keterkaitan atau korelasi antara tiap tiap *item* yang berbeda pada setiap *record* di dalam basis data. Pola asosiasi yang ditemukan nantinya berupa *rule – rule* dengan masing – masing nilai bobot asosiasinya, *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara *itemset* X dan Y[6][8].

Berkaitan dengan data medis serta *data mining*, maka penelitian kali ini akan membahas tentang implementasi atau kegunaan dari *data mining* pada data kunjungan pasien dengan cara menerapkan *Association rule mining* untuk mendapatkan pola – pola asosiasi dari basis data kunjungan pasien yang tersedia menggunakan algoritma apriori dan algoritma *FP Growth* serta *software* bantu WEKA [5], kedua algoritma ini sengaja dipakai untuk menunjukkan sejauh mana kemampuan masing – masing algoritma di dalam penemuan pola – pola asosiasi, sampai saat ini apriori merupakan algoritma yang paling sering dipakai karena kemudahan implementasinya, sedangkan algoritma *FP-Growth* merupakan revisi dari apriori yang mana dilakukan perbaikan ketika proses *scanning dataset* dilakukan[3].

2. Metode Penelitian

Untuk dapat menemukan pola asosiasi keluhan pasien maka tentunya harus didapatkan dulu data yang akan diproses, setelah pengumpulan data telah selesai maka akan dilakukan *preprocess* data kemudian dilanjutkan dengan *data transformation*, langkah selanjutnya yaitu melakukan *association rule Mining* sehingga pola – pola asosiasi keluhan pasien bisa didapatkan. Deskripsi alur penelitian bisa dilihat pada gambar

1, yang diikuti dengan penjelasan mengenai tahapan – tahapan penelitian. Pada penelitian ini dibagi menjadi 3 tahap utama yaitu : Tahap persiapan data, tahap pengolahan data, serta tahap mining dan analisa hasil keluaran.



Gambar 1. Alur Penelitian

Tahap Persiapan Data

Data yang digunakan pada penelitian ini adalah data kunjungan pasien puskesmas Ambulu pada rentang waktu Agustus 2011 sampai Januari 2012 dan diambil dari Dinas Kesehatan Kabupaten Jember. Data yang diambil termasuk data sekunder karena data ini merupakan hasil olahan Dinas Kesehatan yang berasal dari laporan bulanan tiap – tiap puskesmas di Kabupaten Jember. Data yang diperoleh berbentuk file .xls (format Microsoft Excel), jumlah record sebanyak 16.169 buah, yang mempunyai 3 atribut yaitu nomor pasien, keluhan, serta kode penyakit. Data yang diperoleh ini tidak bisa langsung digunakan dalam proses *Association rule* mining, sebelumnya data harus diperiksa apakah tidak ada data yang berulang, apakah terdapat duplikasi data, atau bahkan apakah ada data yang tidak lengkap, selain itu diperiksa juga apakah dari ketiga atribut tersebut dipakai semuanya atau tidak. Proses tersebut akan dilakukan pada tahap selanjutnya yaitu tahap pengolahan data, dimana nantinya akan ada 2 sub proses, *preprocess*, dan *data transformation*.

Tahap Pengolahan Data

Data yang sudah diperoleh kemudian akan dipersiapkan untuk proses *association rule mining* melalui *preprocess* dan *transformation* [4][6][8], kedua proses ini menggunakan alat bantu yaitu *software* Microsoft Excel 2007 (*evaluation copy*). *Preprocess*, bertujuan untuk membersihkan dan memilih data, yang dimaksud dengan membersihkan data adalah memeriksa apakah ada duplikasi data, kesalahan pengetikan, serta data yang tidak lengkap, sedangkan yang dimaksud dengan pemilihan data adalah memilih variabel data yang diperlukan saja. Setelah melalui *preprocess* maka hanya tersedia 670 data saja yang siap untuk proses selanjutnya, hal ini disebabkan karena pada variabel “keluhan” banyak berisi keluhan tunggal, keluhan berupa rujukan, dan keterangan yang kosong. *Transformation*, bertujuan untuk merubah format data yang sudah ada menjadi format yang bisa diproses, format yang dipakai adalah format biner, dimana pada satu transaksi berisi angka 0 (nol) atau 1 (satu), dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”.

Tahap Mining dan Analisis

Pencarian pola asosiasi diantara keluhan pasien akan dikerjakan dengan dua algoritma, *apriori* dan *FP-Growth*, dengan batasan nilai *support* dan *confidence* yang diinputkan oleh *user*, setelah pola asosiasi didapatkan maka tahap yang terakhir adalah melakukan analisa terhadap *rule – rule* yang terbentuk, sehingga diharapkan akan ditemukan informasi baru yang menarik dan berguna. Berdasarkan dataset yang diberikan *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi. Disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara *itemset* X dan Y.

3. Hasil dan Analisis

Pada bab ini akan disampaikan tentang hasil penelitian serta analisis dari hasil penelitian tersebut. Bab ini dibagi menjadi 2 sub bab, yang menjelaskan tentang dua tahapan penelitian yaitu tahap pengolahan data dan tahap mining dan analisis.

3.1. Tahap Pengolahan Data

Tahap pengolahan data ini terbagi menjadi 2 sub proses yaitu *preprocess* dan *transformation*. Tahap *preprocess*, bertujuan untuk membersihkan dan memilih data, yang dimaksud dengan membersihkan data adalah memeriksa apakah ada duplikasi data, kesalahan pengetikan, serta data yang tidak lengkap, sedangkan yang dimaksud dengan pemilihan data adalah memilih variabel data yang diperlukan saja. Tahap *transformation* merubah format data yang sudah ada menjadi format yang bisa diproses, format yang dipakai adalah format biner, dimana pada satu transaksi berisi angka 0 (nol) atau 1 (satu), dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”.

PUSING	PANAS	FLU	BATUK	SESAK	PILEK	MUAL
1	0	0	1	0	0	0
0	1	0	0	0	1	0
1	0	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
1	0	1	0	0	0	0
0	1	0	1	0	0	0
0	1	1	0	0	0	0
0	1	0	1	0	0	0
0	1	1	0	0	0	0
0	0	1	1	0	0	0
0	1	1	0	0	0	0
1	0	1	0	0	0	0
1	1	0	0	0	0	0
0	0	1	0	1	0	0

Gambar 2. Alur Penelitian

Pada gambar 2 merupakan potongan hasil dari proses pengolahan data, dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”, jadi dari gambar 2 dapat diartikan bahwa pada tiap baris mewakili pasien, baris pertama, si pasien mengalami pusing dan batuk, sedangkan pada baris kedua pasien mengalami panas dan pilek, begitu juga seterusnya.

3.2. Tahap Mining dan Analisis

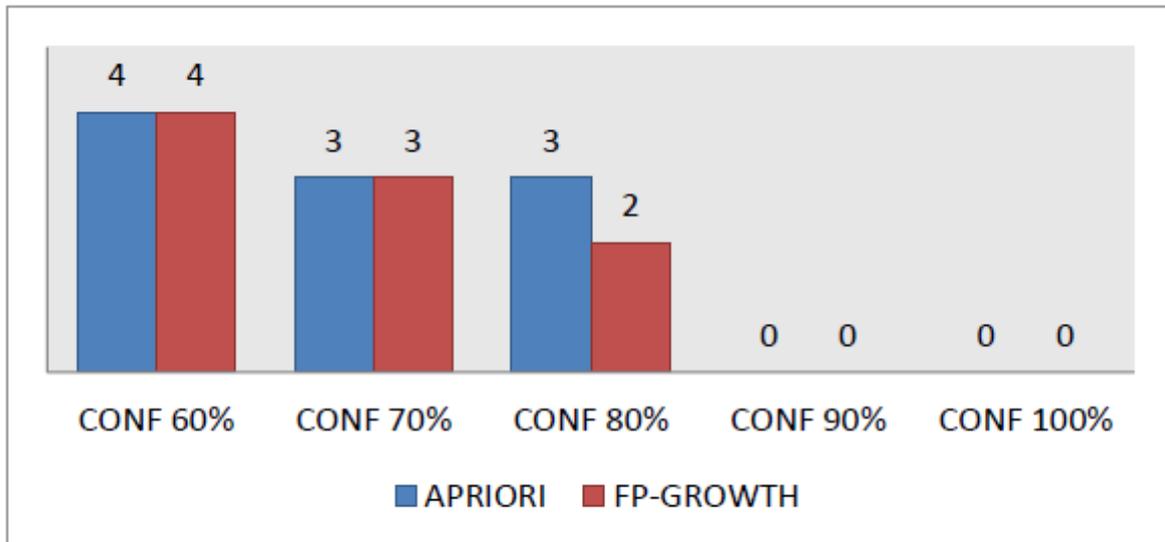
Skenario uji coba memakai 670 record data, dengan 5 nilai *confidence* yang berbeda – beda, yaitu 100%, 90%, 80%, 70%, dan 60%, dengan nilai support yang sama yaitu 1%. Alasan mengapa uji coba ini hanya dilakukan dengan nilai *confidence* yang berbeda – beda adalah karena biasanya pola asosiasi yang menarik ada pada kecenderungan pasangan itemset yang sering terjadi (definisi dari *confidence*). Uji coba nantinya akan dilakukan bergantian, yaitu uji coba terlebih dahulu menggunakan algoritma apriori kemudian menggunakan algoritma FP-Growth. Tabel menunjukkan hasil uji dengan variasi *confidence* yang berbeda – beda,

secara konsep, rule – rule yang dihasilkan juga ditampilkan secara berbeda oleh kedua algoritma.

Tabel. 1 Hasil Uji Coba

Uji ke-	Confidence	Hasil Apriori	Hasil FP-Growth
1	60%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=t 44 ==> PANAS=t 37 conf:(0.84) 2. FLU=t 344 ==> PANAS=t 288 conf:(0.84) 3. SESAK=t 45 ==> BATUK=t 36 conf:(0.8) 4. BATUK=t 259 ==> PANAS=t 168 conf:(0.65)	FPGrowth found 4 rules (displaying top 4) 1. [PILEK=t]: 44 ==> [PANAS=t]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=t]: 344 ==> [PANAS=t]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32) 3. [SESAK=t]: 45 ==> [BATUK=t]: 36 <conf:(0.8)> lift:(2.07) lev:(0.03) conv:(2.76) 4. [BATUK=t]: 259 ==> [PANAS=t]: 168 <conf:(0.65)> lift:(0.83) lev:(-0.05) conv:(0.62)
2	70%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=t 44 ==> PANAS=t 37 conf:(0.84) 2. FLU=t 344 ==> PANAS=t 288 conf:(0.84) 3. SESAK=t 45 ==> BATUK=t 36 conf:(0.8)	FPGrowth found 3 rules (displaying top 3) 1. [PILEK=t]: 44 ==> [PANAS=t]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=t]: 344 ==> [PANAS=t]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32) 3. [SESAK=t]: 45 ==> [BATUK=t]: 36 <conf:(0.8)> lift:(2.07) lev:(0.03) conv:(2.76)
3	80%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=t 44 ==> PANAS=t 37 conf:(0.84) 2. FLU=t 344 ==> PANAS=t 288 conf:(0.84) 3. SESAK=t 45 ==> BATUK=t 36 conf:(0.8)	FPGrowth found 2 rules (displaying top 2) 1. [PILEK=t]: 44 ==> [PANAS=t]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=t]: 344 ==> [PANAS=t]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32)
4	90%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found:	No rules found!
5	100%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found:	No rules found!

Pada gambar 3 disajikan grafik perbandingan secara persentase tentang jumlah rule yang dihasilkan berdasarkan skenario uji coba. Perbedaan hasil uji coba terletak pada jumlah *rule* asosiasi yang ditemukan, hal ini terjadi pada saat uji coba yang dilakukan menggunakan *confidence* sebesar 80%, dengan menggunakan algoritma apriori ditemukan 3 buah *rule* asosiasi, sedangkan ketika digunakan algoritma *FP-Growth* ditemukan 2 buah *rule* asosiasi.



Gambar 3 Perbandingan Uji Coba

4. Kesimpulan

Algoritma apriori dan algoritma FP-Growth dapat digunakan untuk mencari *rule* asosiasi dalam konteks *association rule mining*, algoritma apriori menggunakan prinsip apriori dalam pencarian *frequent itemset*, yaitu semua subset yang tidak kosong dari sebuah *frequent itemset* pasti juga akan merupakan *frequent*, begitu juga dengan semua superset yang tidak kosong dari sebuah *non-frequent itemset* pasti juga akan merupakan *non-frequent*. Sedangkan algoritma FP-Growth menggunakan konsep pembangunan *tree* dalam pencarian *frequent itemsets* yang memungkinkan dapat secara langsung membentuk *frequent itemset* dengan menerapkan prinsip *divide and conquer*.

Baik algoritma apriori dan algoritma FP_Growth menghasilkan output yang sama. Perbedaan hasil uji coba terletak pada jumlah *rule* asosiasi yang ditemukan, dengan menggunakan algoritma apriori ditemukan 3 buah *rule* asosiasi, sedangkan ketika digunakan algoritma FP-Growth ditemukan 2 buah *rule* asosiasi, hal ini terjadi pada saat uji coba yang dilakukan menggunakan *confidence* sebesar 80%.

Penelitian ini bisa dikembangkan dengan menggunakan kombinasi algoritma yang lain dalam bidang *association rule mining*, selain itu bisa juga dilakukan analisa tentang perbandingan waktu proses ketika melakukan pencarian *rule* asosiasi.

Daftar Pustaka

- [1] Danapana, H., Roy, M. S., Effective Data Mining Association Rules for Heart Disease Prediction System, IJCST Vol. 2, October – December, 2011.
- [2] Erwin, Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth, Jurnal Generic Vol. 4 No. 2, Juli 2009.
- [3] Han, J., et al, Mining Frequent Pattern Without Candidate Generation A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, 8, 53–87, 2004.

- [4] Han, J., Kamber, M., Pei, J., Data Mining Concepts and Techniques Third Edition, Morgan Kaufmann Publisher, 2012
- [5] Ordonez, C., Santana, C. A., de Braal, L., Discovering Interesting Association Rules in Medical Data, Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 2000, Hal. 78 – 85.
- [6] Sumathi, S., Sivanandam, S. N., Introduction to Data Mining and its Applications, Springer, 2006.
- [7] Srinivas, K., Rao, G. R., Govardhan, A., Mining Association Rules from Large Datasets Towards Disease Prediction, International Conference on Information and Computer Networks, 2012.
- [8] Witten, I. H., Frank, E., Data Mining Practical Machine Learning Tools and Technique, Morgan Kaufmann Publishers, 2005

Kesimpulan Dan Pendapat Saya

Algoritma apriori dan algoritma *FP-Growth* dapat digunakan untuk mencari *rule* asosiasi dalam konteks *association rule mining*, algoritma apriori menggunakan prinsip apriori dalam pencarian *frequent itemset*, yaitu semua subset yang tidak kosong dari sebuah *frequent itemset* pasti juga akan merupakan *frequent*, begitu juga dengan semua superset yang tidak kosong dari sebuah *non-frequent itemset* pasti juga akan merupakan *non-frequent*. Sedangkan algoritma *FP-Growth* menggunakan konsep pembangunan *tree* dalam pencarian *frequent itemsets* yang memungkinkan dapat secara langsung membentuk *frequent itemset* dengan menerapkan prinsip *divide and conquer*.

Baik algoritma apriori dan algoritma *FP_Growth* menghasilkan output yang sama. Perbedaan hasil uji coba terletak pada jumlah *rule* asosiasi yang ditemukan, dengan menggunakan algoritma apriori ditemukan 3 buah *rule* asosiasi, sedangkan ketika digunakan algoritma *FP-Growth* ditemukan 2 buah *rule* asosiasi, hal ini terjadi pada saat uji coba yang dilakukan menggunakan *confidence* sebesar 80%.

Penelitian ini bisa dikembangkan dengan menggunakan kombinasi algoritma yang lain dalam bidang *association rule mining*, selain itu bisa juga dilakukan analisa tentang perbandingan waktu proses ketika melakukan pencarian *rule* asosiasi.

APLIKASI ASSOCIATION RULE MINING UNTUK MENEMUKAN POLA PADA DATA NILAI MAHASISWA MATEMATIKA ITS

Latar Belakang



Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalah dari Tugas Akhir ini adalah bagaimana mengimplementasikan algoritma *Association Rule* pada data nilai mahasiswa Matematika ITS.

Batasan Masalah

- Data yang digunakan adalah data sekunder
- Data yang digunakan adalah data mahasiswa S1 Matematika ITS yang masih aktif dan data alumni S1 Matematika ITS 5 tahun terakhir

- Algoritma yang digunakan adalah algoritma Apriori
- Software yang digunakan untuk merancang program adalah NetBeans 7.1.2

Tujuan

Tujuan dari tugas akhir ini adalah membuat suatu perangkat lunak yang memuat *Association Rule Mining* sehingga dapat digunakan untuk menemukan pola dari *database* nilai mahasiswa S1 Matematika ITS.

Manfaat

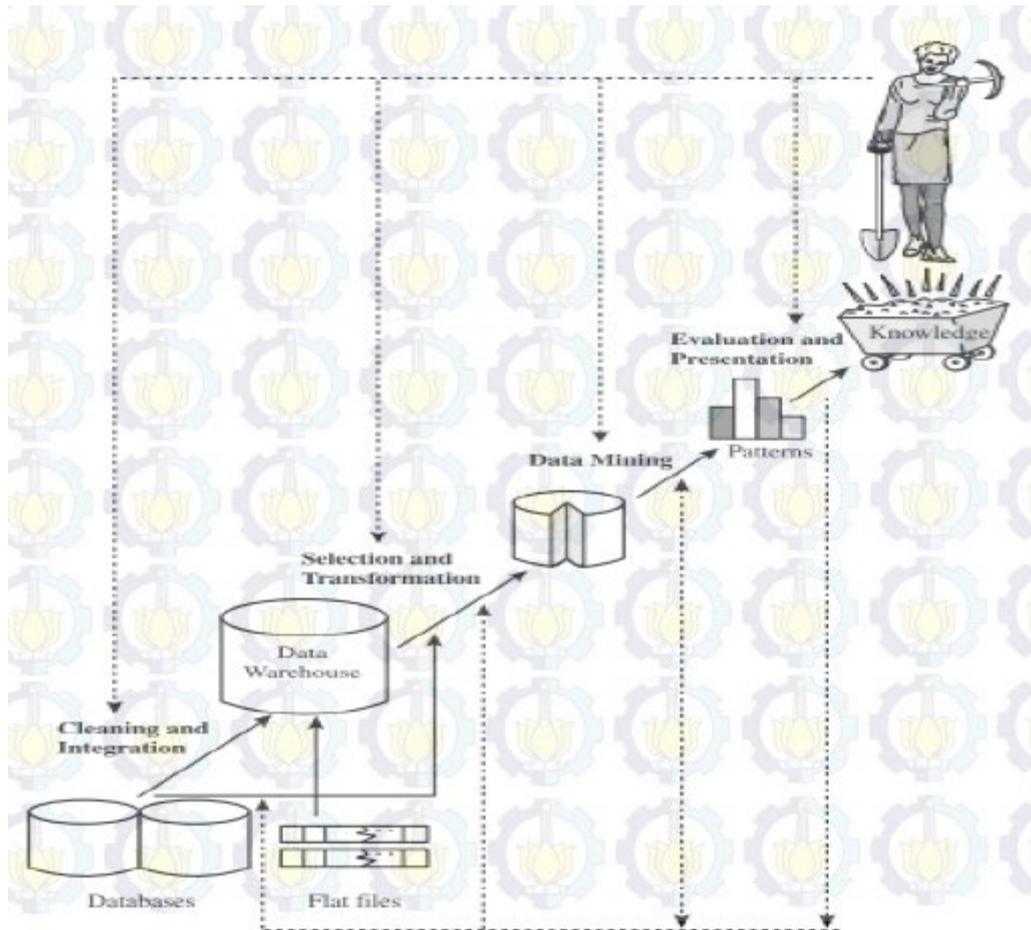
- Mendapatkan pola asosiasi atau hubungan antara satu nilai mata kuliah dengan nilai mata kuliah yang lain. Pola yang ditemukan dapat digunakan sebagai rekomendasi pengambilan keputusan
- Sebagai salah satu referensi penggunaan *Association Rule Mining*

Tinjauan Pustaka

- Data Mining
- Association Rule Mining
- Algoritma Apriori

Data Mining

Data mining merupakan Proses ekstraksi pola-pola yang menarik dari suatu data berukuran besar



Task Dalam Data Mining

- Asosiasi
- Klustering
- Klasifikasi
- Sequence pattern

Association Rule Mining

Asosiasi rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item.

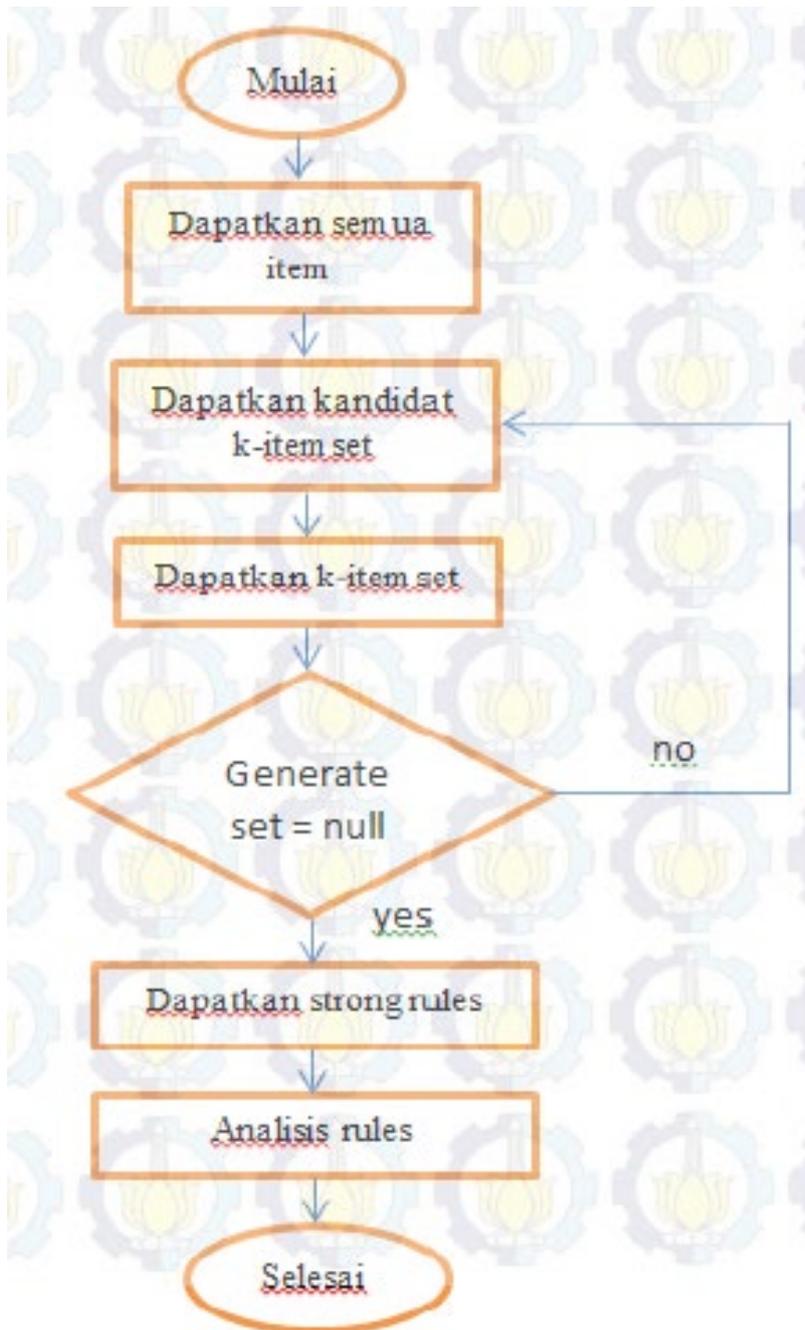
Dua tahapan dalam Association Rule Mining:

1. Temukan semua frequent itemsets
2. Temukan semua aturan asosisasi yang kuat

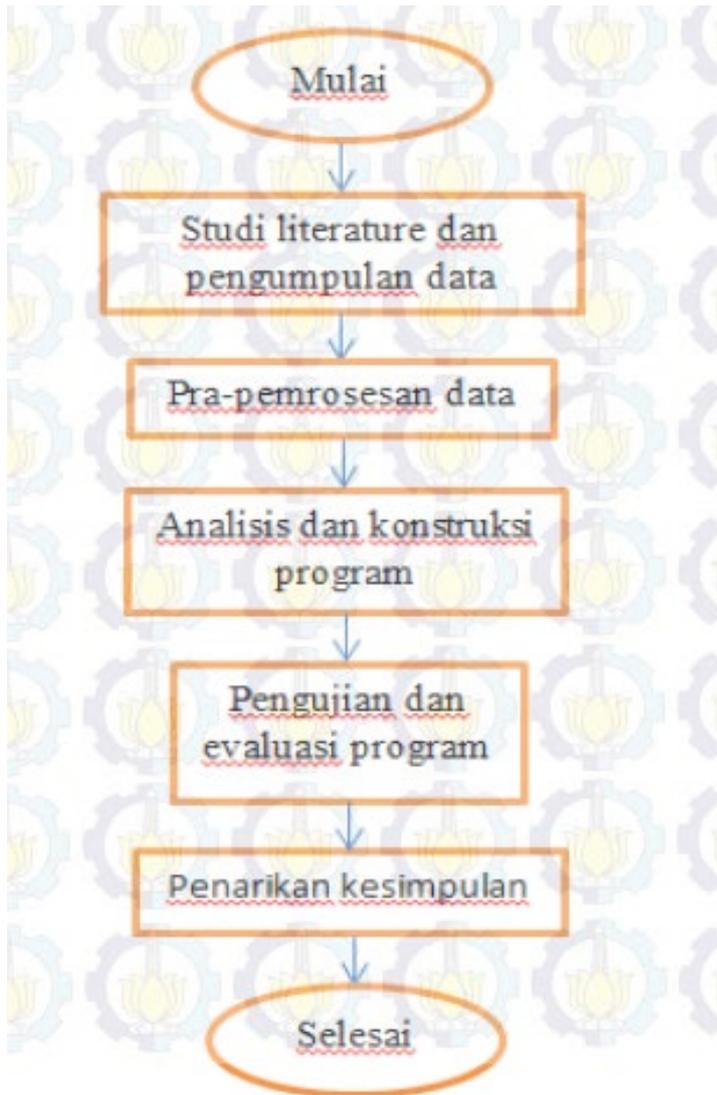
$$\text{Support } (A, B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}}$$

$$\text{Confidence } (A \rightarrow B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A}$$

Algoritma Apriori



Metodologi Penelitian



Analisa dan Perancangan Sistem

Secara umum ada 2 tahap dalam pembuatan program ini, yaitu:

1. Pengolahan data
2. Pemrosesan data

Pengolahan Data

Market basket analysis

Transaksi	Item yang dibeli
1	Kopi, Gula, Susu
2	Gula, Minyak
3	Gula, Roti
4	Kopi, Gula, Minyak
5	Kopi, Roti
6	Gula, Roti
7	Kopi, Roti
8	Kopi, Gula, Roti, Susu
9	Kopi, Gula, Roti

Transaksi	Item
I204100001	Aljabar I_AB, Aljabar Linear_AB, ...
I204100002	Aljabar I_A, Aljabar Linear_AB, ...
I204100003	Aljabar I_BC, Aljabar Linear_BC, ...
I204100004	Aljabar I_A, Aljabar Linear_A, ...
I204100005	Aljabar I_B, Aljabar Linear_A, ...

Pengolahan Data: Activity Diagram



Pengolahan Data 2

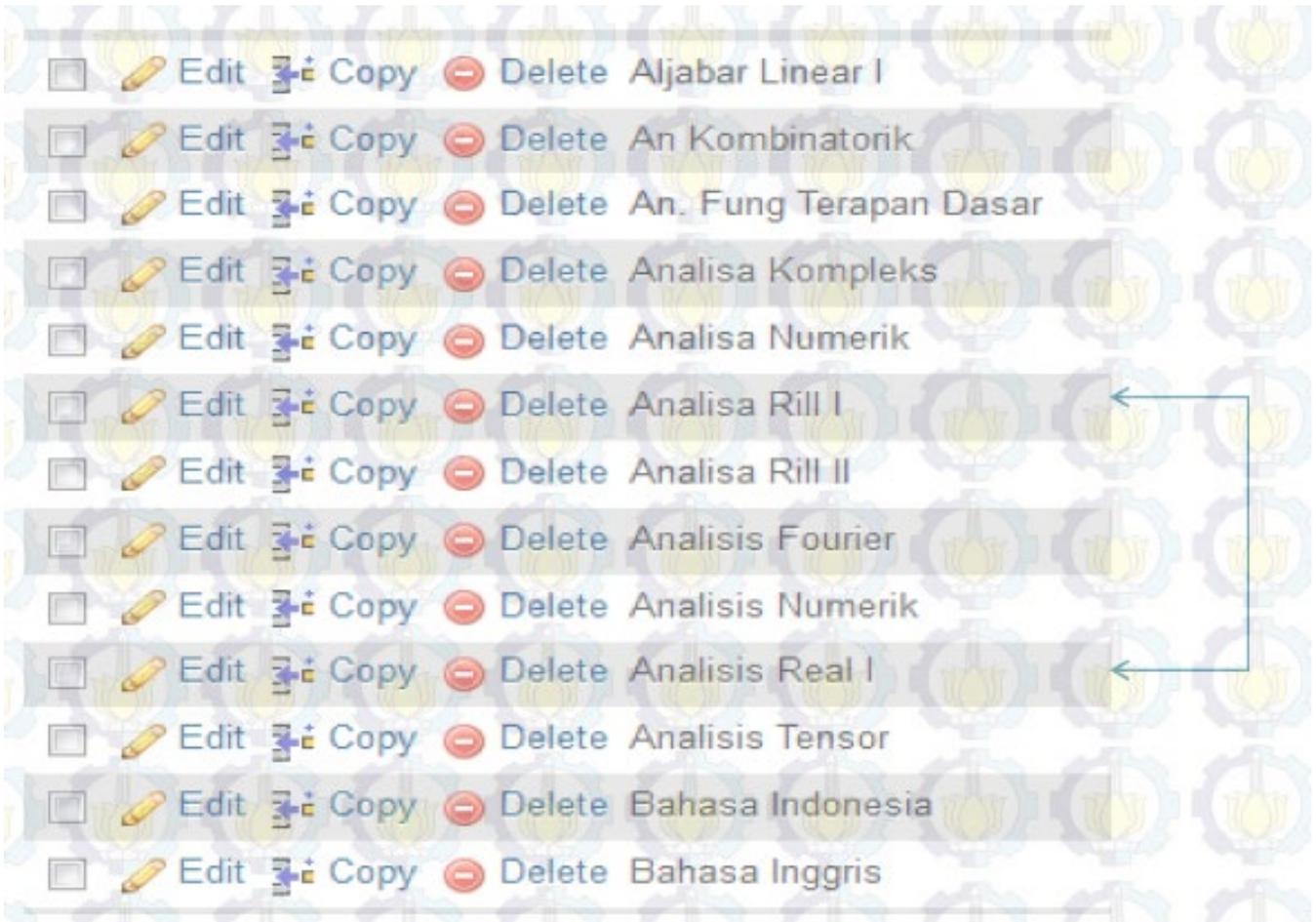
	A	B	C	D	E	F
1	tahun	semester	nrp	kurikulum	kodeMK	nilaiHuruf
2	2005	1	1205100068	2004	SM1303	AB
3	2005	1	1205100068	2004	SM1305	AB
4	2005	1	1205100068	2004	SK1205	BC
5	2005	1	1205100068	2004	UG1201	C
6	2005	1	1205100068	2004	UM1201	D
7	2005	1	1205100068	2004	UF1201	E
8	2005	2	1205100068	2004	UG1301	AB
9	2005	2	1205100068	2004	SM1308	BC
10	2005	2	1205100068	2004	UG1202	BC
11	2005	2	1205100068	2004	UM1202	BC
12	2005	2	1205100068	2004	SM1306	C
13	2005	2	1205100068	2004	SM1304	D
14	2005	2	1205100068	2004	UF1202	E
15	2006	1	1205100068	2004	SM1313	AB
16	2006	1	1205100068	2004	UM1201	B

	A	B	C	D
1	kurikulum	kodeMK	namamk	kredit
2	2004	RI2130	STRATEGI MANUFAKTUR	2
3	2004	RI2131	Perencanaan Fasilitas	2
4	2004	RI2133	Computer System in Manufacturing	2
5	2004	RI2132	Multivariat Statistik	2
6	2004	RM2246	Analisa Pelat dan Cangkang	2
7	2004	RM2405	Kinematika dan Dinamika Lanjut	3
8	2004	RM2417	Desain Produk	3
9	2004	RM2249	Mekanika Patahan	2
10	2004	RM2415	Kontrol Adaptif	3
11	2004	RM2247	Condition Manitoring	2
12	2004	RM2418	Mekanika Material Lanjut	3
13	2004	RM2250	Robotika	2
14	2004	RM2252	Analisa Tegangan Eksperimen	2
15	2004	RM2248	Reliability Centered Maintenance	2
16	2004	RM2414	Tribologi	2
17	2004	RM2251	Desain dan Stabilitas Kendaraan	2
18	2004	RM2416	Optimasi Desain	2

Pengolahan Data 3

1205100068	2004	SM1318	D
1205100068	2004	UF1201	D
1205100068	2004	SM1311	-
1205100068	2004	SM1315	-
1205100068	2004	SM1245	AB
1205100068	2004	SM1241	C
1205100068	2004	SM1309	D

Pengolahan Data 4



Pengolahan Data 5

					Nilai Huruf	Indeks
					A	4
					AB	3,5
					B	3
					BC	2,5
					C	2
					D	1
					E	0

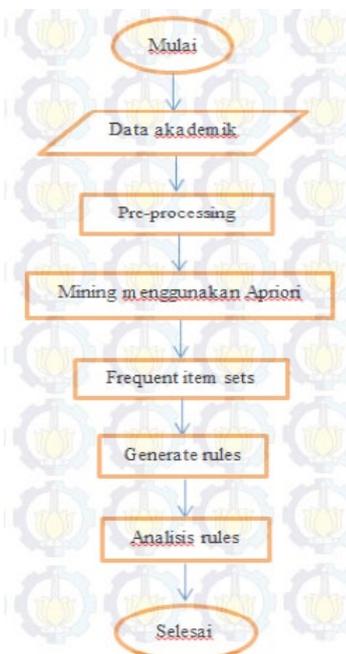
1205100068	2004	SK1205	Kimia Dasar	2,5	BC
1205100068	2004	UG1201	Bahasa Inggris	2	C
1205100068	2004	UM1201	Kalkulus I	1	D
1205100068	2004	UF1201	Fisika Dasar I	0	E
1205100068	2004	UG1301	Agama Islam	3,5	AB
1205100068	2004	SM1308	Geometri Analitik	2,5	BC
1205100068	2004	UG1202	Pengantar Ilmu Lingkungan	2,5	BC
1205100068	2004	UM1202	Kalkulus II	2,5	BC
1205100068	2004	SM1306	Aljabar Linear Elementer	2	C
1205100068	2004	SM1304	Aljabar Linear Elementer	1	D
1205100068	2004	UF1202	Fisika Dasar II	0	E
1205100068	2004	SM1313	Matematika Diskret	3,5	AB
1205100068	2004	UM1201	Kalkulus I	3	B
1205100068	2004	UG1307	Bahasa Indonesia	2,5	BC

Two blue arrows point from the 'Indeks' column of the second table to the 'Nilai Huruf' column of the first table, specifically pointing to the 'D' and 'B' rows.

Pengolahan Data 6

nrp	kurikulum	kodemk	namaMK	nilaiAngka	nilaiHuruf	dataOlah
1205100068	2004	UG1301	Agama Islam	3.5	AB	Agama Islam_AB
1205100068	2009	SM1256	Algoritma Kombinatorik	2	C	Algoritma Kombinatorik_C
1205100068	2004	SM1318	Aljabar 1	2	C	Aljabar 1_C
1205100068	2004	SM1323	Aljabar Linear	2.5	BC	Aljabar Linear_BC
1205100068	2004	SM1304	Aljabar Linear Elementer	3	B	Aljabar Linear Elementer_B
1205100068	2009	SM1322	An Kombinatorik	3	B	An Kombinatorik_B
1205100068	2009	SM1315	Analisis Numerik	2.5	BC	Analisis Numerik_BC
1205100068	2009	SM1321	Analisis Real I	3	B	Analisis Real I_B
1205100068	2004	UG1307	Bahasa Indonesia	2.5	BC	Bahasa Indonesia_BC
1205100068	2004	UG1201	Bahasa Inggris	2	C	Bahasa Inggris_C
1205100068	2004	UG1308	Bahasa Inggris Lanjut	4	A	Bahasa Inggris Lanjut_A
1205100068	2004	UF1201	Fisika Dasar I	1	D	Fisika Dasar I_D
1205100068	2004	UF1202	Fisika Dasar II	1	D	Fisika Dasar II_D
1205100068	2004	SM1312	Fungsi Peubah Kompleks	2.5	BC	Fungsi Peubah Kompleks_BC
1205100068	2004	SM1305	Geometri Analitik	3.5	AB	Geometri Analitik_AB
1205100068	2004	SM1242	Jaringan Komputer	3	B	Jaringan Komputer_B
1205100068	2004	SM1244	Jaringan Syaraf Tiruan	3.5	AB	Jaringan Syaraf Tiruan_AB
1205100068	2004	SM1303	Kalkulus I	3.5	AB	Kalkulus I_AB
1205100068	2004	UM1202	Kalkulus II	2.5	BC	Kalkulus II_BC
1205100068	2004	SM1309	Kalkulus Peubah Banyak	2	C	Kalkulus Peubah Banyak_C

Pemrosesan Data



Pemrosesan Data 2

Dalam perancangan ini yang dimaksud transaksi adalah seluruh mata kuliah yang diambil beserta nilai yang didapatkan untuk setiap mahasiswa (nrp).

Transaksi	Item
1204100001	Aljabar I_AB , Aljabar Linear_AB , ...
1204100002	Aljabar I_A , Aljabar Linear_AB , ...
1204100003	Aljabar I_BC , Aljabar Linear_BC , ...
1204100004	Aljabar I_A , Aljabar Linear_A , ...
1204100005	Aljabar I_B , Aljabar Linear_A , ...

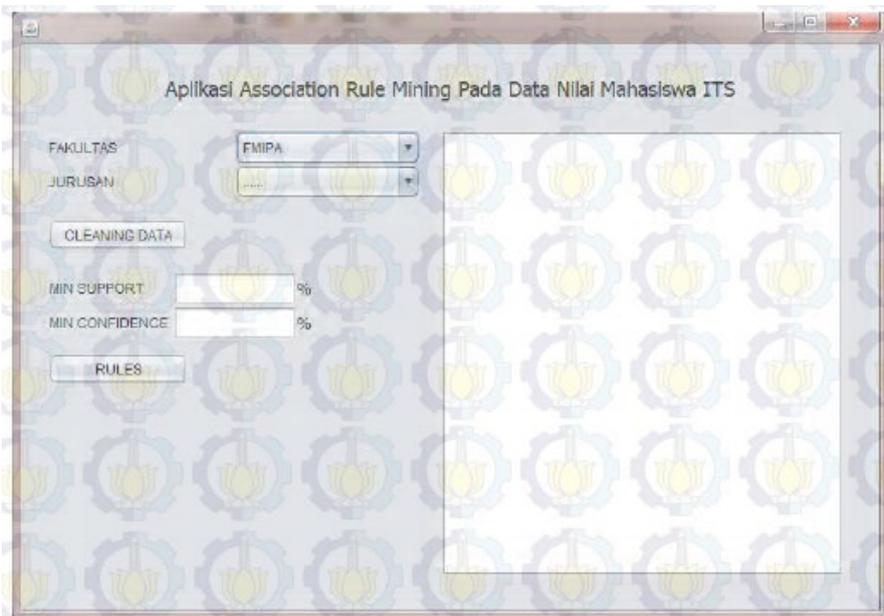
Pemrosesan Data 3

Transaksi-transaksi tersebut disimpan dalam linked list, dimana setiap node berisi NRP kemudian semua mata kuliah yang diambil beserta nilai yang didapatkan.

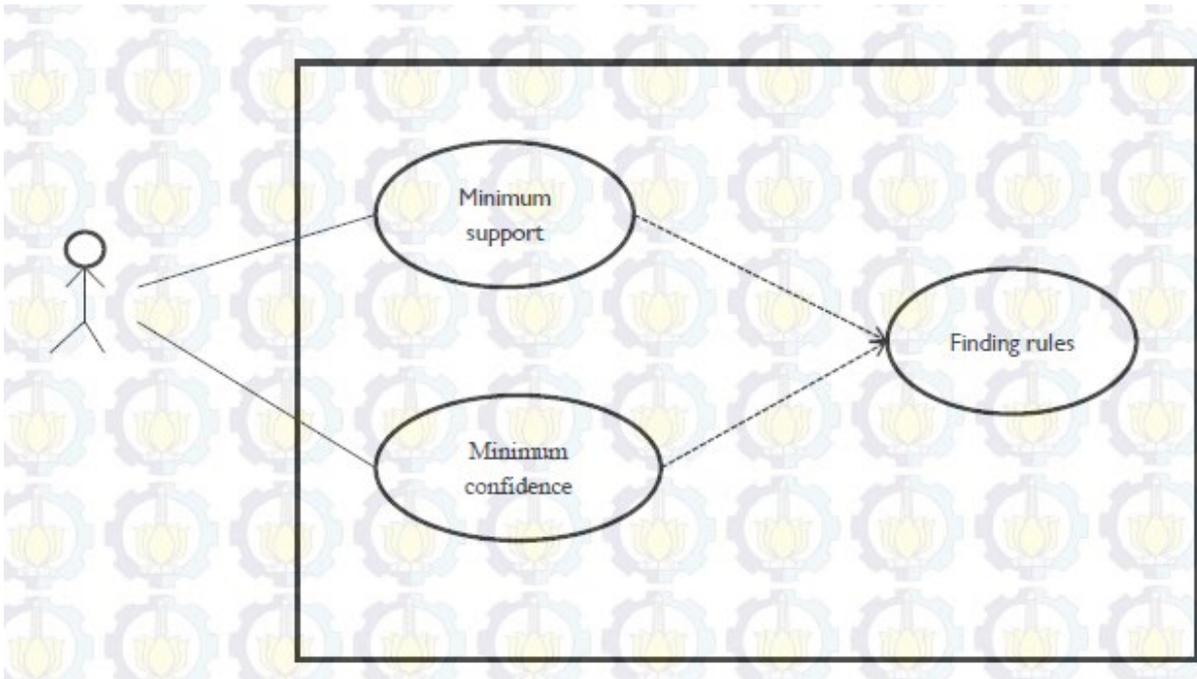


Pemrosesan Data 4

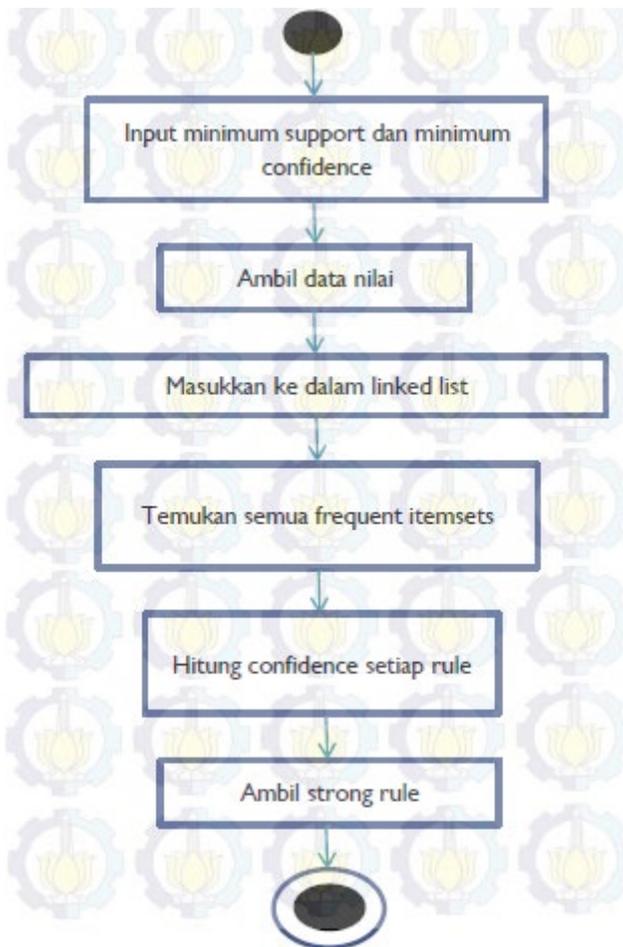
Berikutnya adalah menentukan nilai minimum support dan nilai minimum confidence



Pemrosesan Data 5



Pemrosesan Data 6



Uji Coba Dan Pembahasan

Dilakukan pengujian pada data nilai alumni matematika lima tahun terakhir. Banyak alumni matematika 5 tahun terakhir sebanyak 403 yang berarti terdapat 403 transaksi. Ada tiga nilai minimum support dan tiga nilai minimum confidence yang digunakan dalam uji coba ini.

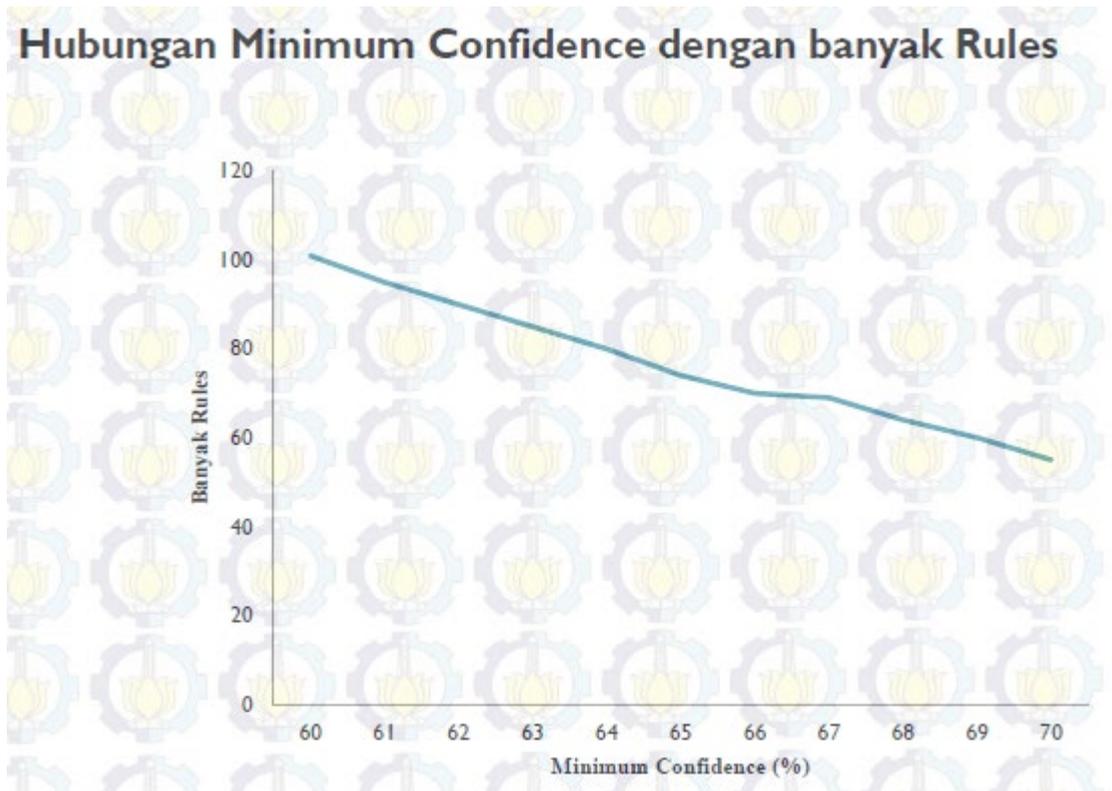
Uji Coba Dan Pembahasan 2



Uji Coba Dan Pembahasan 3



Uji Coba Dan Pembahasan 4



Uji Coba Dan Pembahasan 4

Untuk nilai minimum support 20% dan nilai minimum confidence 60% didapatkan 34 rules:

1. Aljabar Linear Elementer_A \implies Kalkulus I_A Confidence : 86.44%
2. Aljabar Linear Elementer_A \implies Tugas Akhir_A Confidence : 68.64%
3. Aljabar Linear Elementer_AB \implies Kalkulus I_A Confidence : 79.25%
4. Aljabar Linear_AB \implies Kalkulus I_A Confidence : 70.15%
5. An Kombinatorik_B \implies Kalkulus I_A Confidence : 75.21%
6. Analisis Numerik_A \implies Kalkulus I_A Confidence : 85.71%
7. Analisis Numerik_AB \implies Kalkulus I_A Confidence : 81.25%
8. Analisis Real I_AB \implies Kalkulus I_A Confidence : 64.56%
9. Fungsi Peubah Kompleks_AB \implies Kalkulus I_A Confidence : 81.34%
10. Fungsi Peubah Kompleks_AB \implies Pemodelan Matematika_AB Confidence : 61.19%
11. Geometri Analitik_AB \implies Kalkulus I_A Confidence : 86.32%
12. Kalkulus II_A \implies Kalkulus I_A Confidence : 83.72%
13. Kalkulus II_A \implies Tugas Akhir_A Confidence : 64.34%
14. Kimia Dasar_AB \implies Kalkulus I_A Confidence : 80.2%
15. Komputasi Dinamika Fluida_AB \implies Kalkulus I_A Confidence : 69.4%
16. Matematika Diskret_AB \implies Kalkulus I_A Confidence : 71.15%
17. Matematika Sistem_A \implies Kalkulus I_A Confidence : 85.26%

Uji Coba Dan Pembahasan 5

18. Matematika Sistem_AB \implies Kalkulus I_A Confidence : 61.83%
19. Metode Matematika_A \implies Kalkulus I_A Confidence : 82.61%

20. Metode Matematika_AB ==> Kalkulus I_A Confidence : 69.53%
21. Kalkulus I_A ==> Pemodelan Matematika_AB Confidence : 60.51%
22. Pemodelan Matematika_AB ==> Kalkulus I_A Confidence : 74.55%
23. Pengantar Analisis Fungsional_A ==> Kalkulus I_A Confidence : 87.25%
24. Pengantar Analisis Fungsional_AB ==> Kalkulus I_A Confidence : 72.19%
25. Pengantar Analisis Fungsional_B ==> Kalkulus I_A Confidence : 70.63%
26. Penulisan Ilmiah Matematika_AB ==> Kalkulus I_A Confidence : 84.76%
27. Str Data & An Alg_B ==> Kalkulus I_A Confidence : 72.81%
28. Tugas Akhir_A ==> Kalkulus I_A Confidence : 79.41%
29. Metode Matematika_AB ==> Pemodelan Matematika_AB Confidence : 63.28%
30. Pengantar Analisis Fungsional_AB ==> Pemodelan Matematika_AB Confidence : 62.72%
31. Kalkulus I_A Pengantar Analisis Fungsional_AB ==> Pemodelan Matematika_AB Confidence : 67.21%
32. Pemodelan Matematika_AB Pengantar Analisis Fungsional_AB ==> Kalkulus I_A Confidence : 77.36%
33. Kalkulus I_A Tugas Akhir_A ==> Pemodelan Matematika_AB Confidence : 61.11%
34. Pemodelan Matematika_AB Tugas Akhir_A ==> Kalkulus I_A Confidence : 88.39%

Uji Coba Dan Pembahasan 6

Banyak rule dengan konsekuen Kalkulus I_A tapi hanya sekali Kalkulus I_A menjadi anteseden, hal tersebut disebabkan nilai support Kalkulus I_A cukup besar .

Tidak semua rule dapat digunakan, rule yang dapat digunakan adalah rule no 2, 10, 13, 21, 29

Uji Coba Dan Pembahasan 7

2. Aljabar Linear Elementer_A ==> Tugas Akhir_A Confidence : 68.64%
10. Fungsi Peubah Kompleks_AB ==> Pemodelan Matematika_AB Confidence : 61.19%
13. Kalkulus II_A ==> Tugas Akhir_A Confidence : 64.34%
21. Kalkulus I_A ==> Pemodelan Matematika_AB Confidence : 60.51%
29. Metode Matematika_AB ==> Pemodelan Matematika_AB Confidence : 63.28%

Daftar Pustaka

- S. Baha, U. Emine, D. Dursun. 2012. *Predicting and Analyzing Secondary Education Placement Test Scores: A Data Mining Approach*. Expert System with Application, Volume 39, page 9468-9476.
- B. Ali, U. Kerem. 2010. *Data Mininig Application on Students Data*. Procedia Social and Behavioral Sciences, Volume 3, page 5251-5259.
- W. Cheng-Hsiung. 2011. *Mining Fuzzy Specific Rare Itemsets for Education Data*. Knowlwdge-Based Systems, Volume 24, page 697-708.
- Han. J, Kamber. M. 2006. *Data Mining Concepts and Techniques*. Morgan Kaufmann.
- P. Diah. 2012. *Sistem Rekomendasi Bidang Minat Mahasiswa Menggunakan Metode Association Rule dan Algoritma Apriori*. Digilib ITS
- Kusriani, Luthfi E. T. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit Andi

Kesimpulan dan pendapat saya

Berdasarkan hasil pembahasan dapat ditarik kesimpulan sebagai berikut:

1. Association Rule Mining telah berhasil diimplementasikan pada data nilai mahasiswa Matematika ITS
2. Nilai *minimum support* berpengaruh pada hasil dan lama pencarian rules/aturan asosiasi, begitu juga nilai *minimum confidence* berpengaruh pada hasil pencarian aturan asosiasi.
3. Banyak mata kuliah pada semester pertengahan dan akhir yang mendapat nilai A atau AB dengan latar belakang mendapatkan nilai A pada Kalkulus I, sehingga dari hasil pengujian dapat dikatakan Kalkulus I memiliki peran penting sebagai salah satu faktor keberhasilan mahasiswa matematika dalam mencapai nilai yang maksimal pada perkuliahan berikutnya

Tugas 2

Nama : Oman Arrohman
Mata Kuliah : Advanced Database
NIM : 202420042

Judul Jurnal	:	Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang Dijual di E-commerce OrderMas
Peneliti	:	Moh.Sholik, Abu Salam
Publikasi	:	<i>Techno.COM</i>
Tahun	:	2018
Pokok Penelitian/Pokok Bahasan	:	<p>Ketika sebuah toko memiliki stok barang yang lebih banyak daripada jumlah yang diminta konsumen, maka toko akan mengalami kerugian, karena stok barang tidak laku dan habis terjual, terutama barang-barang yang mengandung unsur kadaluwarsa dan mudah rusak. Sebaliknya apabila suatu toko/perusahaan memiliki stok barang yang minimum dibandingkan dengan jumlah permintaan maka berakibat pada toko/perusahaan yaitu konsumen tersebut akan pergi karena barang yang akan dibeli oleh konsumen tidak mencukupi (<i>opportunity loss</i>). Hal yang sering kita jumpai adalah saat konsumen ingin membeli suatu barang dan barang yang ingin dibeli ternyata habis, maka konsumen akan membeli ditempat lain dan untuk waktu yang akan datang konsumen akan lebih banyak mempertimbangkan jika ingin membeli ditoko itu kembali, karena takut barang yang diinginkan habis kembali, tentunya hal ini dapat mengurangi tingkat penjualan. Guna mengatasi masalah tersebut maka dibutuhkan suatu system cerdas data mining yang dapat membantu guna mengambil keputusan berhubungan jenis dan stok barang yang disediakan pada <i>Supply Chain</i> berikutnya menggunakan data mining [1].</p> <p>Data mining berguna untuk memberikan solusi kepada para pemilik perusahaan untuk mengambil keputusan guna meningkatkan bisnis perusahaan. Dengan menggunakan analisis <i>asosiasi</i> dan algoritma <i>apriori</i> dapat menghasilkan suatu rekomendasi yang di harapkan dapat membantu manager dalam mendukung keputusan strategi penjualan [2]. Jadi data maning adalah suatu proses tentang pengambilan data ataupun informasi dalam skala besar dan kemudian dapat diproses menjadi sebuah pengetahuan untuk mendapatkan informasi yang akurat tanpa melalui perhitungan manual [3]. Salah satu hal penting dalam teknik data mining adalah aturan untuk menemukan pola frekuensi tinggi antar himpunan <i>itemset</i> yang disebut dengan fungsi <i>Association Rules (aturan asosiasi)</i> [4][5][6]. Perkembangan teknologi pada data mining tersebut dapat digunakan untuk menganalisa data dan memprediksi jumlah barang [7].</p>

	<p>Dalam sebuah penelitian yang berjudul <i>Analisa Data Mining Dalam Penjualan Sparepart Mobil dengan Menggunakan Metode Algoritma Apriori</i>, proses asosiasi menggunakan algoritma apriori menghasilkan aturan asosiasi keterkaitan yang kuat antar itemset penjualan suku cadang sehingga bisa memberi rekomendasi penyetokan barang dan mempermudah dalam penataan atau penempatan barang yang kuat berkaitan saling ketergantungan [8], Denny Haryanto juga melakukan penelitian pada market basket analysis dengan melakukan pencarian pola relasi penjualan terlebih dahulu sebelum melakukan penawaran, maka penawaran barang yang berpeluang kecil untuk terbeli akan terabaikan untuk ditawarkan [2]. Dalam penelitian lain penerapan data mining untuk market basket analisis menggunakan algoritma Apriori dapat membantu menemukan pola penjualan obat berdasarkan kecenderungan obat yang terjual secara bersamaan, kemudian dari hasil pengujian ini juga dapat membantu pihak apotek dalam mengatur tata letak obat secara berdekatan untuk memudahkan dalam mengetahui keberadaan obat serta dapat membantu dalam mengambil keputusan untuk menentukan persediaan obat. Penelitian Association rule dalam penjualan barangbarang kesehatan juga pernah dilakukan untuk mengetahui pola penjualan yang digunakan sebagai dasar pengambilan keputusan dalam menyiapkan barang yang diprioritaskan dalam penjualan kedepan, Penelitian Association rule juga digunakan untuk mengetahui produk dengan penjualan terbanyak dan keterkaitan produk satu dengan yang lainnya berdasarkan kondisi transaksi yang dilakukan. .</p> <p>Penelitian ini bertujuan menganalisis data transaksi penjualan di e-commerce OrderMas untuk mencari keterkaitan pembelian antar item guna memecahkan masalah tentang pengadaan stok barang oleh supplier OrderMas yang tidak dapat di prediksi kapan supplier harus menyetok barang tersebut, serta berguna untuk penataan barang pada e-commerce OrderMas dengan kecenderungan pola pembelian konsumen .</p>
Metode Penelitian	<p>: Analisis implementasi algoritma apriori untuk mencari asosiasi barang yang dijual pada ecommerce OrderMas akan diselesaikan melalui beberapa tahapan sebagai berikut :</p> <ol style="list-style-type: none"> 1. Literature View <p>Penulis menggunakan jurnal ilmiah, <i>proceedings</i> dan <i>e-book</i> diantaranya adalah <i>“Decision Support on Supply Chain Management System Using Apriori Data Mining Algorithm”</i>[1], <i>“Applications Of Data Mining Techniques In Pharmaceuntical Industry”</i> [3] dan jurnal ilmiah ISSN yang berjudul <i>“Analisa Data Mining Dalam Penjualan Sparepart Mobil dengan Menggunakan Metode Algoritma Apriori”</i>[8], <i>“Data Mining Market Basket Analysis Menggunakan Algoritma Apriori Untuk menentukan Persediaan Obat”</i> [10], serta e-book yang berjudul <i>“Data</i></p>

Mining : Practical Machine Learning Tools and Techniques 3rd Edition" [6], *"Discovering Knowledge in Data, an Introduction to Data Mining"* [5], *"Data Mining: Concepts and Techniques Third Edition"* [4], dan juga beberapa jurnal internasional maupun nasional yang mengenai data mining serta ketersediaan dan keterkaitan antar barang menggunakan algoritma *apriori* lainnya.

2. Objek Penelitian

Pada penelitian ini objek yang akan dijadikan penelitian oleh penulis adalah project Web E-commerce Semarang, Jawa Tengah. Dengan mengambil data dari dataset di internet untuk penelitian, dikarenakan belum adanya transaksi pada web E-commerce tersebut sedangkan transaksi digunakan lebih kurangnya 1000 record data.

3. Instrumen Penelitian

Instrumen peralatan yang digunakan untuk penelitian ini meliputi perangkat keras (Hardware) dan perangkat lunak (software) sebagai berikut:

1) Perangkat Keras (Hardware) meliputi :

- a. Laptop Dell Inspiron N-4050 Prosesor Intel core i3-6006U
- b. CPU dengan kecepatan @2,0GHz
- c. RAM 6 GB
- d. Harddisk 500 GB
- e. Layar 14 inch

2) Perangkat Lunak (software) meliputi :

- a. Delphi 7, digunakan untuk mengolah data mentah menjadi data tabular agar data bisa diolah di aplikasi RapidMiner.
- b. Data Base FireBird, Sistem Manajemen Basis data Relasional untuk database Delphi 7
- c. Rapid Miner, Rapid Miner adalah salah satu software yang digunakan untuk membangun sistem ini sebagai pengolah data dan menampilkan hasil yang ada.
- d. Microsoft Excel, Untuk mempermudah mengumpulkan data dan menampilkan data hasil ekstraksi dari program delphi.

a. Metode Pengumpulan Data

Metode yang digunakan adalah metode Survei dimana mengacu kepada metode pengumpulan informasi mengenai fenomena yang diteliti dari seluruh populasi ataupun sampel yang mewakili melalui pengamatan. Penulis juga menggunakan metode wireless dan website yaitu pengumpulan data-data melalui internet yaitu set data transnasional yang berisi semua transaksi yang terjadi antara 01/12/2010 dan 09/12/2011 untuk bebas-toko ritel online berbasis di Inggris dan terdaftar.

b. Metode Analisis

Data yang telah didapat dan dikumpulkan selama proses pencarian dan pengumpulan data kemudian akan masuk ke tahap selanjutnya yaitu tahap analisis data. Metode analisis data ini yaitu dengan proses membedakan data sesuai dengan jenisnya, adapun

macam dan jenis data antara lain data primer dan data sekunder. Disini data yang digunakan adalah data Sekunder Pengertian atau penjelasan data sekunder merupakan data yang didapat dari proses pencarian data online shop retail yang berbasis di Inggris melalui internet, data ini lah yang nantinya akan dijadikan sebagai acuan dalam proses pengembangan sistem data mining. Selanjutnya pengertian data sekunder merupakan data-data yang terkumpul melalui proses analisis literatur, buku, jural, atau internet yang semua memiliki hubungan dengan data mining. Semua data-data tersebut akan dianalisis agar dapat digunakan sesuai dengan metode perhitungan yang akan peneliti gunakan.

c. Metode Yang Diusulkan

Melakukan pelatihan pada set data menggunakan algoritma apriori dengan cara metode *association rule* untuk menentukan pengelompokan data, Teknik *association rule* yaitu cara dalam data mining untuk menemukan rule dalam kombinasi tertentu untuk satu item. Didalam teknik *association rule* ada 2 parameter antara lain adalah *confidence* dan *support*. *Support* yaitu hasil dari nilai yang ditunjang untuk dilihat dengan tingkatan presentasi dalam itemset tersebut dari keseluruhan transaksi yang dilakukan, sedangkan *confidence* merupakan suatu hasil kepastian yang dapat dilihat kuat tidaknya relasi dengan itemset tersebut merupakan *association rule*.

Algoritma *apriori* yaitu salah satu algoritma yang diajukan Agrawal dan Srikant tahun 1994 yang berfungsi untuk menentukan *frequent itemset* pada asosiasi *boolean* [4]. Algoritma ini bertugas untuk memantau perkembangan kandidat dari *itemset* tersebut [2]. *Pseudo-Code Apriori Algorithm*:

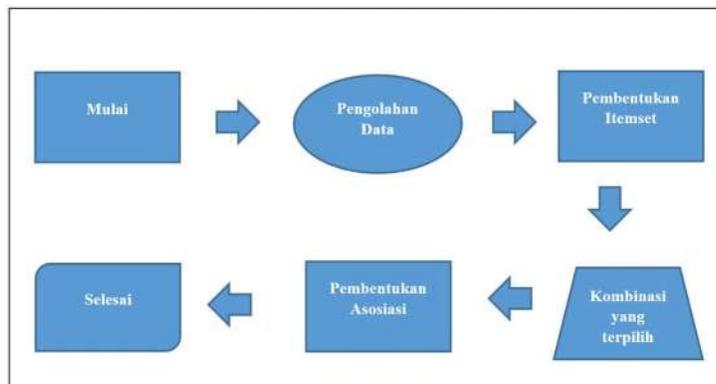
```
? Ck: Candidate itemset dari ukuran k;
? Lk : Frequent itemset dari ukuran k.
? L1 = {frequent items};
? for (k = 1; Lk !=0; k++) do begin
? Ck+1 = {kandidat dibangun dari
? for each transaksi t yang dimuat dalam database do naikkan
  hitungan dari seluruh kandidat dalam Ck+1 yang dimuat dalam
  t
? Lk+1 = {kandidat dalam Ck+1 dengan min_support}
? end
? return .k Lk;
```

Knowledge Discovery in Databases (KDD) merupakan sebutan lain dari tahapan data mining. Langkah-langkah yaitu :

- Seleksi Data

Data utama dalam penelitian ini yaitu data transaksi penjualan online shoop retail yang berbasis di Inggris diambil pada periode 01/12/2010-31/12/210 atau \pm 1 bulan transaksi.

- **Praproses Data /Pembersihan Data**
 Pada tahap ini merupakan tahap yang dapat dikatakan sebagai tahap pembersihan data. Data pertama yang terkumpul biasanya masih banyak terdapat data yang kosong, data yang tidak konsisten, atau data yang sama/ kembar. Data yang masih tercakup seperti yang telah dijelaskan selanjutnya akan disaring hingga mendapatkan output data yang benar sesuai dengan yang dibutuhkan untuk proses selanjutnya.
- **Transformasi Coding**
 Setelah melewati proses pembersihan data kemudian data akan di masukan ke dalam proses data mining. Langkah selanjutnya yaitu mengubah data keformat file Comma Separated Values (CSV) yang dapat terintegrasi pada Excel.
- **Data Mining**
 Kemudian tahap data mining, selanjutnya proses untuk mengolah data menggunakan algoritma. Dalam penelitian ini algoritma apriori dipilih sebagai solusi untuk pengolahan data. Contoh implementasi seperti pada gambar dibawah ini.



Gambar . Proses Data mining Algoritma Apriori

- **Interpretation/Evaluasi**
 Hasil akhir dari proses data mining agar pola informasi dapat dimengerti oleh pihak perusahaan ataupun supplier dapat ditampilkan dalam bentuk sederhana dan mudah dimengerti. Evaluasi yang berjalan pada tahapan ini merupakan memeriksa ulang dan membandingkan data atau informasi yang dirasa berbeda dengan data fakta yang ada dilapangan.

Hasil :

Tahapan awal yang dilakukan pada penelitian ini adalah mempersiapkan data, persiapan data yang akan diolah merupakan data public yang sama dengan dataset OrderMas, dataset tersebut diperoleh dari internet yaitu dataset Online Retail yang berbasis di Inggris. Kemudian

total jumlah data yang digunakan untuk penelitian yaitu diambil sebanyak lebih kurangnya 10.000 *record* data. Data-data tersebut merupakan data transaksi penjualan barang.

Pada proses ini data akan disaring dan dirubah menjadi data tabular oleh aplikasi yang sudah dibuat di delphy7 sehingga menjadi data yang siap untuk diolah dalam proses sistem data mining dengan *algoritma apriori*. Dari data yang berjumlah lebih kurang 10.000 *record* data, setelah disaring data yang siap digunakan berjumlah 324 *record* data dengan atribut yang digunakan yaitu:

1. Cutomer ID (id konsumen)
2. Discription (nama barang)

Gambar Dataset Online Retail dengan Dua Atribut

Pada gambar diatas dapat dijelaskan penulis merubah dataset awal yang memiliki beberpa atribut dan dirubah melalui proses aplikasi sehingga didapat 2 atribut yang akan digunakan untuk langkah penelitian selanjutnya.

Langkah selanjutnya adalah transformasi coding, merupakan pengolahan data yang berjumlah 324 record yang dilakukan dengan memasukkan data kedalam aplikasi RapidMiner kemudian system mengolahnya dengan *algoritma apriori* dengan metode *association rules* untuk mendapatkan hasil akhir sesuai yang diharapkan.

1. Analisis Data Mining

Sebagai perhitungan tentang *algoritma apriori* dengan metode assosiasi penulis mengambil data yang akan diolah sebagai contoh yaitu sebanyak 10 data penjualan barang seperti terlihat pada table dibawah. Penulis memberikan batasan nilai minimum support 0.5 atau sama dengan 50 %. dan confidence 0.5 atau sama dengan 50%. Berikut ini merupkan sampel data barang yang sudah dijelaskan pada table.

Tabel Data Barang

No	Kode Barang	Nama Barang
1	A1	WHITE METAL LANTERN
2	A2	PINK FAIRY CAKE CHILDRENS APRON
3	A3	MAGIC DRAWING SLATE PURDEY
4	A4	SPACEBOY LUNCH BOX
5	A5	VICTORIAN SEWING KIT

6	A6	WOODLAND CHARLOTTE BAG
7	A7	DOORMAT NEW ENGLAND
8	A8	ALARM CLOCK BAKELIKE GREEN
9	A9	DOORMAT FAIRY CAKE
10	A10	MINT KITCHEN SCALES

Data diatas merupakan sample data barang yang akan digunakan pada penelitian ini dan berlaku untuk semua jenis. Akumulasi transaksi penjualan barang dapat dilihat sebagai berikut :

Tabel Data Transaksi

Transaksi	Barang yang dibeli (berdasarkan kode barang)
1	A2,A4,A5,A8,A9
2	A1,A6,A7,A9
3	A1,A2,A3,A4,A7,A8,A9,A10
4	A2,A6,A7,A8,A9
5	A1,A2,A3,A4,A6,A7,A8,A9,A10
6	A3,A4,A5,A8,A10
7	A2,A4,A5,A7
8	A1,A7,A8,A9
9	A2,A3,A9,A10
10	A1,A6,A8,A10

Pada data transaksi penjualan barang dibentuk tabel tabular yang akan memudahkan dalam mengetahui beberapa banyak item yang dibeli disetiap tansaksi.

Tabel Data Transaksi Tabular

Transaksi	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
1	0	1	0	1	1	0	0	1	1	0
2	1	0	0	0	0	1	1	0	1	0
3	1	1	1	1	0	0	1	1	1	1
4	0	1	0	0	0	1	1	1	1	0
5	1	1	1	1	1	0	1	1	1	1
6	0	0	1	1	1	0	0	1	0	1
7	0	1	0	1	1	0	1	0	0	0
8	1	0	0	0	0	0	1	1	1	0
9	0	1	1	0	0	0	0	0	1	1
10	1	0	0	0	0	1	0	1	0	1

Penyelesaian berdasarkan tabel yang ada pada tabel 2 proses pembuatan C1 atau bisa dikatakan dengan 1 item dan jumlah minimum support = 50% .

Tabel Support dari Setiap Barang

Kode Barang	Proses Support	Support
A1	$(5/10)*100$	50%
A2	$(6/10)*100$	60%
A3	$(4/10)*100$	40%
A4	$(5/10)*100$	50%
A5	$(3/10)*100$	30%
A6	$(4/10)*100$	40%
A7	$(6/10)*100$	60%
A8	$(7/10)*100$	70%
A9	$(7/10)*100$	70%
A10	$(5/10)*100$	50%

Dari Proses pembentukan item pada tabel diatas dengan minimum support 50% dapat diketahui yang memenuhi standar minimum support yaitu ada 7 Jenis, dari ke 7 jenis tersebut kemudian dibentuk kombinasi item 2 item. Tabel berikut merupakan jenis item yang memenuhi standart minimum support :

Tabel Data Barang yang Memenuhi Support

Kode Barang	Support minimal 50%
A1	50%
A2	60%
A4	50%
A7	60%
A8	70%
A9	70%
A10	50%

Pembentukan C2 atau bisa dikatakan dengan 2 item dengan jumlah minimum support =50% dan dapat terselesaikan dengan rumus Support (A,B) = P(A B) Support(A,B).

Tabel Data dari 2 itemset dengan Minimum Support 50%

Kode Barang	Jumlah	Proses Support	Support %
A1-A2	2	(2/10)*100	20%
A1-A3	2	(2/10)*100	20%
A1-A4	2	(2/10)*100	20%
A1-A5	0	(0/10)*100	00%
A1-A6	3	(3/10)*100	30%
A1-A7	4	(4/10)*100	40%
A1-A8	4	(4/10)*100	40%
A1-A9	4	(4/10)*100	40%
A1-A10	3	(3/10)*100	30%
A2-A3	3	(3/10)*100	30%
A2-A4	4	(4/10)*100	40%
A2-A5	2	(2/10)*100	20%
A2-A6	2	(2/10)*100	20%
A2-A7	4	(4/10)*100	40%
A2-A8	4	(4/10)*100	40%
A2-A9	5	(5/10)*100	50%
A2-10	3	(3/10)*100	30%
A3-A4	3	(3/10)*100	30%
A3-A5	1	(1/10)*100	10%
A3-A6	1	(1/10)*100	10%
A3-A7	2	(2/10)*100	20%
A3-A8	3	(3/10)*100	30%
A3-A9	3	(3/10)*100	30%
A3-A10	4	(4/10)*100	40%
A4-A5	3	(3/10)*100	30%
A4-A6	1	(1/10)*100	10%
A4-A7	3	(3/10)*100	30%
A4-A8	4	(4/10)*100	40%
A4-A9	3	(3/10)*100	30%
A4-A10	3	(3/10)*100	30%
A5-A6	0	(0/10)*100	0%
A5-A7	1	(1/10)*100	10%
A5-A8	2	(2/10)*100	20%
A5-A9	1	(1/10)*100	10%
A5-A10	1	(1/10)*100	10%
A6-A7	3	(3/10)*100	30%

A6-A8	3	$(3/10)*100$	30%
A6-A9	3	$(3/10)*100$	30%
A6-10	2	$(2/10)*100$	20%
A7-A8	3	$(3/10)*100$	30%
A7-A9	5	$(5/10)*100$	50%
A7-A10	2	$(2/10)*100$	20%
A8-A9	5	$(5/10)*100$	50%
A8-A10	4	$(4/10)*100$	40%
A9-A10	3	$(3/10)*100$	30%

Dari kombinasi 2 itemset dengan minimum support 50% maka diketahui item yang memenuhi standart minimum support yaitu sebagai berikut :

Tabel Data 2 itemset yang Memenuhi Support 50%

Kode Barang	Nama Barang	Support
A2	PINK FAIRY CAKE CHILDRENS APRON	50%
A7	DOORMAT NEW ENGLAND	50%
A9	DOORMAT FAIRY CAKE	50%

Dari kombinasi 2 item akan dibentuk 3 item set dan diambil beberapa proses untuk proses pencarian 3 itemset sebagai berikut:

Tabel Data dengan Kombinasi 3 Itemset

Kode Barang	Jumlah Barang	Proses Support	Support
A1-A2-A3	2	$(2/10)*100$	20%
A3-A4-A5	1	$(1/10)*100$	10%
A5-A6-A7	0	$(0/10)*100$	0%
A6-A7-A-8	2	$(2/10)*100$	20%
A7-A8-A9	4	$(4/10)*100$	40%
A8-A9-A10	2	$(2/10)*100$	20%

Dalam gabungan tiga item set semua data belum bisa mencakup minimal 50%, oleh karena itu gabungan dua item set untuk memcukupi dalam bentuk asosiasi.

Hal tersebut untuk memenuhi aturan dan menghitung confidence dalam asosiatif antar item dengan minimum confidence adalah 70% nilai yang didapat.

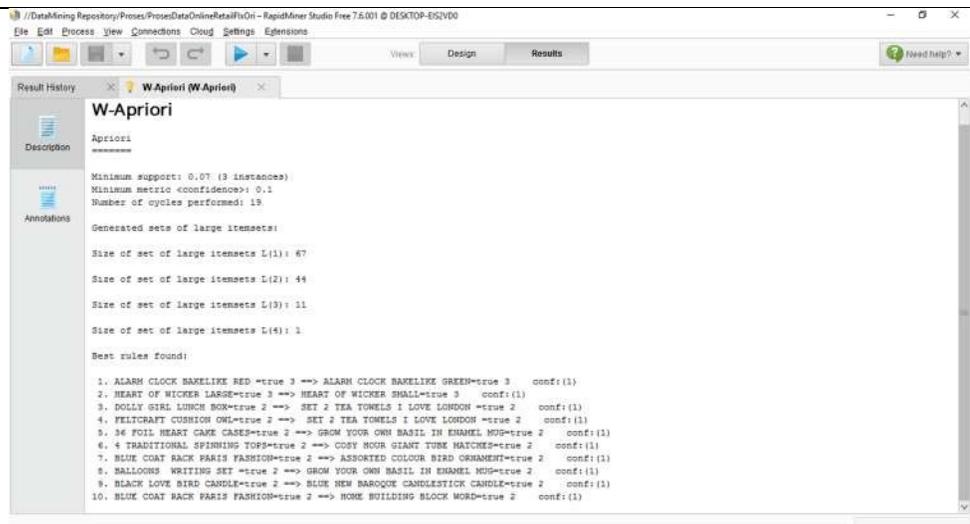
Tabel Data Hasil Perhitungan Confidence Asosiasi

Hubungan Antar Item	Proses Confidence	Nilai Confidence
A2=>A9	$(5/5)*100$	100%
A7=>A9	$(5/6)*100$	83%

Dari tabel diatas menunjukkan nilai asosiasi, bahwa konsumen sering membeli Pink Fairy Cake, Doormat New englang dan Doormat Fairy Cake karena data dari item tersebut telah memenuhi nilai minimum confidence yang ditentukan.

2. Simulasi Data dengan Aplikasi RapidMiner

Dari hasil proses data menjadi data tabular pada table 3, kemudian akan diproses menggunakan aplikasi RapidMiner dengan extenstion W-apriori akan menghasilkan aturan asosiasi dan beberapa rules berikut ini .



Gambar Rules Hasil Simulasi RapidMiner

Setelah data melewati tahapan pre-prosesing data oleh RapidMiner, kemudian data dapat diolah dengan menentukan nilai minimal support 50% dan minimal confidence 50% dan menentukan jumlah rule yang akan di tampilkan yaitu 10 rules.

Hasil keseluruhan dari proses implementasi apriori yang disimulasikan pada RapidMiner terhadap data penjualan/transaksi pada dataset OnlineRetail sebanyak 324 data dengan memberikan batasan nilai minimum support dan confidence yang sudah ditentukan, sehingga mendapatkan hasil asosiasi bahwa jenis barang yang paling banyak terjual adalah “Alarm Clock Bakelike Red, Alarm Clock Bakelike Green, Heart of Wicker Large dan Heart of Wicker Small”.

Jadi e-commerce OrderMas akan memberikan sebuah informasi/notifikasi kepada supplier yang menjual barang-barang yang paling laku terjual untuk segera menyetok barang tersebut.

Asosiasi apriori akan maksimal apabila semakin tinggi nilai minimum support dan minimum confidence maka nilai akurasi semakin tinggi pula, karena hanya item-item yang memiliki asosiasi yang kuat dalam pembelian konsumen yang akan dimunculkan dalam hasil perhitungan.

Kelebihan	:	Dari pola hubungan yang ada e-commerce OrderMas memberikan sebuah informasi/notifikasi kepada supplier yang menjual barang-barang yang paling laku terjual untuk segera menyetok barang tersebut. Asosiasi apriori akan maksimal apabila semakin tinggi nilai minimum support dan minimum confidence maka nilai akurasi semakin tinggi pula, karena hanya item-item yang memiliki asosiasi yang kuat dalam pembelian konsumen yang akan dimunculkan dalam hasil perhitungan.
Kekurangan	:	Dalam penggunaan aplikasi RapidMiner untuk analisis data menggunakan W-apriori , banyak tools dan operator yang mempunyai fungsi tertentu, jadi di perlukan pemahaman tentang aturan asosiasi dan Data Mining pada aplikasi RapidMiner.

Perbandingan dengan metode usulan/penelitian usulan	:	Perbandingan metode yang diterapkan pada paper ini yaitu <i>algoritma apriori</i> untuk penelitian usulan tetap akan menggunakan <i>algoritma apriori</i> tetapi akan mengembangkannya ke dalam objek yang berbeda, dan dengan menggunakan dataset yang lebih besar. Dengan menggunakan dataset yang lebih besar, bisa mendapatkan nilai support dan confidence yang bervariasi, sehingga dapat memperoleh banyak asosiasi rules antar item.
---	---	--

Tugas 02

Nama : Puspita Dewi Setyadi

NIM : 202420011

Implementasi *Association Rule Mining* Untuk Menentukan Menu Paket Makanan Dengan Algoritma *FIN* Menggunakan *Nodesets* (Studi Kasus R.M. Lesehan Nova Sragen)

Rumah makan nova menyediakan 2 macam jenis menu yaitu menu utama dan menu paket. Menu utama berisi daftar seluruh item, makanan yang berjumlah 156 item dan minuman yang berjumlah 44 item sedangkan menu paket terdiri dari 30 item paket makanan dan minuman. Untuk tipe pengunjung dengan pola pemesanan menu memilih sendiri-sendiri. Pihak rumah makan menyediakan buku menu utama sebagai media memilih menu kepada tiap anggota rombongan. Disebabkan pemilihan menu pilihan yang banyak, setiap anggota rombongan dalam memilih menu memerlukan waktu yang lama. Nota pesanan yang dihasilkan oleh tipe rombongan ini memiliki variasi item pesanan yang tinggi. Sehingga apabila variasi item pesanan yang tinggi maka akan membutuhkan waktu yang lebih lama dalam proses pembuatan setiap item pesanan. Menggunakan metode *Association Rule Mining* untuk menemukan kombinasi item makanan dan minuman *Association Rule* melakukan mining terhadap data transaksi penjualan yang tercatat pada struck pembayaran yang dicetak oleh mesin kasir dalam kurun waktu tertentu.

Algoritma *FIN* adalah algoritma *frequent itemset* yang menggunakan struktur data *novel* berupa *nodesets*. Langkah-langkah algoritma *FIN* sebagai berikut :

- a. Menghitung nilai *support* untuk setiap *item* pada daftar menu.
- b. Mengurutkan *item* berdasarkan nilai *support* terbesar sampai terkecil.
- c. Mengeliminasi seluruh *item* yang memiliki nilai *support* kurang dari minimum *support*.
- d. Mengeliminasi *item-item* setiap transaksi yang memiliki nilai *support* kurang dari minimum *support*.
- e. Mengurutkan *item-item* yang tersisa untuk setiap transaksi sesuai dengan urutan *L*.
- f. Memasukkan *item-item* pada setiap transaksi pada *POC-Tree*.
- g. Memindai *POC-Tree* untuk menyusun *frequent 1-itemset*.
- h. Memindai *POC-Tree* untuk menyusun *frequent 2-itemset* yang berdasar pada *frequent 1-itemset*.
- i. Memindai *POC-Tree* untuk menyusun *frequent 3-itemset* yang berdasar pada *frequent 2-itemset*.
- j. Menyusun kombinasi menu paket makanan dan minuman berdasarkan *frequent 3-itemset*.

Pembahasan tentang domain masalah yang dibahas dalam paper :

Internet merupakan jaringan komputer terbesar di dunia, dengan jumlah pengguna lebih dari 580 juta. Dari perspektif pengguna, setiap *resource* atau jaringan yang lebih kecil tentunya memiliki nama yang unik sehingga bisa dicari untuk proses komunikasi data, salah satu nama unik tersebut disebut sebagai *domain name*. *domain name* dikelola oleh perangkat komputer yang disebut sebagai *domain name server* *DNS* merupakan salah satu layanan yang berisiko mengalami serangan, beberapa serangan *DNS* yang ditemukan adalah *DNS cache snooping*, *DNS poisoning*, *DNS spoofing*, dan sebagainya tentunya tujuannya berbagai macam seperti membelokkan arah *browsing* menuju *website* yang salah. Salah satu ancaman yaitu *Cache snooping* bisa digunakan untuk menentukan situs/*host*, siapa *klien* dan penggunanya, dan informasi lain yang berguna bagi penyerang. Selain banyaknya ancaman seperti yang sudah disebutkan sebelumnya ancaman terhadap *DNS* bisa juga digunakan untuk melihat *software* yang digunakan sebuah *host* dari *resource record* yang berisi alamat *update software*.

Pendapat saya : Berdasarkan besarnya ancaman tersebut tentunya sudah seharusnya dilakukan proses manajemen keamanan jaringan dengan salah satu cara melakukan pengukuran terhadap jaringan yang digunakan, Ada beberapa alat ukur yang dapat digunakan seperti *nessus*, *openVAS*, dan *nmap*. *Nessus* sendiri merupakan sebuah program yang memang dirancang untuk *security scanner* yang berfungsi untuk mengaudit keamanan sebuah sistem. Selain itu juga dapat melakukan pengujian

manajemen *DNS Server* Menggunakan Standar Keamanan *National Institute of Standards and Technology (NIST)* dengan langkah-langkah pengujian yang akan dilakukan.

Nama : Robby Prabowo

NIM : 202420001

Kelas MTIA1

JUDUL

Analysis of Attribute Association in Heart Disease Using Data Mining Techniques

PENULIS

K.Srinivas 1, G.Raghavendra Rao 2 and A.Govardhan 3

TAHUN

2012

JURNAL & VOL

International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
<http://www.ijera.com> Vol. 2, Issue4, July-August 2012, pp.1680-1683

ABSTRAK

Di data mining, penggalian rule asosiasi menunjukkan sebuah teknik yang menjanjikan untuk menemukan pola yang tersembunyi di dalam database yang besar. Isu utama tentang penggalian rule asosiasi pada data kesehatan yang jumlahnya ratusan banyak ditemukan, tetapi kebanyakan tidak relevan. DSS ditujukan untuk diagnosis CVD.

Dataset yang digunakan untuk pembangkitan DSS dan evaluasi ada 1897 subyek, masing-masing 21 fitur yang meliputi demografi dan data rekaman seperti uji laboratorium. Beberapa jumlah rule membuat pencarian lambat. Bagaimanapun, tidak semua rule yang dibangkitkan menarik, ada beberapa rule yang ditolak. Di bidang kesehatan, rule asosiasi berhubungan ukuran data penyakit dengan factor resiko dan penyakit sekarang. Rule asosiasi medis signifikan dievaluasi dengan support dan confidence. Rule asosiasi dibandingkan dengan rule prediksi yang digali dengan decision tree, dengan teknik mesin learning . Pada paper ini kami mengusulkan system baru untuk menemukan kekuatan asosiasi antara atribut dengan set data yang diberikan. Sistem ditujukan mempunyai keuntungan sejak dibangkitkan, hal ini menyediakan diagnosis CVD berdasarkan dengan mudah dan tidak “invasive” fitur dikenali dan bisa digunakan untuk interpretasi pembuatan keputusan

USULAN KERJA

Setelah dihitung frekuensi dari beberapa atribut di dataset memiliki banyak asosiasi di antara atribut

age				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	25-35	7	2.3	2.3
	36-45	56	18.5	20.8
	46-55	88	29.0	49.8
	56-65	119	39.3	89.1
	66-75	31	10.2	99.3
	>75	2	.7	100.0
	Total	303	100.0	

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	97	32.0	32.0	32.0
	Male	206	68.0	68.0	100.0
	Total	303	100.0	100.0	

Tabel di atas menunjukkan bahwa resiko CVD lebih pada laki-laki dengan range umur antara 56 dan 65. Alat ujur yang sama diterapkan untuk menghitung stastitika pada frekuensi setiap atribut dan kami dapat mengaplikasikan untuk prediksi serangan jantung.

KESIMPULAN

Pada paper ini, kami mengusulkan alat ukur baru untuk menemukan asosiasi di anantara atribut yang bervariasi di dataset. Metode ini membangkitkan rule asosiasi yang valid dengan mengambil ukur peluang. Kami melakukan eksperimen pada sintesis dan set data real. Kami telah menerapkan ukuran pada kedua frequent dan infrequent itemset juga memiliki asosiasi di antara atributnya. Tipe dari asosiasi mungkin pada kasus penyakit. Pasien mungkin memiliki penyakit dan dapat dirawat dengan gejala yang jarang seperti 18 tahun laki-laki muda mendapat serangan jantung. Pada kerja selanjutnya, kami berharap melakukan eksperimen pada dataset kesehatan yang real time untuk prediksi penyakit seperti serangan jantung dan membandingkan performa dengan algoritma kami dengan algoritma yang lain.

Pendapat : Penelitian ini dapat mengetahui perbandingan algoritma mana yang lebih baik untuk mendeteksi penyakit jantung.

Nama : Shabila Fitri Aulia
Nim : 202420024
Mata Kuliah : Advanced Database
Dosen : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

SOAL

Silahkan cari satu paper yang membahas tentang association rule dan buat ringkasan dan pembahasan tentang domain masalah yang dibahas dalam paper dan berikan pendapat anda (review paper). Format bebas.

Judul Penelitian : Implementasi Data Mining Untuk Pengaturan Layout Minimarket dengan Menerapkan Association Rule

Penulis : Maharani, Nelly Astuti Hasibuan, Natalia Silalahi, Surya Darma Nasution, Mesran, Suginan, Diana U Sutiksno, Heri Nurdiyanto, Efori Buulolo, Yuhandri

Publikasi : Jurnal Riset Komputer(JURIKOM) Vol.4 No.4, Agustus 2017
ISSN 2407-389X (Media Cetak)

Reviewer : Shabila Fitri Aulia

Latar Belakang : PT.Indomarco Pristama adalah perusahaan yang bergerak dalam bidang retail yang menghasilkan banyak kegiatan transaksi. Data pembelian yang selalu bertambah setiap hari hanya akan menjadi tidak berguna jika tidak diolah kembali sehingga data-data tersebut perlu diolah kembali agar dapat menambah pengetahuan yang terdapat dalam data tersebut.

Tujuan Penelitian : Penelitian ini bertujuan untuk mengolah data-data yang terkumpul yang membentuk pola pembelian sehingga dari pola ini bisa diketahui minat beli dan kebiasaan berbelanja pelanggan.

Metode Penelitian : Metode yang digunakan dalam penelitian ini adalah metode association rule yang dapat menentukan atribut yang akan didapatkan bersama. Association Rule bertugas untuk mencari aturan yang tidak mengcover untuk mengukur hubungan antara dua atau lebih atribut.

Hasil Penelitian : Berdasarkan hasil perhitungan dari nilai *support*, *confidence*, dan *rule*, dapat dilihat bahwa produk yang sering dibeli pelanggan secara bersamaan dalam transaksi penjualan adalah produk dalam dept dan dept 8. *Rule* yang didapat salah satunya bisa digunakan untuk memperbaiki *layout* idm.

Pendapat reviewer : Menurut saya, jurnal sudah dituliskan dengan baik dan dibuat dengan sesuai ketentuan akan tetapi untuk perkembangan aplikasi kedepannya sebaiknya di bagian analisa tabel dikembangkan lagi agar tidak terjadi sistem error. Misal letakny disesuaikan dengan letak ciki atau brand produk sehingga meminimalisir kebingungan data. Selain itu, kedepannya agar lebih menambahkan item lebih dari 3 karena lebih banyak item maka sistem akan lebih baik.

ANALISIS DATA GANGGUAN KERUSAKAN MESIN PRODUKSI MENGUNAKAN TEKNIK ASSOCIATION RULES

Association rules merupakan salah satu teknik data mining yang digunakan untuk menentukan korelasi dari sebuah dataset terkait keputusan yang akan diambil. Berdasarkan data yang diperoleh dilakukan pengkategorian untuk variabel analisis terpilih. Berikutnya data yang diperoleh akan diolah dengan menggunakan teknik Association Rule untuk melihat pola yang terjadi dengan bantuan packages arules software R. Hasil analisis menunjukkan bahwa terdapat hubungan yang kuat antara kerusakan pada shift 2 untuk jenis kerusakan mekanik, serta kerusakan pada shift 3 dengan lama perbaikan sedang dan tingkat resiko yang tinggi memiliki peluang kejadian yang lebih tinggi jika dibandingkan kejadian lainnya.

Dari penelitian jurnal ini menggunakan data gangguan sistem produksi berupa kerusakan mesin yang terjadi pada PT Z untuk bulan Januari - Agustus tahun 2018. Berdasarkan data yang diperoleh dilakukan peng-kategorian untuk variabel analisis terpilih. Berikutnya data yang diperoleh akan diolah dengan menggunakan teknik association rules untuk melihat pola yang terjadi. Pada kasus ini software yang digunakan untuk memudahkan proses analisis data yaitu software R dengan bantuan packages rules untuk mendapatkan pola data.

Pembahasan: berdasarkan tahapan pengumpulan data, didapatkan data kerusakan mesin yang terjadi antara bulan Januari 2018 sampai dengan tanggal 13 Agustus 2018 sebanyak 314 data kejadian kerusakan yang terjadi dan lama kerusakan 10,570 menit atau 176.17 jam. Tabel 2 berikut ini adalah data kejadian kerusakan mesin dengan lama perbaikan kerusakan yang lama yang perlu dieksplorasi lebih lanjut oleh divisi maintenance terkait akar penyebabnya serta solusi perbaikannya untuk mengurangi lamanya waktu perbaikan. Hasil analisis menunjukkan bahwa terdapat hubungan yang kuat antara kerusakan pada shift 2 untuk jenis kerusakan mekanik, serta kerusakan pada shift tiga dengan lama perbaikan sedang dan tingkat resiko yang tinggi memiliki peluang kejadian yang lebih tinggi jika dibandingkan kejadian lainnya.

Permasalahan: terjadi kejadian breakdown sebanyak 314 data dengan lama kerusakan 10,570 menit atau 176.17 jam yang telah melebihi KPI untuk Divisi Maintenance PT Z. Hasil pemetaan data gangguan kerusakan mesin produksi menunjukkan bahwa jenis kerusakan yang paling sering terjadi adalah kerusakan mekanik dan didominasi dengan kerusakan dengan tingkat resiko sedang, serta kebanyakan terjadi pada shift 1 dan shift 3.

Dalam penelitian ini ada beberapa tinjauan pustaka yaitu:

a. **Algoritma Apriori**

Algoritma Apriori merupakan algoritma dasar yang diusulkan oleh Agrawal penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu: support dan confidence.

b. **Association Rule**

Association rules merupakan salah satu teknik yang ada di dalam data mining yang digunakan untuk menemukan kombinasi item yang sering terjadi dalam sebuah dataset. Konsep dasar dari association rules mencari pola yang sering muncul atau terjadi di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item sehingga teknik ini akan mendukung rekomendasi sistem melalui penemuan pola antar item dalam transaksi-transaksi yang terjadi.

Review jurnal : jurnalnya bagus dan mudah di mengerti, tapi di bagian permasalahannya kurang efektif. Lebih di pikirkan lagi tentang kerusakan mesinnya. Sekian dan terima kasih.

2.1. Sumber Data

... 9 ... 2009 ... 2571 ... 14 ... World Meteorological Organization¹⁵ ... 0 ...

2.2. Pemilihan Perangkat lunak dan teknik Data mining

... Orange Ailab ... Association Rule, Classification Tree, ... Random Forest ... supervised ... 4 ...

2.3. Pengolahan Data

... d ... d ... d ... d ... d ... d ...

Seksi 0

... 1 ... 99 ...

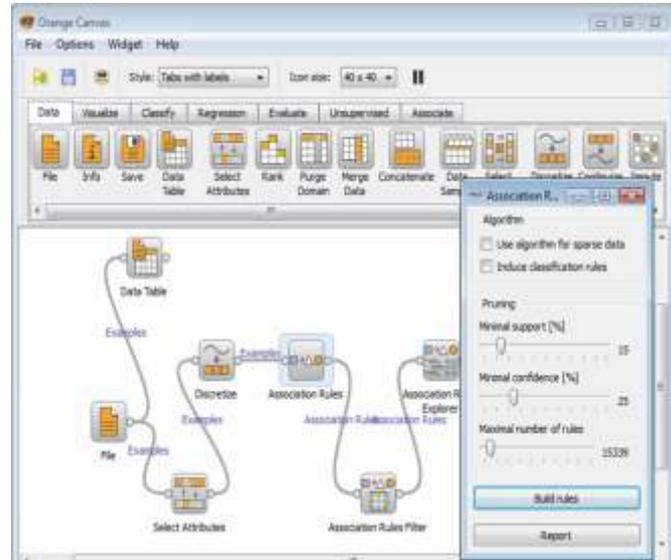
Seksi 1

... dd ... 00 ... 2 ... d ... d ... 29 ... 3 ... 4 ... 4 ... 3 ... 5 ... 6 ... 7 ... 1 ... 2 ... 7 ... 8 ... M ... 9 ...

2.4. Pembuatan Model

2.4.1. Metode Association Rule

... support ... 15 ... confidence ... 25 ... maximum rule ... 15399 ... support ... 15 ... confidence ... 25 ... rule ... Lift ratio ... 1 ... 2 ...



... 2 ... Association Rule ... Orange Ailab

2.4.2. Metode C4.5, Classification Tree, dan Random Forest

... d ... d ...

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan mengevaluasi efektivitas algoritma data mining dalam memprediksi cuaca. Hasilnya menunjukkan bahwa algoritma yang digunakan mampu meningkatkan akurasi prediksi cuaca dibandingkan dengan metode konvensional. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi cuaca yang lebih akurat dan efisien.

V. DAFTAR PUSTAKA

1. *Data mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*.
2. *European Journal of Scientific Research*, 47(2) 295-308.
3. *Proceedings of Conference on Intelligent data Understanding* 213-227.
4. *Presentation/webcast*.
5. *Journal of Earth Science Informatics*, 1(2) 49-57.
6. *International Journal of Advanced*

Networking and Applications. 2(2) 608-613.

7. *Proceedings of IADIS European Conference Data Mining* 81-88.
8. *Journal of Intelligent System*. 18(3) 193-209.
9. *International Journal of Informatics and Communication Technology Research*, 1(3) 96-100.
10. *International Journal of World Academy of science, engineering and Technology*. v41-43 248-253.
11. *Proceedings of International Conference on Data Mining and Knowledge Discovery* 18-2011.
12. *Penerapan Data mining di Badan Meteorologi dan Geofisika untuk memprediksi cuaca di Jakarta*.
13. *Proceedings of International Conference on Data Mining and Knowledge Discovery* 8-2010.
14. *Proceedings of International Conference on Data Mining and Knowledge Discovery* 01-2010.
15. *Guide to Marine Meteorological Services*.
16. *Decision Support and Business Intelligence Systems. Ninth Edition*.

REVIEW JURNAL

Nama : Surta Wijaya

NIM : 202420014

JUDUL	PEMANFATAN DATA MINING UNTUK PRAKIRAAN CUACA
Nama Jurnal	Jurnal Meteorologi Dan Geofisika
Volume dan halaman	Volume 12 Nomor 2, 189 - 195
Tahun	2011
Penulis	Subekti Mujiasih
Reviewer	Surta Wijaya
Tanggal Reviewer	5 Oktober 2020
Latar Belakang	BMKG memiliki sekitar 10 stasiun meteorologi maritim dan 3 (tiga) stasiun yang diperbantukan untuk memberikan pelayanan meteorologi maritim. Sebagian besar stasiun tersebut melakukan pengamatan sinoptik dan sebagian diantaranya memberikan pelayanan analisa dan prakiraan cuaca maritim. Data pengamatan ini sangat penting untuk melihat karakteristik cuaca setempat dan pembuatan informasi prakiraan beberapa hari ke depan.
Tujuan Penelitian	Tujuan dari penelitian ini adalah untuk memperoleh model prakiraan yang sesuai agar memudahkan proses analisa dan prakiraan cuaca
Permasalahan	<p>Dalam proses pembuatan informasi prakiraan cuaca, terdapat beberapa kendala. Pertama, sulitnya membuat informasi prakiraan karena melibatkan banyak sumber data seperti data pengamatan, data model aplikasi cuaca, data gambar kondisi awan dari satelit, data kondisi awan dari radar. Kedua, prakiraan cuaca maritim umumnya mengandalkan kemampuan dari prakirawan, sehingga intepretasi yang dihasilkan bisa berbeda antar prakirawan satu dengan yang lain karena bergantung dari pengalaman masing masing. Perbedaan interpretasi dapat membingungkan pengguna yang pada akhirnya berpeluang menurunkan kualitas informasi yang disampaikan.</p> <p>Berdasarkan masalah tersebut penulis bermaksud melakukan kajian model prakiraan untuk memperoleh model yang sesuai sehingga memudahkan proses analisa dan prakiraan cuaca maritim</p>
Metodologi Penelitian	Data sinoptik berasal dari pengamatan 9 (sembilan) stasiun meteorologi maritim tahun 2009 berukuran (2571 raw data)14). Data Sinoptik adalah data pengamatan cuaca permukaan yang dikirim dari stasiun-stasiun pengamatan cuaca di seluruh Indonesia setiap tiga jam. Data tersebut meliputi suhu udara, jumlah curah hujan, arah dan kecepatan angin, tekanan udara dan sebagainya.

	<p>Perangkat lunak yang dipilih adalah <i>Orange Ailab</i> dengan menggunakan metode <i>Association Rule</i>, <i>Classification Tree</i>, dan <i>Random Forest</i>.</p> <p>Pada pengujian ini dipilih <i>support</i> = 15% dan <i>confidence</i> = 25%, <i>maximum rule</i> = 15399, dengan alasan nilai <i>support</i> sebesar 15% akan memperkecil jumlah <i>rule</i> yang terjadi dan menghemat memori komputer. Adapun <i>confidence</i> sebesar 25% adalah untuk memperkuat asosiasi antar <i>rule</i> sehingga <i>rule</i> yang dihasilkan dapat lebih bermanfaat untuk prediksi.</p> <p>Komponen data sinoptik yang menjadi masukan adalah suhu udara, suhu titik embun, perbedaan suhu udara dan suhu titik embun serta kecepatan angin. Metode yang digunakan untuk membuat model dengan masukan komponen data cuaca tersebut adalah <i>C4.5</i>, <i>Classification Tree</i>, dan <i>Random Forest</i>. Model keluaran masing-masing metode tersebut diuji dengan sebagian data masukan untuk melihat kehandalan model. Setelah itu, hasilnya dibandingkan untuk mendapatkan akurasi yang tertinggi, dan memutuskan model prediksi yang paling baik</p>
<p>Hasil Penelitian</p>	<p>Berdasarkan hasil pembuatan dan pengujian model pada teknik <i>Association rule</i> dan <i>Classification</i>, maka penulis berpendapat bahwa model prediksi yang dihasilkan dari dua metode tersebut tidak memiliki perbedaan yang besar baik dari komponen cuaca yang menyusunnya ataupun dari nilai syarat batas masing-masing komponen cuaca. Hasil <i>Association rule</i> menunjukkan factor penentu hujan atau tidak hujan adalah suhu udara dan suhu titik embun. Sedangkan <i>Classification</i> yang diwakili <i>C4.5</i> menunjukkan faktor penentu hujan atau tidak hujan adalah suhu udara, suhu titik embun, dan tutupan awan. Namun interval syarat batasnya tidak jauh berbeda. Menurut metode <i>Assocition Rule</i>, nilai syarat batas suhu udara ≥ 23.9 dan suhu titik embun ≥ 23.6. Pada metode <i>C4.5</i>, syarat terjadinya hujan jika suhu udara ≤ 25.9, suhu titik embun ≥ 22.3 ditambah tutupan awan ≥ 6. Dengan membandingkan tingkat akurasinya, <i>Association Rule</i> mempunyai tingkat akurasi 60.9%, sedangkan <i>C.45</i> mempunyai tingkat akurasi 68.5%. Dengan demikian model prediksi yang disarankan adalah model prediksi <i>C4.5</i>.</p>
<p>Kelebihan</p>	<p>Dengan adanya penelitian ini maka kita dapat mengetahui dan membandingkan besarnya tingkat akurasi penggunaan metode <i>association rule</i> dan <i>C.45</i> sehingga kita dapat memutuskan penggunaan metode yang lebih akurat dalam memprediksi cuaca</p>
<p>Kekurangan</p>	<p>Tingkat akurasi sebesar 68.5% ini sebenarnya masih mengandung resiko apakah prediksi cuaca sesuai dengan kenyataan. Untuk meningkatkan tingkat akurasi tampaknya diperlukan data seluruh stasiun meteorologi maritim yang memiliki interval 5 - 10 tahun</p>

Kesimpulan	Untuk memenuhi kecepatan dan ketepatan prakiraan, diperoleh model prediksi yang dapat digunakan selanjutnya adalah model keluaran C4.5 dengan tingkat akurasi 68.5%. Komponen cuaca yang dominan memungkinkan terjadinya hujan adalah suhu udara ≤ 25.9 , suhu titik embun ≥ 22.3 , dan tutupan awan ≥ 6 . Pada kajian model selanjutnya, diperlukan data dengan interval 5 - 10 tahun untuk memperbaiki akurasi model.
------------	--

Nama : Trada Ayang Pratiwi
 NIM : 202420020
 Mata Kuliah : Advance Database
 Dosen Pengasuh : Tri Basuki Kurniawan, S.Kom., M.eng., Ph.D.
 Judul Tugas : Silahkan cari satu paper yang membahas tentang association rule dan buat ringkasan dan pembahasan tentang domain masalah yang dibahas dalam paper dan berikan pendapat anda (review paper)

Judul Jurnal	Perancangan Market Basket Analysis Menggunakan Association Rule Untuk Pendukung Keputusan Promosi pada Sistem Penjualan Sun Young Cell
Volume	Vol 12, No.1, April 2017
Tahun	2017
Penulis	Rina Halim, dkk.
Reviewer	Trada Ayang Pratiwi
Tanggal	30 September 2020
Pendahuluan	Sun Young Cell merupakan toko yang bergerak di bidang penjualan handphone, tablet, aksesoris gadget, dan perangkat lainnya. Seringkali pemilik tidak melakukan analisis yang tepat dalam melakukan promosi yang disebabkan oleh pencatatan dan pengolahan transaksi tidak terorganisir dan tidak akurat sehingga menimbulkan kerugian yaitu, adanya produk-produk tertumpuk yang telah melewati masa penjualan, pengeluaran biaya promosi tidak efisien, dan berkurangnya profit yang berdampak pada terhambatnya pengembangan dan pemasaran kelancaran usaha yang dijalankan Berdasarkan permasalahan yang dihadapi Sun Young Cell, dibutuhkan solusi pemecahan, yaitu sistem penjualan dengan yaitu sistem penjualan dengan Market Basket Analysis yang dapat menganalisa transaksi penjualan dan menghasilkan pendukung untuk pengambilan keputusan promosi.

Tujuan	<ol style="list-style-type: none"> 1. Untuk mengatasi adanya produk-produk yang tertumpuk yang telah melewati masa penjualan 2. Untuk meningkatkan pengembangan dan pemasaran kelancaran usaha
Metedologi	<ol style="list-style-type: none"> 1. Metodologi pengumpulan data yang dilakukan penulis yaitu dengan cara observasi, wawancara dan dokumentasi pada toko Sun Young Cell 2. Metode pengembangan sistem yang dipakai penulis menggunakan metode air terjun (<i>waterfall</i>).
Hasil	<p>Berdasarkan uraian permasalahan yang telah dibahas, solusi pemecahan masalah yang diajukan penelitian ini adalah pembangunan sistem penjualan yang dilengkapi dengan aplikasi Market Basket Analysis yang berfungsi sebagai:</p> <ol style="list-style-type: none"> 1. Sarana pengelolaan transaksi 2. Sarana penyajian informasi 3. Sarana pendukung pengambilan keputusan <p>Untuk mendukung proses tersebut, maka sistem yang akan dibangun dapat melakukan proses sebagai berikut:</p> <ol style="list-style-type: none"> 1. Melakukan pengelolaan data maupun produk dan stok dan transaksi yang berkaitan dengan pembelian dan penjualan sehingga dihasilkan informasi yang akurat dan dinamis 2. Menyajikan informasi dalam bentuk rekap dan laporan berdasarkan pengelolaan data dan transaksi secara aktual, akurat, dan nyata secara cepat dan mudah dipahami. 3. Menghasilkan analisis Market Analysis yang dapat memberikan gambaran mengenai produk berdasarkan kecenderungan penjualan, sehingga dapat dilihat pola transaksi penjualan pelanggan yang dapat digunakan untuk membantu pengambilan keputusan promosi oleh pemilik
Kesimpulan	<ol style="list-style-type: none"> 1. Permasalahan yang dihadapi toko Sun Young Cell dengan sistem penjualan saat ini adalah pencatatan dan pengolahan transaksi penjualan yang tidak terorganisir dengan baik menimbulkan kesulitan

	<p>bagi pemilik untuk mengetahui kondisi sebenarnya maupun menganalisa data-data yang ada untuk pengambilan keputusan yang tepat terkait pengembangan usaha, sehingga diperlukan sistem terkomputerisasi untuk pengolahan data-data penjualan yang dapat menghasilkan analisis dari penjualan untuk mendukung pengambilan keputusan.</p> <p>2. Penelitian ini menghasilkan perancangan Market Basket Analysis yang diterapkan pada sistem penjualan Sun Young Cell untuk pengolahan dan penyajian data-data penjualan yang dapat memberikan gambaran tingkat penjualan produk tertentu berdasarkan kecenderungan penjualan dengan produk lain untuk mendukung keputusan promosi. Analisa dihasilkan menggunakan teknik association rules berupa aturan-aturan yang menggambarkan pola transaksi penjualan. Rules yang dihasilkan dari perhitungan data yang dimasukkan ke sistem pada tanggal 17 Desember 2016 s.d 17 Januari 2017 dengan menggunakan minimum support 4% dan minimum confidence 10% menghasilkan rule tertinggi untuk produk, yaitu: if Samsung Galaxy J1 and Waterproofs Bag Case then Tempered Glass Screen Protector dengan nilai support 4.00% dan confidence 97.5610% 3.</p> <p>3. Sistem baru dibangun berbasis web menggunakan bahasa pemrograman PHP dan Javascript dengan basis data MySQL.</p>
--	--

Dari ringkasan jurnal diatas dapat disimpulkan bahwa

Association rule merupakan salah satu algoritma klasik data mining. *Association rule* digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset.

Market basket analysis menganalisis hubungan antara item pada transaksi untuk menemukan pengetahuan bermakna yang dapat digunakan untuk strategi pemasaran untuk meningkatkan penjualan item secara bersamaan.

Dengan adanya *Association Rule* pada toko Sun Young Cell maka dapat dihasilkan sebuah sistem penjualan menggunakan *market basket analysis* yang membantu pemilik melihat perkembangan usaha, mengetahui hasil informasi transaksi yang akurat, melakukan perbandingan dan membantu dalam pengambilan keputusan sehingga dapat meningkatkan penjualan maupun promosi pada toko Sun Young Cell.

Artinya manfaat dari *Association Rule* adalah :

1. Dapat memberikan masukan data dalam suatu market material / barang apa yang akan dijadikan promosi
2. Memberikan analisa kebiasaan konsumen dengan mencari asosiasi dan korelasi dari data transaksi sehingga dapat meningkatkan penjualan

Association Rule

Association rule mining adalah teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi *item*. *Interestingness measure* yang dapat digunakan dalam *asosiasi* adalah:

- Support* adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi.
- Confidence* adalah suatu ukuran yang menunjukkan hubungan antar dua *item* secara *conditional* (berdasarkan suatu kondisi tertentu).

Metodologi Dasar Association Rule

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap (Ulmer, David, 2002) :

- Analisis pola frekuensi tinggi

Merupakan tahap pembentukan kombinasi *item* yang memenuhi syarat *minimum support* yang ditentukan. Nilai *support* pada satu *itemset* diperoleh dengan rumus dibawah ini.

$$Support(A) = \sum \frac{Jumlah\ Transaksi\ mengandung\ A}{Total\ Transaksi} \quad \dots(2.1)$$

Pada Rumus diatas memiliki arti untuk menentukan nilai *support* pada satu *itemset* jumlah transaksi yang mengandung *item* A dibagi dengan jumlah transaksi yang ada pada *database*. Sedangkan pada dua *itemset* atau lebih dapat diperoleh dengan rumus seperti di bawah ini.

$$Support(A \cup B) = \sum \frac{Jumlah\ Transaksi\ mengandung\ A\ dan\ B}{Total\ Transaksi} \quad \dots(2.2)$$

Pada rumus diatas untuk menentukan nilai *support* pada dua *itemset* atau lebih, jumlah transaksi yang mengandung *item* A dan B dibagi dengan total transaksi yang terjadi pada *dataset*.

- Pembentukan Aturan Asosiasi

Langkah yang dilakukan setelah menentukan nilai *support* pada *itemset* berfrekuensi tinggi lalu dibentuk aturan asosiasi yang menyatakan kuatnya hubungan kombinasi *itemset* pada transaksi. Untuk menentukan aturan asosiasi yang terbentuk minimal *itemset* harus memiliki dua kandidat A dan B. Pada *rules* yang terbentuk berlaku hukum *asosiatif* A->B tidak berlaku B->A. Untuk menentukan aturan A->B digunakan rumus seperti di bawah ini.

$$Confidence = P(B | A) = \frac{Jumlah\ Transaksi\ mengandung\ A\ dan\ B}{Jumlah\ Transaksi\ mengandung\ A} \quad \dots(2.3)$$

Pada rumus diatas untuk menentukan nilai *confidence* pada aturan asosiasi, jumlah transaksi yang mengandung *item* A dan B dibagi dengan jumlah transaksi yang mengandung *item* A.

Langkah-langkah untuk melakukan perhitungan untuk menentukan *association rule* terdiri dari 2 langkah utama yaitu :

Langkah 1 : Temukan *Frequent Itemset*, yaitu : suatu *set item* yang memenuhi *minimum support* dan *confidence* yang telah ditentukan.

Langkah 2 : Gunakan *Frequent Itemset* tersebut untuk menghasilkan *association rules*.

Pembentukan Pola Transaksi

Berdasarkan data transaksi dilakukan akumulasi transaksi penjualan, Pola Transaksi *Itemset*

Transaksi	Item yang dibeli
1	A, B, C, E
2	E, G
3	D, F, A, B
4	E, C, A, B
5	D, A
6	E, G, C, A
7	A, E
8	A, C, G
9	D, C, E
10	A, F, G

Tabulasi Data Transaksi

Pada data transaksi penjualan di bentuk tabel *tabular* yang akan mempermudah dalam mengetahui berapa banyak *item* yang ada dibeli dalam setiap transaksi

Transaksi	A	B	C	E	F	G	H
-----------	---	---	---	---	---	---	---

Format Tabular Data Transaksi	1	1	1	1	0	1	0	0
	2	0	0	0	0	1	0	1
	3	1	1	0	1	0	1	0
	4	1	1	1	0	1	0	0
	5	1	0	0	1	0	0	0
	6	1	0	1	0	1	0	1
	7	1	0	0	0	1	0	0
	8	1	0	1	0	0	0	1
	9	0	0	1	1	1	0	0
	10	1	0	0	0	0	1	1
Jumlah	8	3	5	3	6	2	4	

Pembentukan *Itemset*

Pembentukan 1 *Itemset*

Berikut ini adalah penyelesaian berdasarkan data yang sudah disediakan pada tabel 2 proses pembentukan C1 atau disebut dengan 1 *itemset* dengan jumlah *minimum support* 40% Dengan rumus sebagai berikut

$$Support (A) = \frac{\sum \text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}} \times 100\%$$

Support dari setiap *Itemset*

Nama <i>Item</i>	Jumlah	<i>Support</i>
A	8	80%

B	3	30%
C	5	50%
D	3	30%
E	6	60%
F	2	20%
G	4	40%

Dari Proses Pembentukan *itemset* Pada Tabel 3.5 dengan *minimum support* 40%, maka dapat diketahui *item* yang memenuhi standar *minimum support* yaitu : 1

$Itemset = \{\{A\} \{C\} \{E\} \{G\}\}$.

Pembentukan *support* dari 1 *itemset*

Nama Item	Jumlah Transaksi Item	Support
A	8	80%
C	5	50%
E	6	60%
G	4	40%

Setelah dari hasil pembentukan 1 *itemset* telah ditemukan maka selanjutnya akan dilakukan kombinasi 2 *itemset*.

Pembentukan Kombinasi 2 *Itemset*

Proses Pembentukan C2 Atau disebut dengan 2 *itemset* dengan jumlah *minimum support* 40% Dapat diselesaikan dengan rumus berikut :

$$Support (A,B) = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{Total transaksi}} \times 100\%$$

Support dari 2 *itemset*

Nama Item	Jumlah Transaksi Item	Support
A, C	4	40%
A, E	4	40%

A, G	3	30%
C, E	4	40%
C, G	2	20%
E, G	2	20%

Dari Proses Pembentukan 2 *itemset* Pada Tabel 3.5 dengan *minimum support* 40%, maka dapat diketahui *item* yang memenuhi standar *minimum support* yaitu : 2

Itemset = {{A,C} {A,E} {C,E}}.

Pembentukan *support* dari 2 *itemset*

Nama Item	Jumlah Transaksi Item	Support
A, C	4	40%
A, E	4	40%
C, E	4	40%

Setelah dari hasil pembentukan 2 *itemset* telah ditemukan maka selanjutnya akan dilakukan kombinasi 3 *itemset*.

Pembentukan Kombinasi 3 *Itemset*

Proses Pembentukan C3 Atau disebut dengan 3 *itemset* dengan jumlah *minimum support* 40% Dapat diselesaikan dengan rumus berikut :

$$Support (A,B) = \frac{\sum \text{transaksi mengandung A, B, dan C}}{\sum \text{Total transaksi}} \times 100\%$$

Tabel 3.7 *Support* Dari 3 *itemset*

Nama Item	Jumlah Transaksi Item	Support
A, C, E	3	30%

Karena Kombinasi 3 *itemset* tidak ada yang memenuhi *minimal support* 40%, maka pencarian *itemset* dihentikan. Dan kombinasi *itemset* yang memenuhi pembentukan asosiasi adalah 2 *itemset*.

Pembentukan Aturan Asosiasi

Perhitungan *Support* dan *Confidence*

Setelah pola frekuensi tinggi ditentukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif A->B.

Minimum Confidence = 70%

Nilai *Confidence* dari aturan A->B diperoleh

$$Support = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{Total transaksi}} \times 100\%$$

$$Confidence = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}} \times 100\%$$

Perhitungan *Support* dan *Confidence*

Aturan	<i>Support</i>	<i>Confidence</i>
Jika membeli A maka membeli C	$(4/10) \times 100\% = 40\%$	$(4/8) \times 100\% = 50\%$
Jika membeli C maka membeli A	$(4/10) \times 100\% = 40\%$	$(4/5) \times 100\% = 80\%$
Jika membeli A maka membeli E	$(4/10) \times 100\% = 40\%$	$(4/8) \times 100\% = 50\%$

Jika membeli E maka membeli A	$(4/10) \times 100\% = 40\%$	$(4/6) \times 100\% = 66,66\%$
Jika membeli C maka membeli E	$(4/10) \times 100\% = 40\%$	$(4/5) \times 100\% = 80\%$
Jika membeli E maka membeli C	$(4/10) \times 100\% = 40\%$	$(4/6) \times 100\% = 66,66\%$

Hasil Dari Pembentukan *Support* dan *Confidence*

Aturan	<i>Support</i>	<i>Confidence</i>
Jika membeli C maka membeli A	40%	80%
Jika membeli C maka membeli E	40%	80%

Berdasarkan dari Hasil pencarian dengan *minimum support* 40% dan *minimum confidence* sebesar 70%, maka dapat disimpulkan bahwa Produk yang paling sering dibeli oleh konsumen adalah C, A dan C, E.

Dengan diketahuinya produk yang paling sering dibeli konsumen, maka perusahaan dapat menyusun strategi pemasaran dalam menentukan pembelian produk untuk menjaga ketersediaan stok produk yang dibutuhkan konsumen dan dapat mengatur tata letak produk berdasarkan kombinasi *itemset* produk yang terbentuk.

NAMA : WIDIA ASTUTI

NIM : 202420021

MATA KULIAH : ADVANCED DATABASE

Silahkan cari satu paper yang membahas tentang association rule dan buat ringkasan dan pembahasan tentang domain masalah yang dibahas dalam paper dan berikan pendapat anda (review paper)

Jawab :

Judul : Penerapan metode association rule mining pada data transaksi penjualan produk kartu perdana kuota internet menggunakan algoritma apriori

Pendahuluan

- Data transaksi penjualan produk kartu perdana kuota internet dijadikan sebagai bahan acuan untuk mengetahui seberapa besar tingkat penjualan produk yang telah dipasarkan oleh beberapa operator telekomunikasi seluler. Tujuannya untuk menemukan aturan asosiasi kombinasi antar item produk operator telekomunikasi seluler mana saja yang paling laku terjual di wilayah penjualan Priangan Timur meliputi cluster Ciamis, Garut dan Tasikmalaya
- Persaingan antara operator telekomunikasi seluler di Indonesia didominasi dan dikuasai oleh penguasa pangsa pasar industri telekomunikasi. Operator telekomunikasi seluler Telkomsel memiliki peredaran berjumlah 178 juta unit produk atau sekitar 45,3 %, selanjutnya diikuti Indosat Ooredoo sebanyak 96,4 juta unit atau 24,54 persen, Tri 56,8 juta unit disusul XL Axiata 50,5 juta unit dan Smartfren 11 juta unit
- Hal ini tentunya dibutuhkan suatu aplikasi yang dapat melakukan analisa terhadap data transaksi penjualan kartu perdana kuota internet. Salah satunya yaitu dengan menerapkan metode Association Rule Mining yaitu teknik data mining untuk menemukan aturan asosiasi suatu kombinasi item yang tersembunyi dalam database.
- Metode Association Rule Mining menggunakan algoritma apriori untuk melakukan analisa terhadap data transaksi penjualan produk kartu kuota internet yang diimplementasikan pada aplikasi yang dibangun. Hal ini untuk mempermudah dalam melakukan analisa terhadap daya saing produk operator telekomunikasi seluler mana saja yang memiliki tingkat penjualan produk yang paling laku terjual secara bersamaan dengan produk operator lainnya. Dengan demikian hasil yang diperoleh dapat digunakan untuk membantu pengambil keputusan dalam meningkatkan pemasaran dan promosi produk yang lebih baik.
- Algoritma apriori merupakan salah satu algoritma yang digunakan untuk melakukan proses pencarian frequent itemset dengan association rules. Algoritma apriori menggunakan pendekatan level – wise search, dimana k-itemset digunakan untuk memperoleh (k+1) itemset. Proses ini dilakukan hingga tidak ada lagi kombinasi yang dapat dibentuk. Pembentukan pola asosiasi oleh algoritma apriori terdiri dari dua tahap yaitu tahap pertama mencari frequent itemset (himpunan item yang memenuhi nilai minimum support) dan tahap kedua membentuk pola asosiasi dari frequent itemset yang telah didapat dengan menggunakan nilai confidence.

Metodelogi Penelitian

- Pengumpulan data
- Identification penerapan Association rule mining algoritma apriori dengan mencari kombinasi itemset pada data transaksi penjualan produk kartu perdana kuota internet yang menjadi objek penentuan aturan asosiasi untuk menemukan aturan dasar kombinasi item- item produk operator mana saja yang paling laku terjual dan menghitung banyaknya transaksi yang terjadi dalam setiap kombinasi tersebut
- metodologi dasar analisis asosiasi terbagi menjadi dua tahap:
 1. Analisa pola frekuensi tinggi
Pada tahap ini dicari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi untuk A}}{\text{Total Transaksi}}$$

Sedangkan untuk mencari nilai support dari 2-itemset dan seterusnya menggunakan rumus berikut:

$$\text{Support (A, B)} = \frac{\Sigma \text{Transaksi untuk A dan B}}{\text{Total Transaksi}}$$

$$\text{Support (A, B, C)} = \frac{\Sigma \text{transaksi untuk A,B dan C}}{\text{Total Transaksi}}$$

2. Pembentukan Aturan Asosiasi
Setelah semua pola frekuensi tinggi ditemukan, berikutnya mencari aturan asosiasi yang memenuhi syarat minimum untuk mengetahui confidence dengan menghitung nilai confidence aturan asosiatif $A \rightarrow B$. Nilai confidence dari aturan $A \rightarrow B$ diperoleh dari rumus sebagai berikut:

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\Sigma \text{transaksi untuk A dan B}}{\Sigma \text{Transaksi A}}$$

Pengembangan Sistem

Model proses untuk pengembangan perangkat lunak menggunakan model Extreme Programming (XP), dimana model tersebut model yang paling banyak digunakan untuk pengembangan perangkat lunak cepat. Model tersebut memiliki tahapan-tahapan yang terdiri dari : Planning (Perencanaan), Design (Perancangan), Coding (Pengkodean), dan Testing.

Pengujian

Pada tahap penelitian ini dilakukan untuk menguji penerapan dari metode association rule mining pada aplikasi yang telah diimplementasikan menggunakan perhitungan algoritma apriori. Pengujian ini bertujuan untuk membandingkan perhitungan yang dilakukan antara perhitungan manual dengan perhitungan sistem yang telah dibangun, hal ini berguna untuk mengetahui bahwa perhitungan yang dilakukan oleh sistem sudah sesuai dengan konsep perhitungan algoritma apriori itu sendiri.

Hasil dan Pembahasan

1. Analisa aturan Asosiasi perhitungan Algoritma Apriori

Data yang dianalisa adalah data transaksi penjualan produk kartu perdana kuota internet. Total sampel data yang diambil adalah 614 data transaksi dari jumlah data transaksi keseluruhan dari bulan Maret sampai bulan Juli 2018, kemudian dibagi menjadi tiga berdasarkan wilayah penjualan di Priangan Timur, meliputi cluster Ciamis memiliki 208 data transaksi, cluster Tasikmalaya memiliki 205 data transaksi dan cluster Garut memiliki 201 data transaksi. Terdapat 6 item atau atribut data yang digunakan dalam melakukan aturan asosiasi dengan nilai minimum support yang ditentukan yaitu 35% dan nilai minimum confidence yaitu 80%. Berikut adalah tabel 1 dari jumlah data transaksi proses mining aturan asosiasi dalam perhitungan algoritma apriori.

Tabel 1. Jumlah data transaksi

No	Cluster	Jumlah Data Transaksi	Min_Support	Min_Confidence
1	Ciamis	208	35%	80%
2	Tasikmalaya	205	35%	80%
3	Garut	201	35%	80%

2. Proses perhitungan Algoritma Apriori

Pada Tabel 2. Diketahui contoh sampel data transaksi penjualan produk kartu perdana kuota internet yang telah dilakukan pendataan ke setiap pasar outlet yang ada diwilayah penjualan Priangan Timur meliputi cluster Ciamis, Tasikmalaya dan Garut.

Tabel 2. Sampel data transaksi penjualan produk kartu perdana kuota internet

No	Tanggal	Nama Outlet	Cluster	Tsel	XL	Isat	Tree	Axis	Smartfren
1	23-03-2018	Agro Cell	Ciamis	150	125	130	95	80	45
2	23-03-2018	Budi Cell	Ciamis	185	0	0	0	0	0
3	23-03-2018	Phoenix Cell	Tasikmalaya	125	85	75	65	90	45
4	23-03-2018	Hoky Cell	Tasikmalaya	140	0	0	70	100	39
5	23-03-2018	Agro Cell	Ciamis	146	0	100	92	125	0
6	23-03-2018	Budi Cell	Garut	125	0	95	90	110	40
7	23-03-2018	WM Cell	Garut	80	0	87	50	86	16
8	23-03-2018	Yoma Cell	Tasikmalaya	135	120	100	90	95	45
...
614	30-07-2018	Nanda Cell	Tasikmalaya	125	88	95	45	75	35

3. Pembentukan kandidat 1-Itemset

Proses pembentukan kandidat 1-Itemset pada masing – masing setiap wilayah dengan jumlah minimum support yang ditentukan sebesar 35% kemudian hitung jumlah kemunculannya pada setiap transaksi

4. Pembentukan kandidat 2-Itemset

Pembentukan kandidat 2-itemset atau kombinasi dari 2-itemset data untuk semua item sehingga tidak bisa lagi dikombinasikan (disebut calon kombinasi 2-itemset atau calon F2) dengan memasang satu item dengan item lainnya. Kombinasi yang telah dibentuk kemudian hitung jumlah kemunculannya pada setiap transaksi. Selanjutnya pemangkasan atau pembuangan itemset yang memiliki nilai kurang dari minimum support 35% yang telah ditentukan. Hal ini dilakukan untuk menemukan F2 akhir.

5. Pembentukan kandidat 3-Itemset

Setelah menemukan Frekuensi 2-itemset akhir, kemudian melanjutkan untuk menemukan kembali kombinasi 3-itemset. Buat kombinasi 3-itemset untuk semua item sehingga tidak bisa lagi dikombinasikan dengan cara yang sama yaitu memasang item satu dengan item lain sehingga membentuk calon kandidat 3-itemset. Selanjutnya pemangkasan atau pembuangan itemset yang memiliki nilai kurang dari minimum support 35% yang telah ditentukan. Hal ini dilakukan untuk menemukan F3 akhir.

6. Nilai Confidence dari Frekuensi 2-Itemset (F2) dan Frekuensi 3-Itemset (F3)

Sebelum menghitung nilai confidence dilakukan terlebih dahulu pertukaran itemset. Misal suatu kombinasi pada itemset 2 yaitu $A \rightarrow B$, maka dibalik menjadi $B \rightarrow A$. contoh lainnya adalah suatu kombinasi pada itemset 3, yaitu $A,B \rightarrow C$, itemset tersebut bisa dibalik menjadi $A,C \rightarrow B$ dan $B,C \rightarrow A$. Nilai support pada masing masing itemset tersebut tetap sama, akan tetapi kemungkinan nilai confidence bisa berbeda. Hal tersebut juga agar dapat diketahui mana nilai confidence yang terbesar dari tiap itemset tersebut . Selanjutnya pemangkasan atau pembuangan itemset yang memiliki nilai kurang dari minimum confidence 80% yang telah ditentukan. Pembentukan Aturan

7. Asosiasi Final

Setelah didapat nilai Support dan Confidence untuk masing-masing kandidat lakukan perkalian antara Support dan Confidence. Setelah didapat hasil perkalian antara nilai Support dan Confidence pilihlah yang hasil perkaliannya paling besar. Hasil paling besar dari perkalian - perkalian tersebut merupakan rule yang di pakai dalam aturan Asosiasi final.

Hasil Pengujian

Berdasarkan hasil pengujian yang telah dilakukan bahwa secara keseluruhan penerapan metode association rule mining menggunakan perhitungan algoritma apriori yang diimplementasikan pada aplikasi web dapat dijalankan dengan baik dan menampilkan informasi hasil analisa aturan asosiasi yang dilakukan oleh sistem sudah sesuai dengan konsep perhitungan algoritma apriori itu sendiri. Hal ini tentunya dapat membantu supervisor untuk melakukan analisa terhadap market share penjualan produk kartu kuota internet di setiap pangsa pasar outlet.

REVIEW PAPER

1. Data transaksi penjualan produk kartu perdana kuota internet dapat dijadikan sebagai bahan acuan untuk mengetahui seberapa besar tingkat penjualan produk yang telah dipasarkan oleh beberapa operator telekomunikasi seluler.
2. Data dapat dianalisa dan dimanfaatkan menjadi sebuah informasi untuk membantu dalam melakukan pengembangan strategi pemasaran produk. Dimana tujuannya untuk menemukan aturan asosiasi kombinasi antar item produk operator telekomunikasi seluler mana saja yang paling laku terjual di wilayah penjualan Priangan Timur meliputi cluster Ciamis, Garut dan Tasikmalaya.
3. Perhitungan Algoritma Apriori pada aturan asosiasi ini dihitung melalui tiga tahap iterasi pembentukan kandidat k-itemset. Hasil analisa aturan asosiasi yang terbentuk dari perhitungan algoritma apriori dengan menentukan nilai minimum support 35% dan nilai minimum confidence 80%, menghasilkan 9 aturan asosiasi final terbaik pada cluster Ciamis, 21 aturan asosiasi final untuk cluster Tasikmalaya dan 7 aturan asosiasi final untuk cluster Garut.
4. Ketiga wilayah penjualan tersebut produk yang paling sering laku terjual dipasarkan outlet adalah produk dari operator kartu kuota internet XL dengan Telkomsel dan produk Indosat dengan Telkomsel. Hasil penelitian dapat digunakan organisasi untuk pengambilan keputusan dalam meningkatkan penjualan produk yang lebih baik.
5. Aturan asosiasi yang terbaik Cluster Ciamis ditemukan yaitu menunjukkan bahwa "Jika outlet menjual produk kartu kuota internet XL, maka ada kemungkinan outlet tersebut menjual produk kartu kuota internet Telkomsel dengan nilai support 52,40% dan nilai confidence 95,61% serta nilai akurasi perkalian support dan confidence adalah 0,5010".
6. Cluster Tasikmalaya memiliki 21 aturan asosiasi final, aturan asosiasi yang terbaik ditemukan yaitu menunjukkan bahwa "Jika outlet menjual produk kartu kuota internet Indosat, maka ada kemungkinan outlet tersebut akan menjual produk kartu kuota internet Telkomsel dengan nilai support 58,05% dan nilai confidence 92,25% serta nilai akurasi perkalian support dan confidence adalah 0,5355".
7. Terakhir 7 aturan asosiasi final untuk wilayah penjualan Garut, aturan asosiasi yang terbaik ditemukan yaitu menunjukkan bahwa "Jika outlet menjual produk kartu kuota internet XL, maka ada kemungkinan outlet tersebut akan menjual produk kartu kuota internet Telkomsel dengan nilai support 52,74% dan nilai confidence 97,25% serta nilai akurasi perkalian support dan confidence adalah 0,5129".
8. Hasil yang diperoleh dalam penelitian ini dapat digunakan untuk membantu proses pengambilan keputusan. Hasil lainnya juga, implementasi ini bisa membantu untuk mengetahui daya saing operator telekomunikasi seluler mana saja yang memiliki tingkat pangsa pasar penjualan yang terbaik di setiap wilayah Priangan Timur.

TUGAS ADVANCE DATABASE

Association Rule Review Paper



Dibuat Oleh
Aan Novrianto

Dosen Pengampu
TRI BASUKI, S.KOM. M.Eng. Ph.d

Program Pasca Sarjana
Universitas Binadarma Palembang
2020/2021

Pencarian Pola Asosiasi Keluhan Pasien Menggunakan Teknik Association Rule Mining

Ulya Anisatur Rosyidah*, Hardian Oktavianto**

Pada Jurnal berikut yang saya review adalah berkenaan asosiasi rule mining ditulis oleh Ulya Anisatur Rosida dan Hardiaan Oktavianto pada Dinas Kesehatan Kabupaten Jember.

Perkembangan dan pertumbuhan data di bidang kesehatan semakin meningkat dan bertambah, baik dari kualitas maupun kuantitas, dilihat dari sisi kualitas, perkembangan data ini mengalami perubahan dari bentuk dokumen tulis menjadi dokumen digital atau yang biasanya kita sebut dengan *file*, baik file dokumen, gambar, maupun video, sedangkan apabila dilihat dari sisi kuantitas pertumbuhan data di bidang kesehatan mengalami penambahan jumlah dari waktu ke waktu, sebagai contoh, data yang paling sering kita temui adalah data kunjungan pasien, data rekam pasien, data transaksi obat, data penyakit, sampai dengan data citra medis (hasil *USG*, hasil *CT-Scan*, hasil *Rontgen*), pada umumnya tumpukan data ini disimpan dan digunakan oleh pihak terkait untuk berbagai keperluan sesuai kebutuhan, data rekam medis biasanya digunakan sebagai sarana kontrol terhadap pasien, data transaksi obat untuk sumber informasi persediaan obat serta distribusinya, data penyakit untuk mengetahui jenis – jenis penyakit serta cara penanggulangannya dengan pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *data mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *data mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna. Fungsi *data mining* yang sering digunakan adalah untuk klasifikasi, klusterisasi, estimasi, prediksi, serta penemuan pola asosiasi. *Association rule mining* merupakan salah satu teknik dalam *data mining* yang berguna untuk menemukan pola asosiasi tersembunyi dalam suatu basis data, pola yang dimaksud disini adalah keterkaitan atau korelasi antara tiap tiap *item* yang berbeda pada setiap *record* di dalam basis data. Pola asosiasi yang ditemukan nantinya berupa *rule – rule* dengan masing – masing nilai bobot asosiasinya, *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara *itemset* X dan Y.

2. Metode Penelitian

Untuk dapat menemukan pola asosiasi keluhan pasien maka tentunya harus didapatkan dulu data yang

akan diproses, setelah pengumpulan data telah selesai maka akan dilakukan *preprocess* data kemudian

dilanjutkan dengan *data transformation*, langkah selanjutnya yaitu melakukan *association rule mining* sehingga pola – pola asosiasi keluhan pasien bisa didapatkan. Deskripsi alur penelitian bisa dilihat pada gambar 1, yang diikuti dengan penjelasan mengenai tahapan – tahapan penelitian. Pada penelitian ini dibagi menjadi 3 tahap utama yaitu : Tahap persiapan data, tahap pengolahan data, serta tahap mining dan analisa hasil keluaran.

Tahap Persiapan Data

Data yang digunakan pada penelitian ini adalah data kunjungan pasien puskesmas Ambulu pada rentang waktu Agustus 2011 sampai Januari 2012 dan diambil dari Dinas Kesehatan Kabupaten Jember. Data yang diambil termasuk data sekunder karena data ini merupakan hasil olahan Dinas Kesehatan yang berasal dari laporan bulanan tiap – tiap puskesmas di Kabupaten Jember. Data yang diperoleh berbentuk file .xls (format Microsoft Excel), jumlah record sebanyak 16.169 buah, yang mempunyai 3 atribut yaitu nomor pasien, keluhan, serta kode penyakit. Data yang diperoleh ini tidak bisa langsung digunakan dalam proses *association rule mining*, sebelumnya data harus diperiksa apakah tidak ada data yang berulang, apakah terdapat duplikasi data, atau bahkan apakah ada data yang tidak lengkap, selain itu diperiksa juga apakah dari ketiga atribut tersebut dipakai semuanya ataukah tidak. Proses tersebut akan dilakukan pada tahap selanjutnya yaitu tahap pengolahan data, dimana nantinya akan ada 2 sub proses, *preprocess*, dan *data transformation*.

Tahap Pengolahan Data

Data yang sudah diperoleh kemudian akan dipersiapkan untuk proses *association rule mining* melalui *preprocess* dan *transformation*, kedua proses ini menggunakan alat bantu yaitu *software* Microsoft Excel 2007 (*evaluation copy*). *Preprocess*, bertujuan untuk membersihkan dan memilih data, yang dimaksud dengan membersihkan data adalah memeriksa apakah ada duplikasi data, kesalahan pengetikan, serta data yang tidak lengkap, sedangkan yang dimaksud dengan pemilihan data adalah memilih variabel data yang diperlukan saja. Setelah melalui *preprocess* maka hanya tersedia 670 data saja yang siap untuk proses selanjutnya, hal ini disebabkan karena pada variabel “keluhan” banyak berisi keluhan tunggal, keluhan berupa rujukan, dan keterangan yang kosong. *Transformation*,

bertujuan untuk merubah format data yang sudah ada menjadi format yang bisa diproses, format yang dipakai adalah format biner, dimana pada satu transaksi berisi angka 0 (nol) atau 1 (satu), dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”.

Tahap Mining dan Analisis

Pencarian pola asosiasi diantara keluhan pasien akan dikerjakan dengan dua algoritma, apriori dan FP-*Growth*, dengan batasan nilai *support* dan *confidence* yang diinputkan oleh *user*, setelah pola asosiasi didapatkan maka tahap yang terakhir adalah melakukan analisa terhadap *rule – rule* yang terbentuk, sehingga diharapkan akan ditemukan informasi baru yang menarik dan berguna. Berdasarkan dataset yang diberikan *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara *itemset* X dan Y.

3. Hasil dan Analisis

Pada bab ini akan disampaikan tentang hasil penelitian serta analisis dari hasil penelitian tersebut. Bab ini dibagi menjadi 2 sub bab, yang menjelaskan tentang dua tahapan penelitian yaitu tahap pengolahan data dan tahap mining dan analisis.

3.1. Tahap Pengolahan Data

Tahap pengolahan data ini terbagi menjadi 2 sub proses yaitu *preprocess* dan *transformation*. Tahap *preprocess*, bertujuan untuk membersihkan dan memilih data, yang dimaksud dengan membersihkan data adalah memeriksa apakah ada duplikasi data, kesalahan pengetikan, serta data yang tidak lengkap, sedangkan yang dimaksud dengan pemilihan data adalah memilih variabel data yang diperlukan saja. Tahap *transformation* merubah format data yang sudah ada menjadi format yang bisa diproses, format yang dipakai adalah format biner, dimana pada satu transaksi berisi angka 0 (nol) atau 1 (satu), dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”.

PUSING	PANAS	FLU	BATUK	SESAK	PILEK	MUAL
1	0	0	1	0	0	0
0	1	0	0	0	1	0
1	0	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	0	0	0
1	0	1	0	0	0	0
0	1	0	1	0	0	0
0	1	1	0	0	0	0
0	1	0	1	0	0	0
0	1	1	0	0	0	0
0	0	1	1	0	0	0
0	1	1	0	0	0	0
1	0	1	0	0	0	0
1	1	0	0	0	0	0
0	0	1	0	1	0	0

Pada gambar diatas merupakan potongan hasil dari proses pengolahan data, dimana 0 (nol) merepresentasikan “tidak ada”, sedangkan 1 (satu) merepresentasikan “ada”, jadi dari gambar diatas dapat diartikan bahwa pada tiap baris mewakili pasien, baris pertama, si pasien mengalami pusing dan batuk, sedangkan pada baris kedua pasien mengalami panas dan pilek, begitu juga seterusnya.

3.2. Tahap Mining dan Analisis

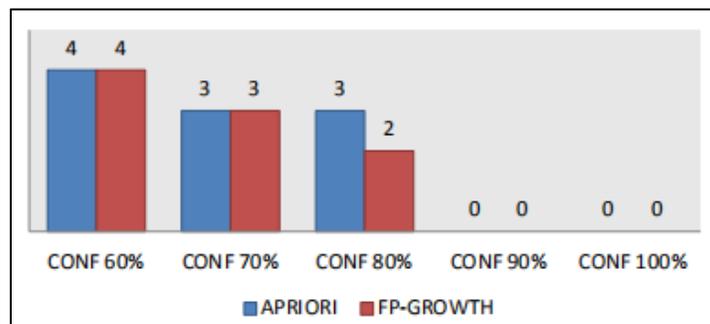
Skenario uji coba memakai 670 record data, dengan 5 nilai *confidence* yang berbeda – beda, yaitu 100%, 90%, 80%, 70%, dan 60%, dengan nilai support yang sama yaitu 1%. Alasan mengapa uji coba ini hanya dilakukan dengan nilai *confidence* yang berbeda – beda adalah karena biasanya pola asosiasi yang menarik ada pada kecenderungan pasangan itemset yang sering terjadi (definisi dari *confidence*). Uji coba nantinya akan dilakukan bergantian, yaitu uji coba terlebih dahulu menggunakan algoritma apriori kemudian menggunakan algoritma FP-*Growth*. Tabel menunjukkan hasil uji dengan variasi *confidence* yang berbeda – beda, secara konsep, rule –rule yang dihasilkan juga ditampilkan secara berbeda oleh kedua algoritma.

Tabel. 1 Hasil Uji Coba

Uji ke-	Confidence	Hasil Apriori	Hasil FP-Growth
---------	------------	---------------	-----------------

1	60%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=1 44 ==> PANAS=1 37 conf:(0.84) 2. FLU=1 344 ==> PANAS=1 288 conf:(0.84) 3. SESAK=1 45 ==> BATUK=1 36 conf:(0.8) 4. BATUK=1 259 ==> PANAS=1 168 conf:(0.65)	FPGrowth found 4 rules (displaying top 4) 1. [PILEK=1]: 44 ==> [PANAS=1]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=1]: 344 ==> [PANAS=1]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32) 3. [SESAK=1]: 45 ==> [BATUK=1]: 36 <conf:(0.8)> lift:(2.07) lev:(0.03) conv:(2.76) 4. [BATUK=1]: 259 ==> [PANAS=1]: 168 <conf:(0.65)> lift:(0.83) lev:(-0.05) conv:(0.62)
2	70%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=1 44 ==> PANAS=1 37 conf:(0.84) 2. FLU=1 344 ==> PANAS=1 288 conf:(0.84) 3. SESAK=1 45 ==> BATUK=1 36 conf:(0.8)	FPGrowth found 3 rules (displaying top 3) 1. [PILEK=1]: 44 ==> [PANAS=1]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=1]: 344 ==> [PANAS=1]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32) 3. [SESAK=1]: 45 ==> [BATUK=1]: 36 <conf:(0.8)> lift:(2.07) lev:(0.03) conv:(2.76)
3	80%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found: 1. PILEK=1 44 ==> PANAS=1 37 conf:(0.84) 2. FLU=1 344 ==> PANAS=1 288 conf:(0.84) 3. SESAK=1 45 ==> BATUK=1 36 conf:(0.8)	FPGrowth found 2 rules (displaying top 2) 1. [PILEK=1]: 44 ==> [PANAS=1]: 37 <conf:(0.84)> lift:(1.08) lev:(0) conv:(1.21) 2. [FLU=1]: 344 ==> [PANAS=1]: 288 <conf:(0.84)> lift:(1.07) lev:(0.03) conv:(1.32)
4	90%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found:	No rules found!
5	100%	Generated sets of large itemsets: Size of set of large itemsets L(1): 6 Size of set of large itemsets L(2): 8 Best rules found:	No rules found!

Pada gambar diatas, disajikan grafik perbandingan secara persentase tentang jumlah rule yang dihasilkan berdasarkan skenario uji coba. Perbedaan hasil uji coba terletak pada jumlah *rule* asosiasi yang ditemukan, hal ini terjadi pada saat uji coba yang dilakukan menggunakan *confidence* sebesar 80%, dengan menggunakan algoritma apriori ditemukan 3 buah *rule* asosiasi, sedangkan ketika digunakan algoritma FP-*Growth* ditemukan 2 buah *rule* asosiasi.



Review Jurnal diatas :

Pada Jurnal ini, menggunakan 2 metode association rule mining yakni algoritma apriori menggunakan prinsip apriori dalam pencarian frequent itemset, yaitu semua subset yang tidak kosong dari sebuah frequent itemset pasti juga akan merupakan frequent, begitu juga dengan semua superset yang tidak kosong dari sebuah non-frequent itemset pasti juga akan merupakan non-frequent. Sedangkan algoritma FP-Growth menggunakan konsep pembangunan tree dalam pencarian frequent itemsets yang memungkinkan dapat secara langsung membentuk frequent itemset dengan menerapkan prinsip divide and conquer. Baik algoritma apriori dan algoritma FP_Growth menghasilkan output yang sama. Perbedaan hasil uji coba terletak pada jumlah rule asosiasi yang ditemukan, dengan menggunakan algoritma apriori ditemukan 3 buah rule asosiasi, sedangkan ketika digunakan algoritma FP-Growth ditemukan 2 buah rule asosiasi, hal ini terjadi pada saat uji coba yang dilakukan menggunakan confidence sebesar 80%.

Penelitian ini bisa dikembangkan dengan menggunakan kombinasi algoritma yang lain dalam bidang association rule mining, selain itu bisa juga dilakukan analisa tentang perbandingan waktu proses ketika melakukan pencarian rule asosiasi.

.....

NAMA : AHMAD ALI MA'MUN
NIM : 202420037

Association Rule Mining

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien, yaitu analisis pola frekuensi tinggi (frequent pattern mining) [5].

Secara umum association rule mempunyai bentuk : LHS => RHS dimana LHS dan RHS tersebut adalah himpunan item; jika setiap item-item dalam LHS terdapat dalam transaksi maka item-item dalam RHS juga terdapat dalam transaksi.

Aturan asosiasi biasanya dinyatakan dalam bentuk[5]:

$$\{A,B\} \Rightarrow \{C\} \text{ (support = 10\%, confidence = 50\%)}$$

Support

Support dari suatu *association rule* adalah presentasi kombinasi *item* tersebut dalam database, dimana jika mempunyai *item* A dan *item* B maka support adalah proporsi dari transaksi dalam database yang mengandung A dan B. Rumus untuk menghitung nilai support dari dua *item* tersebut adalah sebagai berikut[5][6]:

$$\text{Support}(A, B) = P(A \cap B)$$

$$\text{Support}(A, B) = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi}}$$

2.3 Confidence

Confidence dari *association rule* adalah ukuran ketepatan suatu *rule*, yaitu presentasi transaksi dalam database yang mengandung A dan mengandung B. Dengan adanya *confidence* kita dapat mengukur kuatnya hubungan antar-*item* dalam *association rule*. Rumus untuk menghitung nilai *confidence* dari dua *item* tersebut adalah sebagai berikut[5][6]:

$$\text{Confidence} = P(B | A)$$

$$\text{Confidence} = \frac{\sum \text{transaksi mengandung A dan B}}{\sum \text{transaksi mengandung A}}$$

Algoritma Apriori

Algoritma apriori adalah langkah untuk proses menemukan *frequent-itemset* dengan melakukan iterasi pada data. Dimana *itemset* adalah himpunan *item-item* yang berada di dalam himpunan yang diolah oleh sistem, sedangkan *frequent-itemset* menunjukkan *itemset* yang memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan (ϕ). Pada iterasi ke-*k*, semua *itemset* yang ditemukan yang memiliki *k* *item* disebut *k-itemset*. Setiap iterasi terdiri dari dua tahap yaitu pembangkitan kandidat dan pembangkitan *rule*.

Pada tahap pembangkitan kandidat (*candidate generation*) dimana himpunan semua *frequent(k - 1)-itemset* yang digunakan pada pass ke-*(k-1)* digunakan untuk membangkitkan kandidat *itemset*

. Prosedur pembangkitan kandidat menjamin bahwa adalah superset dari himpunan. Kemudian di-scan dalam tahap perhitungan *support* (*support counting*). Pada akhir pass diperiksa untuk menentukan kandidat mana yang muncul, menghasilkan. Perhitungan *support* berakhir ketika atau kosong.

Pada tahap membangkitkan *rule*, akan dibangkitkan lebih dahulu *candidate rule*. *Candidate rule* berisi semua kemungkinan *rule* yang memiliki *support > minimum support* karena inputan *candidate rule* adalah *frequent-itemset*. Kemudian *candidate rule* akan di-join dengan table F untuk menemukan *support antecedent*. *Confidence rule* dihitung dengan cara membandingkan *support rule* dengan *support antecedent rule*. Hanya *rule* yang mempunyai *confidence > minimum confidence* yang disimpan dalam table *rule* (table R)[6].

Lift Ratio

Lift ratio adalah suatu ukuran untuk mengetahui kekuatan aturan asosiasi (*association rule*) yang telah

terbentuk. Nilai *lift ratio* biasanya digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid. Untuk menghitung *lift ratio* digunakan rumus sebagai berikut:

$$\text{Lift Ratio} = \frac{\text{Confidence (A, B)}}{\text{Benchmark Confidence (A,B)}}$$

Untuk mendapatkan nilai benchmark confidence sendiri dapat dihitung menggunakan rumus sebagai berikut:

$$\text{Benchmark Confidence} = \frac{N_c}{N}$$

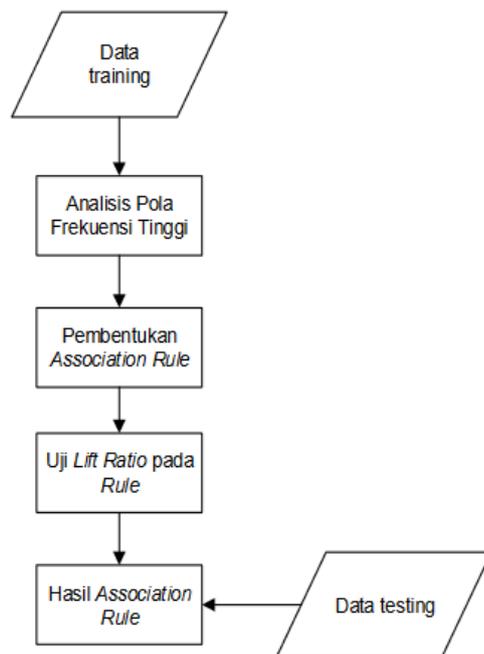
Keterangan:

- N_c = jumlah transaksi dengan item yang menjadi consequent

- N = jumlah transaksi basis data

Analisis dan Perancangan Sistem

Gambar 1. Perancangan Umum Sistem



Skenario keseluruhan sistem :

1. Data training dan data testing adalah data yang sudah melalui proses preprocessing secara manual.
2. Dengan menggunakan algoritma apriori akan dilakukan training terhadap itemset yang terdapat di dalam data klimatologi yaitu dengan membentuk kandidat-kandidat itemset atau biasa disebut dengan k-itemset.
3. Setiap k-itemset yang terbentuk sebelumnya kemudian dilakukan perhitungan nilai support untuk menghitung jumlah transaksi yang memuat semua item di dalam k-itemset tersebut. K-itemset yang nilai supportnya lebih tinggi dari nilai minimum support kemudian dijadikan kandidat sebagai pembentukan aturan asosiasi (association rule).
4. Dilakukan perhitungan *confidence* dari tiap k-itemset untuk menentukan apakah kandidat tersebut dapat dijadikan sebagai aturan asosiasi (association rule) atau tidak.
5. Rule yang terbentuk di evaluasi kekuatannya dengan cara uji lift ratio. Nilai lift ratio dari sebuah rule didapatkan melalui perbandingan *confidence* rule tersebut dengan *benchmark confidence*.
6. Selanjutnya dikeluarkan hasil akhir aturan asosiasi dengan format jika (k-itemset) maka hujan atau tidak hujan. Nilai k-itemset tergantung berdasarkan k-itemset yang terbentuk.

4. Pengujian

Tujuan Pengujian

Tujuan dari pengujian yang dilakukan adalah sebagai berikut ini:

1. Menganalisa pengaruh *minimum support* terhadap jumlah *frequent itemset* yang dibangkitkan algoritma apriori.
2. Menganalisa pengaruh *minimum confidence* terhadap jumlah *rule* yang dihasilkan berdasarkan *frequent itemset*.
3. Menganalisa pengaruh *minimum support* dan *minimum confidence* terhadap akurasi dari sistem yang telah dibangun.

4. Mengetahui kekuatan dari tiap *rule* yang dihasilkan.

Dataset

Dataset yang digunakan oleh sistem yang dibangun berupa *file* bertipe excel yang berisikan data cuaca yang diambil dari BMKG stasiun geofisika kelas 1 Bandung. Penginputan data ke *file* excel dilakukan secara manual karena data yang di dapat dari BMKG merupakan data fisik. Lalu data akan masuk ke proses *data preprocessing* yang akan menghasilkan dataset yang akan dipakai oleh sistem.

5. Analisis Hasil Pengujian

Analisis Hasil Pengaruh *Minimum Support* dan *Minimum Confidence* terhadap *Frequent Itemset*

Tabel 1. *frequent itemset*

mincof\minsup	10%	20%	30%	40%	50%
50%	73	40	19	17	7
60%	73	40	19	17	7
70%	73	40	19	17	7

Berdasarkan tabel 1, dapat dilihat jumlah *frequent itemset* terbanyak dibangkitkan oleh *minimum support* terkecil dalam pengujian ini yaitu 10%, sedangkan *frequent itemset* paling sedikit dibangkitkan oleh *minimum support* sebesar 50%. Hal ini dikarenakan nilai *minimum support* adalah nilai acuan dalam perhitungan sebuah *itemset*. Hanya *itemset* yang nilai supportnya sama atau lebih besar dari nilai *minimum support* saja yang dijadikan sebagai *frequent itemset*. Sedangkan *minimum confidence* tidak memiliki pengaruh dalam pembangkitan *frequent itemset* karena tidak terjadi perhitungan *confidence* pada tahap ini. Maka dapat disimpulkan bahwa semakin tinggi *minimum support* yang digunakan maka akan semakin sedikit jumlah *frequent itemset* yang dibangkitkan.

Analisis Hasil Pengaruh *Minimum Support* *Minimum* dan *Confidence* terhadap *Rule*

Tabel 2. Jumlah *rule*

mincof\minsup	1%	2%	3%	4%	5%
50%	9	5	2	2	0
60%	3	1	0	0	0
70%	2	0	0	0	0

Berdasarkan tabel 2, dapat dilihat jumlah aturan asosiatif yang terbentuk terbanyak dihasilkan oleh *minimum support* 10% dan *minimum confidence* 50% yaitu sebanyak 9 *rule*. Pada setiap pertambahan nilai *minimum support* dan *minimum confidence* yang diujikan terlihat *rule* yang terbentuk semakin berkurang. Penurunan ini dipengaruhi oleh jumlah *frequent itemset* yang dibangkitkan pada pengujian sebelumnya. Pada *minimum support* 50% dapat dilihat tidak terbentuk satu pun aturan asosiatif, hal ini diakibatkan nilai *confidence* dari *frequent itemset* yang dibangkitkan tidak ada yang memenuhi syarat *minimum confidence* yang diujikan. Oleh karena itu, dapat dilihat bahwa *minimum confidence* berpengaruh pada pembentukan *rule*. Pada pengujian ini dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka *rule* yang dihasilkan akan semakin sedikit.

Analisis Hasil Pengaruh *Minimum Support* *Minimum* dan *Confidence* terhadap akurasi

Tabel 3. Akurasi data training

mincof\minsup	10%	20%	30%	40%	50%
50%	76.78	76.71	49.73	49.73	0
60%	42.67	26.71	0	0	0
70%	15.96	0	0	0	0

Berdasarkan tabel 3 di atas, akurasi tertinggi diperoleh sebesar 76.78% dengan *minimum support* 10% dan *minimum confidence* 50%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya tidak dapat dihitung besarnya akurasi karena tidak ada rule yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data training diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data training. Dalam pengujian akurasi data training dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 76.71%.

Tabel 4. Akuarasi data testing

mincof\minsu p	10%	20%	30%	40%	50%
50%	75.89	75.89	50.41	50.41	0
60%	30.69	24.11	0	0	0
70%	6.58	0	0	0	0

Berdasarkan tabel 4 di atas, akurasi tertinggi diperoleh sebesar 75.89% terdapat pada dua *minimum support* berbeda yaitu *minimum support* 10% dan *minimum support* 20% dengan *minimum confidence* 50%. Akurasi berikutnya juga diperoleh pada dua *minimum support* yang berbeda yaitu *minimum support* 30% dan *minimum support* 40% dengan *minimum confidence* 50% sebesar 50.41%. Sedangkan pada *minimum support* 50%, dan yang bernilai akurasi 0 lainnya tidak dapat dihitung besarnya akurasi karena tidak ada rule yang terbentuk. Dari grafik diatas dapat disimpulkan semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka akurasi yang dihasilkan akan semakin kecil. Hal ini terjadi karena *rule-rule* yang dihasilkan semakin kuat keterkaitan antar *item*-nya. Sehingga pada saat data testing diuji kedalam *rule* dan dicocokkan kembali dengan data nilai kebenarannya semakin berkurang. Nilai tersebut di prosentasekan dan dijadikan nilai akurasi dari data testing. Dalam pengujian akurasi data testing dapat dilihat sistem berfungsi optimal pada *minimum support* sebesar 20% dan *minimum confidence* sebesar 50% yang menghasilkan akurasi sebesar 75.89%.

Analisis Hasil Lift Ratio terhadap Hasil Rule

Tabel 5. Lift ratio tiap rule

mini mum supp ort	minim um confid ence	aturan asosiatif	confidence	lift ratio
10%	50%	Agak Panas(temperature) E(windDirection) -> tidak hujan	73.08	1.46
10%	50%	E(windDirection) - -> tidak hujan	72.81	1.45
10%	50%	W(windDirection) -> hujan	62.9	1.26
10%	50%	Agak Basah(humidity) Agak Kencang(windSpee d) -> tidak hujan	53.41	1.06
10%	50%	Agak Kencang(windSpee d) -> tidak hujan	53.38	1.06
10%	50%	Agak Panas(temperature) Pelan(windSpeed) -> tidak hujan	51.94	1.03

10%	50%	<i>Pelan(windSpeed) - -> tidak hujan</i>	50.88	1.01
10%	50%	<i>Agak Basah(humidity) -> tidak hujan</i>	50.66	1.01
10%	50%	<i>Agak Panas(temperature) -> tidak hujan</i>	50.62	1.01
20%	50%	<i>W(windDirection) -> hujan</i>	62.9	1.26
20%	50%	<i>Agak Panas(temperature) Pelan(windSpeed) -> tidak hujan</i>	51.94	1.03
20%	50%	<i>Pelan(windSpeed) - -> tidak hujan</i>	50.88	1.01
20%	50%	<i>Agak Basah(humidity) -> tidak hujan</i>	50.66	1.01
20%	50%	<i>Agak Panas(temperature) -> tidak hujan</i>	50.62	1.01
30%	50%	<i>Agak Basah(humidity) -> tidak hujan</i>	50.66	1.01
30%	50%	<i>Agak Panas(temperature) -> tidak hujan</i>	50.62	1.01
40%	50%	<i>Agak Basah(humidity) -> tidak hujan</i>	50.66	1.01
40%	50%	<i>Agak Panas(temperature) -> tidak hujan</i>	50.62	1.01

Berdasarkan tabel 5 diatas *lift ratio* untuk semua *rule* yang berhasil dibentuk memiliki nilai lebih besar dari 1 (*lift ratio* >1). Hal ini menunjukkan bahwa semua *rule* tersebut bersifat kuat dan valid untuk digunakan sebagai acuan dalam memprediksi Hujan.

KESIMPULAN

Berdasarkan pengujian yang dilakukan diatas, dapat disimpulkan bahwa :

1. Metode *association rule mining* dengan menggunakan algoritma apriori dapat diterapkan pada sistem simulasi prediksi hujan.
2. Semakin tinggi *minimum support* dan *minimum confidence* yang digunakan maka semakin sedikit jumlah *frequent itemset* dan *rule* yang terbentuk serta akurasi semakin berkurang.
3. Semua rule yang dihasilkan pada penelitian ini memiliki nilai *lift ratio* lebih dari 1.00 sehingga dapat digunakan sebagai acuan dalam memprediksi hujan.

SARAN

Saran yang diperlukan untuk pengembangan sistem lebih lanjut adalah sebagai berikut :

1. Kategori yang dilakukan pada penelitian ini adalah hujan dan tidak hujan. Untuk kedepannya dapat dilakukan pengkategorian yang lebih luas sehingga dapat diketahui intensitas hujan yang turun.
2. Untuk selanjutnya dapat menggunakan dataset yang jauh lebih besar dari sekarang, misalnya data yang diolah sebanyak 10 tahun keatas.
3. Diharapkan untuk penelitian selanjutnya dapat menerapkan algoritma pembangkitan *frequent itemset* yang berbeda, seperti *FP-Growth*.

ADVANCE DATABASE
“REVIEW PAPER TENTANG ASSOCIATION RULE”

TUGAS 2

OLEH :

202420017 Ainun Hilaliyyah

DOSEN PENGAMPUH :

Bpk. Tri Basuki Kurniawan, S.Kom., M.eng.,Ph.D

Program Studi IT Infrastructure



UNIVERSITAS BINA DARMA
PALEMBANG
2020

PENERAPAN *ASSOCIATION RULE MINING* UNTUK ANALISIS PENEMPATAN TATA LETAK BUKU DI PERPUSTAKAAN MENGUNAKAN ALGORITMA APRIORI

I. PENDAHULUAN

Data mining adalah suatu teknik yang digunakan untuk mengelola himpunan data menggunakan metode *data mining* untuk menghasilkan informasi atau pengetahuan. Perpustakaan adalah institusi pengelola koleksi karya tulis, karya cetak, dan atau karya rekam secara profesional dengan sistem yang baku guna memenuhi kebutuhan pendidikan, penelitian, pelestarian, informasi, dan rekreasi para pemustaka [1].

Perpustakaan Soeman HS merupakan perpustakaan terbesar di Provinsi Riau dan dikelola oleh Dinas Perpustakaan dan Kearsipan Provinsi Riau. Perpustakaan Soeman HS menyediakan bahan bacaan seperti buku, majalah umum, dan kamus. Jumlah koleksi buku yang ada di perpustakaan Soeman HS 60.583 judul dan 335.745 eksemplar buku dan dikelompokkan menjadi 10 kategori buku.

Pada saat ini proses penempatan letak buku yang dilakukan di perpustakaan disusun dalam rak buku berdasarkan kategori buku tersebut, dalam penempatan letak buku belum diatur dengan melihat tingkat keseringan pengunjung membaca atau meminjam buku tersebut. Selain itu, pengunjung kesulitan dalam mencari keberadaan buku yang sering dibaca atau dipinjam karena letak buku yang dibutuhkan berada dirak yang berbeda-beda.

Untuk menyelesaikan permasalahan yang ada, menggunakan metode *Association Rule Mining*. Algoritma yang digunakan dalam penelitian ini adalah algoritma apriori. Algoritma apriori merupakan salah satu yang paling populer dalam pendekatan *data mining* untuk menemukan *frequent itemset* untuk aturan asosiasi *boolean*.

II. LANDASAN TEORI

2.1 KDD (*Knowledge Discovery in Database*)

Knowledge Discovery in Database (KDD) didefinisikan sebagai ekstraksi informasi potensial. Proses *Knowledge Discovery in Database* melibatkan hasil proses data mining (proses pengekstrak kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. KDD (*Knowledge Discovery in Database*) terbagi menjadi tiga proses utama, yaitu:

1. *Preprocessing*
2. *Data mining*
3. *Post processing*

2.2 Data Mining

Data Mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Data mining* menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

2.3 Association Rule Mining

Association rules mining adalah suatu metode yang digunakan untuk mencari hubungan antar item suatu dataset yang telah ditentukan. *Association rules* mining mencari dan menemukan hubungan antar item yang ada pada suatu dataset. Penerapan data mining dengan aturan asosiasi bertujuan menemukan informasi item-item yang saling berhubungan dalam bentuk aturan/rule. Perhitungan nilai *support* dan *confidence* dapat dijelaskan sebagai berikut:

1. Analisa Pola Frekuensi Tinggi *Support*

Mencari Nilai *support* sebuah item diperoleh dengan menggunakan rumus berikut:

$$Support(A) = \frac{\sum \text{Jumlah transaksi yang mengandung A}}{\sum \text{Total Transaksi}} \times 100$$

Nilai *support* 2 item diperoleh dengan rumus:

$$Support(A,B) = P(A \cap B)$$

$$Support(A,B) = \frac{\sum \text{Jumlah transaksi yang mengandung A dan B}}{\sum \text{Total Transaksi}} \times 100$$

2. Pembentukan Aturan Asosiasi *Confidence*

Pembentukan aturan asosiasi *confidence* diperoleh dengan rumus:

$$Confidence(A,B) = \frac{\sum \text{Jumlah transaksi yang mengandung A dan B}}{\sum \text{Transaksi mengandung A}} \times 100$$

2.4 Algoritma Apriori

Algoritma Apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan *frequent itemset* untuk aturan asosiasi boolean. *Algoritma apriori* termasuk jenis aturan asosiasi pada *Data Mining*. Untuk penerapan *Algoritma Apriori*, secara umum dibutuhkan struktur data untuk menyimpan *candidate frequent itemset* untuk suatu iterasi ke k dan untuk menyimpan *frequent itemset* yang dihasilkan.

III. PEMBAHASAN

Berdasarkan hasil proses Algoritma Apriori pada data transaksi peminjaman buku di Perpustakaan Soemanh Hs Provinsi Riau dapat disimpulkan.

1. Penelitian ini berhasil menganalisa data transaksi peminjaman buku untuk mengetahui buku-buku yang sering dipinjam secara bersamaan menggunakan algoritma Apriori. Hasil dari analisa pada semua data transaksi peminjaman buku dari tahun 2016-2018 dapat disimpulkan bahwa jika pengunjung meminjam buku dengan kategori buku agama maka akan meminjam buku dengan kategori buku ilmu sosial aturan ini memiliki nilai *support* 11,71% dan *confidence* 41,43% dan jika pengunjung meminjam buku dengan kategori buku teknologi maka akan meminjam buku dengan kategori buku ilmu sosial aturan ini memiliki nilai *support* 13,8% dan *confidence* 40,75%.
2. Hasil dari analisa data transaksi peminjaman buku perbulan dari tahun 2016- 2018 dapat disimpulkan bahwa jika pengunjung meminjam buku dengan kategori buku agama maka akan meminjam buku dengan kategori buku ilmu sosial aturan ini memiliki nilai *support* 9,10% dan *confidence* 44,61%. Selain itu jika pengunjung meminjam buku dengan kategori buku filsafat maka akan meminjam buku dengan kategori buku ilmu sosial aturan ini memiliki nilai *support* 12% dan *confidence* 63%. Dan jika pengunjung meminjam buku dengan kategori buku teknologi maka akan meminjam buku dengan kategori buku ilmu sosial aturan ini memiliki nilai 12,53% dan *confidence* 60%.
3. Dengan hasil analisa pada semua data dan perbulan dari tahun 2016-2018 didapatkan buku dengan kategori buku agama, teknologi, filsafat dan ilmu sosial paling sering dipinjam secara bersamaan. Dengan hasil ini disarankan pada pihak pustakawan perpustakaan Soeman Hs untuk menempatkan letak kategori buku tersebut pada rak yang berdekatan. Selain itu pihak pustakawan bisa memberikan rekomendasi kepada pengunjung yang meminjam buku dengan kategori buku agama, filsafat, teknologi agar meminjam juga meminjam buku dengan kategori buku ilmu sosial.

IV. REVIEW

Algoritma apriori merupakan salah satu algoritma klasik data mining. Algoritma apriori digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset.

Dengan adanya algoritma apriori, pengelola perpustakaan dapat mengetahui pola dalam penempatan letak buku yang diatur dengan melihat tingkat keseringan pengunjung membaca atau meminjam buku tersebut. Dan mempermudah pengunjung dalam mencari keberadaan buku yang sering dibaca atau dipinjam karena letak buku yang dibutuhkan berada dirak yang telah disusun.

Nama : Andry Meylani

NIM : 202420009

JUDUL : An Efficient Algorithm for Disease Prediction with Multi Dimensional Data

PENULIS : Smitha.T, dan V.Sundaram,

TAHUN : 2013

**JURNAL & VOL : International Journal of Computer Applications (0975 – 8887) Volume 63– No.9,
Feb 2013**

Abstrak :

Tujuan utama dari penelitian ini adalah membuat algoritma yang cepat, mudah dan efisien untuk prediksi penyakit, dengan error rate yang rendah dan dapat diaplikasikan dengan set data yang besar dan menunjukkan pola yang layak (masuk akal) dengan variable yang mempengaruhi.

Untuk mengidentifikasi penyakit dan prediksi di data mining, sebuah algoritma hybrid dibangun. Algoritma DIP yang menggabungkan decision tree dan rule asosiasi digunakan untuk prediksi kesempatan dari beberapa serangan penyakit pada beberapa fakta-fakta. Ini juga menunjukkan hubungan antara parameter yang berbeda untuk prediksi. Untuk mengimplementasikan algoritma ini menggunakan software VB.Net yang juga dikembangkan.

1. INTRODUCTION

Data mining bertujuan untuk mengekstrak informasi dari data set dan mengubahnya ke dalam struktur yang bisa dipahami untuk kegunaan selanjutnya.(1)

Tujuan dari teknik ini juga menemukan pola yang sebelumnya belum diketahui dan menggunakannya untuk menyelesaikan sejumlah masalah.(9)

Tujuan dari seseorang menggunakan data mining adalah agar dapat memprediksi pola atau sifat dari sesuatu yang bisa dianalisis, sehingga bisa membuat strategi keputusan yang mengantarkannya untuk meraih tujuannya.

2. LITERATURE REVIEW

Elham menjelaskan penerapan baru dari penggunaan data mining untuk manajemen komplain customer menggunakan penggalian rule asosiasi. Data komplain masyarakat di Tehran dianalisis. Hasilnya menunjukkan ada reasi yang kuat antara gender – pendidikan dan pola kejadian komplain. (1)

K.Srinivas pada penelitiannya menjelaskan kegunaan potensial dari klasifikasi berdasarkan teknik data mining seperti rule based, decision tree, naïve bayes dan ANN untuk data healthcare yang sangat besar menyediakan relasi yang signifikan antara factor medis berhubungan dengan penyakit jantung(2).

Sunita Soni,dkk(3) menganalisa masalah mendesak dan menyimpulkan algoritma yang berbeda dari data mining yang digunakan pada prediksi medis yang didiskusikan. Fokus pada penggunaan algoritma yang berbeda untuk beberapa atribut target untuk kecerdasan dan prediksi serangan jantung yang efektif menggunakan data mining. Untuk prediksi serangan jantung signifikan dengan 15 atribut dan dengan basic teknik data mining seperti penerapan time series,clustering dan rule asosiasi, penerapan soft computing dapat juga disatukan. Hasil dari prediksi teknik data mining pada dataset yang sama menunjukkan decision tree dan klasifikasi Bayesian yang memiliki nilai akurasi yang sama tapi metode prediksi seperti KNN, NN, klasifikasi berdasar clustering tidak menunjukkan performa yang bagus. Kesimpulan kedua bahwa akurasi decision tree dan klasifikasi Bayesian ke depan dapat diperbaiki setelah menerapkan algoritma genetika untuk mengurangi ukuran data yang actual untuk mendapatkan subset yang optimal dari atribut yang cukup untuk prediksi sakit jantung (3).

Shelly Gupta menyimpulkan beberapa review dan artikel teknikal pada diagnosis kanker jantung dan prognosis. Dia menunjukkan gambaran tentang penelitian yang menggunakan teknik data mining untuk memperbaiki diagnosis dan ramalan (prognosis) kanker jantung (4). Penelitian tersebut

Menemukan bahwa akurasi untuk analisis diagnosis bervariasi diterapkan teknik klasifikasi data mining dapat sangat diterima dan dapat membantu profesional media dalam pembuatan keputusan untuk diagnosis awal dan menghindari biopsy. Masalah "ramalan" adalah pada analisa dengan ANN dan akurasinya lebih tinggi dibanding teknik klasifikasi lain yang diterapkan pada hal yang sama. Tetapi model yang efisien juga disediakan untuk masalah "ramalan" seperti pewarisan fitur terbaik dari model yang didefinisikan. Pada kedua kasus dapat dikatakan bahwa model terbaik dapat diperoleh setelah membangun beberapa tipe model yang berbeda, atau dengan mencoba teknologi dan algoritma yang berbeda.

Nadiami, dkk (5) mendeteksi kerasnya serangan di dalam dataset berdasarkan kddcup99 dataset dari MIT Lincoln Laboratory. Marek (6) telah menunjukkan sebuah algoritma evolusi baru (EA) untuk induksi dari decision tree yang dicampur. Pada node yang bukan terminal dari mixed tree, tipe berbeda dari uji dapat diletakkan BE, berjarak dari uji tipe yang inequality ke uji yang tak langsung pada pesawat yang hiperaktif terpisah. Kontrasnya pada metode klasikal top down, system yang diusulkan mencari tree yang optimal pada cara global yang belajar struktur tree dan menemukan uji pada 1 run dari EA. Khususnya pada operator genetic yang dikembangkan, yang membolehkan system berubah bagian dari tree, membangkitkan sub tree yang baru. Skema aplikasi mutasi yang diinformasikan dikenalkan dan jumlah dari modifikasi yang tidak menguntungkan direduksi. Semua kerjanya menggunakan algoritma yang berbeda untuk prediksi.

3. METODOLOGI

Data mining dapat dipakai pada penerapan yang berbeda dan membangun model yang berbeda tergantung pada tipe data yang terlibat dan tujuannya. Penelitian ini berdasar pada algoritma data mining yang berbeda pada analisis data multi dimensional. Model yang sekarang digunakan pada data mining prediktif meliputi rule asosiasi, metode clustering, decision tree, rule klasifikasi dan perangkat mining statistika.

Untuk identifikasi penyakit dan prediksi pada data mining membuat sebuah algoritma hybrid yang baru. Algoritma identifikasi dan prediksi penyakit (DIP) yang mengombinasikan decision tree dan rule asosiasi untuk memprediksi kesempatan dari beberapa penyakit menyerang beberapa fakta yang ada. Hal ini juga menunjukkan adanya hubungan antara parameter yang berbeda untuk prediksi. Untuk mengimplementasikannya menggunakan VB.NET.

Pada level konseptual, penggunaan decision tree membuat model data mining yang non-linear untuk variable yang berkorelasi. Untuk melintang ke decision tree dari root ke leaves rule prediksi dapat langsung diperoleh. Ketegantungan logika diantara beberapa atribut dari entitas menggunakan rule asosiasi dari prinsip apriori dibangun. Untuk mengukur confidence dan support kekuatan asosiasi dapat diukur. Menggunakan threshold minimum confidence dan support, algoritma rule-mining mengidentifikasi semua asosiasi yang memuaskan dengan parameter yang ditentukan dan menemukan ketergantungan di antara perbedaan atribut dari entitas yang sama. Kita dapat mengaplikasikan rule dan data ke teknik statistika pada analisis cluster untuk mengekstrak semua kemungkinan cluster dari data yang belum diberi label/nama. Hasilnya dapat ditunjukkan pada grafik untuk analisis.

Fase 1 : membuat DM model non linear untuk parameter yang berhubungan menggunakan DT dan mengembangkan rule prediksi. Ada T sebagai training data set dengan label kelas $\{c_1, c_2, \dots, c_k\}$ dan X adalah atribut yang bukan kelas dari T. Dari daftar atribut dari X w.r.t T dan diurutkan pada daftar atribut dan menggunakan analisis relevansi untuk mengurangi atribut menggunakan threshold. Kita dapat mengukur koefisien uncertainty atribut X dengan rumus $UC(x) = \text{gain}(x, T) / \text{info}(T)$ dan $\text{Gain}(x, T) = \text{info}(T) - \text{info}(x, T)$.

Fase 2: Menemukan frequent item dan menyimpan pada struktur lengkap. Menggabungkan set dan mendaftarkannya sebagai hitungan jika ada transaksi yang multiple. Menscan database dan menemukan semua frequent item dan supportnya. Membuat tree dengan root null. Mendapatkan transaksi pertama dari DB, menghilangkan semua item yang non-frequent, dan mendata sisanya berdasar pada urutan item frequent. Menggunakan transaksi untuk membangun cabang pertama pada tree dengan setiap node menghubungkan ke frequent items. Membaca transaksi berikutnya dan menghilangkan non-frequent item, memasukkan ke dalam tree dan menambah item count. Dilanjutkan sampai semua transaksi selesai diproses.

4. IMPLEMENTASI

Tabel hasil

Table 1: Results obtained from DIP

Disease	Area	Gender	Age group	Sanitation	Food source	water source	type of food	Hereditary	Income	Nature of job
Typhoid	Urban	Male	0-10, 11-20	Average, poor	Outside	Open source	Non-Vegetarian	No	NA	NA, self emp
Dengue	Urban	Male	21-30	Good	Home	NA	Non-Vegetarian		NA	NA, self emp
Cholera	Rural	Female	0-10	Average, poor	Outside	Open source	Non-Vegetarian	No	NA	NA, self emp
Heart Attack	Urban	Male	40 above	Good	Outside	NA	Non-Vegetarian	Yes	High salaried	self emp

Pendapat :

Algoritma hybrid baru yang diusulkan ini unik dan berbeda dari umumnya yang digunakan untuk algoritma prediksi pada data mining. Metode yang diusulkan ini mengatasi kerugian dari metode yang ada pada jumlah frequent item yang kurang. Algoritma baru ini membuktikan efisien pada waktu dan kompleksitas ruang dan membuktikan akurasi saat dibandingkan dengan tool standar analisis statistika seperti SPSS.

TUGAS ADVANCE DATABASE

Implementasi Association Rules dengan Algoritma Apriori pada Dataset Kemiskinan

Nama : Ari Hardiyantoro Susanto

Nim : 202420015

Pengumpulan Data

Dataset yang digunakan merupakan data sekunder dari BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan Tahun 2009-2014. Adapun dataset yang berjumlah 46 atribut ini terdiri dari:

1. Jumlah penduduk miskin (A1).
2. Pengeluaran rata-rata perkapita untuk kelompok makanan (A2).
3. Pengeluaran rata-rata perkapita untuk kelompok non makanan (A3).
4. Produk Domestik Regional Bruto (PDRB) yang terbagi menjadi Atas Dasar Harga Berlaku (ADHB) dan Atas Dasar Harga Konstan (ADHK) pada setiap sektornya sejumlah 14 sektor, yaitu PDRB Sektor Bangunan; PDRB Sektor Industri Pengolahan; PDRB Sektor Jasa-Jasa; PDRB Sektor Keuangan; PDRB Sektor Persewaan dan Jasa Perusahaan; PDRB Sektor Listrik dan Gas; PDRB Sektor Air Bersih; PDRB Sektor Pengangkutan; PDRB Sektor Komunikasi; PDRB Sektor Perdagangan, Hotel dan Restoran; PDRB Sektor Pertambangan dan Penggalian; PDRB Sektor Pertanian dan Tanaman Pangan; PDRB Sektor Perkebunan; PDRB Sektor Peternakan; PDRB Sektor Kehutanan; serta PDRB Sektor Perikanan (A4 sampai dengan A35).
5. Inflasi (A36).
6. Luas panen tanaman padi dan palawija yang terdiri dari 7 kelompok, yaitu padi; jagung; kedelai; kacang tanah; kacang hijau; ubi kayu; dan ubi jalar (A37 sampai dengan A43).
7. Jumlah industri manufaktur besar dan sedang (A44).
8. Jumlah pengangguran (A45).
9. Jumlah penduduk (A46). Semua atribut pada dataset merupakan data numerik. Namun data inflasi dan luas panen tanaman berupa data bulanan, sedangkan data yang lain berupa data tahunan. Sehingga data inflasi dan luas panen tanaman diubah menjadi data tahunan dengan cara menghitung rata-rata dari masing-masing data tersebut.

Preprocessing Data

Tahapan preprocessing data dibagi menjadi 2 tahap, yaitu cleaning data dan transformasi data kategorikal. Cleaning data untuk menghilangkan data yang tidak konsisten. Atribut ADHB (A6) dan ADHK (A22) Sektor Jasa-jasa untuk Tahun 2009-2010 dan Tahun 2011-2014 mempunyai pengumpulan data jasa-jasa yang berbeda sehingga kedua atribut harus dihilangkan dari dataset. Selain kedua atribut tersebut, atribut jumlah industri besar dan sedang (A44) juga perlu dihilangkan dari dataset karena data untuk tahun 2011-2014 tidak diketahui. Sehingga dataset sekarang berjumlah 43 atribut. Berdasarkan kajian, maka proses transformasi data dimulai dengan melakukan perhitungan selisih dari data yang telah terurut dan mencari persentase data.

Hasil persentase data kemudian dibagi menjadi 3 bagian yaitu kenaikan, tetap, dan penurunan. Untuk pergerakan yang tidak berubah diberikan kategori Tetap. Sedangkan untuk menentukan kategori kenaikan dan penurunan, data difuzzifikasi menggunakan fungsi keanggotaan segitiga tipe bahu untuk nilai persentase kenaikan dan penurunan. Setiap fungsi keanggotaan kenaikan dan penurunan dibagi menjadi 3 kategori, yaitu Rendah, Sedang, dan Tinggi. Hal ini menyebabkan setiap atribut numerik mempunyai kemungkinan berubah menjadi 7 data kategorikal, sehingga dataset yang awalnya berukuran 43×5 sekarang menjadi dataset berukuran 301×5 .

Association Rules dengan Algoritma Apriori

Setelah dataset ditransformasi menjadi atribut kategorikal, berikutnya diterapkan association rules dengan algoritma Apriori. Langkah-langkah pada tahap ini adalah sebagai berikut:

1. Menentukan minimum support dan minimum confidence.
2. Menghitung k- itemset, dengan k adalah jumlah kombinasi itemset.
3. Menentukan frequent(k) itemset, yaitu itemset yang memenuhi minimum support.
4. Menentukan rules.

Hasil Penelitian

Dataset hasil preprocessing yang menghasilkan 43 atribut numerik disajikan pada Tabel 1.

Tabel 1. Dataset Hasil *Preprocessing* Data

No	Nama Atribut	Atribut	Tahun					
			2009	2010	2011	2012	2013	2014
1	Penduduk Miskin	A1	14039	13654	14104	14698	15218	13856
2	Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan	A2	308434	345529	364140	424891	474028	474164
3	Pengeluaran Rata-Rata per Kapita untuk Kelompok Non Makanan	A3	264873	242825	326409	383895	370033	369209
...
41	Luas Panen Tanaman Ubi Jalar (Ha)	A43	151	174	180	203	205	228
42	Jumlah Pengangguran	A45	7992	5991	7962	6519	4589	4775
43	Jumlah Penduduk	A46	274526	297814	303190	308510	131725	319098

Dataset pada Tabel 1 kemudian ditransformasi menjadi data kategorikal.

Apabila hasil selisih data menunjukkan kenaikan maka digunakan fungsi keanggotaan fuzzy yang disajikan pada Persamaan 4, 5, dan 6. Fungsi keanggotaan untuk kategori Naik Rendah (NR):

$$\mu_{NR}(x) = \begin{cases} 1 & , x \leq 33.33 \\ \frac{40-x}{6.67} & , 33.33 < x < 40 \\ 0 & , x \geq 40 \end{cases} \quad (4)$$

Fungsi keanggotaan untuk kategori Naik Sedang (NS):

$$\mu_{NS}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-33.33}{16.67} & , 33.33 \leq x < 50 \\ \frac{66.67-x}{16.67} & , 50 < x < 66.67 \\ 0 & , x \geq 66.67 \end{cases} \quad (5)$$

Fungsi keanggotaan untuk kategori Naik Tinggi (NT):

$$\mu_{NT}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-60}{6.67} & , 60 < x < 66.67 \\ 1 & , x \geq 66.67 \end{cases} \quad (6)$$

Sedangkan apabila hasil selisih data menunjukkan penurunan maka digunakan Persamaan 7, 8, dan 9.

Fungsi keanggotaan untuk kategori Turun Rendah (TR):

$$\mu_{TR}(x) = \begin{cases} 1 & , x \geq -33.33 \\ \frac{x+40}{6.67} & , -40 < x < -33.33 \\ 0 & , x \leq -40 \end{cases} \quad (7)$$

Fungsi keanggotaan untuk kategori Turun Sedang (TS):

$$\mu_{TS}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{-x-33.33}{16.67} & , -50 \leq x < -33.33 \\ \frac{x+66.67}{16.67} & , -66.67 < x < -50 \\ 0 & , x \leq -66.67 \end{cases} \quad (8)$$

Fungsi keanggotaan untuk kategori Turun Tinggi (TT):

$$\mu_{TT}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{x+60}{6.67} & , -66.67 < x < -60 \\ 1 & , x \leq -66.67 \end{cases} \quad (9)$$

Berdasarkan hasil penerapan association rules dengan algoritma Apriori pada dataset kemiskinan menggunakan minimum support 30% dan minimum confidence 80% diperoleh 4614 rules hubungan antar indikator. Namun, banyaknya jumlah rules yang dihasilkan mengakibatkan banyaknya pengetahuan mengenai pola hubungan antar indikator. Pengetahuan yang terlalu luas ini berakibat pada kesulitan pengambilan keputusan untuk menentukan pola hubungan antar indikator mana yang lebih berpengaruh terhadap kemiskinan. Guna menghasilkan pola hubungan indikator yang lebih spesifik maka akan diterapkan Principal Component Analysis sebagai salah satu preprocessing data pada penerapan Algoritma Apriori untuk penelitian selanjutnya.

ASSOCIATION RULE (ALGORITMA PRIORI)

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah market basket analysis.

Suatu aturan asosiatif dapat diketahui dengan dua parameter :

1. Support (nilai penunjang)
yaitu persentase kombinasi item tersebut dalam database.
2. Confidence (nilai kepastian)
yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Aturan asosiatif biasanya dinyatakan dalam bentuk :

{roti, mentega} \Rightarrow {susu} (support = 40%, confidence = 50%)

Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu." Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini." Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

a. Analisa pola frekuensi tinggi

Yaitu mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database.

Nilai support sebuah item diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

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

b. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif $A \rightarrow B$

Nilai confidence dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$\text{Confidence} = P(B | A) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}}$$

Sebagai contoh ambil suatu data transaksi yang didapat dari penjualan sayur dengan data transaksi sebagai berikut :

Transaksi	Item yang di beli
1	Broccoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Broccoli
6	Squash, Asparagus, Beans, Tomatoes
7	Tomatoes, corn
8	Broccoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Broccoli, Beans, Squash
12	Asparagus, Bean, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green Peppers, Tomatoes, Beans, Broccoli

I. Definisi-definisi yang terdapat pada Association Rule

1. I adalah himpunan yang tengah dibicarakan.
Contoh: {Asparagus, Beans, ..., Tomatoes}
2. D adalah Himpunan seluruh transaksi yang tengah dibicarakan
Contoh: {Transaksi 1, transaksi 2, ..., transaksi 14}

3. Proper Subset adalah Himpunan Bagian murni

Contoh: Ada suatu himpunan $A = \{a, b, c\}$

Himpunan Bagian dari A adalah

Himpunan Kosong = $\{\}$

Himpunan 1 Unsur = $\{a\}, \{b\}, \{c\}$

Himpunan 2 Unsur = $\{a, b\}, \{a, c\}, \{b, c\}$

Himpunan 3 Unsur = $\{a, b, c\}$

Proper subset nya adalah Himpunan 1 Unsur dan Himpunan 2 Unsur

4. Item set adalah Himpunan item atau item-item di I

Contoh:

Ada suatu himpunan $A = \{a, b, c\}$

Item set nya adalah $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$

5. K- item set adalah Item set yang terdiri dari K buah item yang ada pada I. Intinya K itu adalah jumlah unsur yang terdapat pada suatu Himpunan

Contoh : 3-item set adalah yang bersifat 3 unsur

6. Item set Frekuensi adalah Jumlah transaksi di I yang mengandung jumlah item set tertentu. Intinya jumlah transaksi yang membeli suatu item set.

Contoh :

Kita gunakan tabel transaksi penjualan sayur di atas

- frekuensi Item set yang sekaligus membeli Beans dan Broccoli adalah 3

- frekuensi item set yang sekaligus membeli Beans, Squash dan Tomatoes adalah 2.

7. Frekuensi Item Set adalah item set yang muncul sekurang-kurangnya "sekian" kali di D. Kata "sekian" biasanya di simbolkan dengan Φ . Φ merupakan batas minimum dalam suatu transaksi.

Contoh :

Pertama kita tentukan $\Phi = 3$, karena jika tidak di tentukan maka frekuensi item set tidak dapat di hitung.

Jika $\Phi=3$ untuk {Asparagus, Beans} apakah frekuensi Item set?

Jika kita hitung maka jumlah transaksi yang membeli asparagus sekaligus membeli beans adalah 5. Karena $5 \geq 3$ maka {Asparagus, Beans} merupakan Frekuensi Item set.

8. F_k adalah Himpunan semua frekuensi Item Set yang terdiri dari K item.

II. Langkah-langkah algoritma pada **Association Rule**

1. Tentukan Φ
2. Tentukan semua Frekuensi Item set
3. Untuk setiap Frekuensi Item set lakukan hal sbb:
 - i. Ambil sebuah unsur, namakanlah s
 - ii. Untuk sisanya namakanlah ss-s
 - iii. Masukkan unsur-unsur yang telah di umpamakan ke dalam rule If (ss-s) then s

Untuk langkah ke 3 lakukan untuk semua unsur.

Contoh Pembahasan :

Terdapat transaksi seperti dibawah ini

Transaksi	Item yang di beli
1	C, E, D
2	A, F, D
3	D, G, B, F
4	E, D, G, B
5	B, A, C
6	F, A, B, G
7	G, D
8	C, G, E
9	F, A, B
10	B, D

1. Pisahkan masing-masing item yang dibeli

Item yang di beli
A
B
C
D
E
F
G

2. Kemudian Buat Tabel seperti dibawah ini:

Transaksi	A	B	C	D	E	F	G
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0

3. Kemudian hitung jumlah banyaknya pembelian untuk setiap item.

Transaksi	A	B	C	D	E	F	G
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0
Σ	4	6	3	6	3	4	5

4. Tentukan Φ .

Misalkan kita tentukan $\Phi = 3$, maka kita dapat menentukan frekuensi itemset. Dari tabel di atas diketahui total Φ untuk transaksi $k = 1$, semuanya lebih besar dari Φ . Maka:

$$F_1 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{G\}\}$$

Untuk $k = 2$ (2 unsur), diperlukan tabel untuk tiap-tiap pasang item. Himpunan yang mungkin terbentuk adalah: $\{A,B\}, \{A,C\}, \{A,D\}, \{A,E\}, \{A,F\}, \{A,G\}, \{B,C\}, \{B,D\}, \{B,E\}, \{B,F\}, \{B,G\}, \{C,D\}, \{C,E\}, \{C,F\}, \{C,G\}, \{D,E\}, \{D,F\}, \{D,G\}, \{E,F\}, \{E,G\}, \{F,G\}$.

Tabel-tabel untuk calon 2 item set:

T	A	B	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	1	S
		Σ	3

T	A	C	f
1	0	1	S
2	1	0	S
3	0	0	S
4	0	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	0	0	S
		Σ	1

T	A	D	f
1	0	1	S
2	1	1	P
3	0	1	S
4	0	1	S
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	1	0	S
10	0	1	S
		Σ	1

T	A	E	f
1	0	1	S
2	1	0	S
3	0	0	S
4	0	1	S
5	1	0	S
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	0	0	S
		Σ	0

T	A	F	f
1	0	0	S
2	1	1	P
3	0	1	S
4	0	0	S
5	1	0	S
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	0	S
		Σ	3

T	A	G	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	0	0	S
		Σ	1

T	B	C	f
1	0	1	S
2	0	0	S
3	1	0	S
4	1	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	1	0	S
		Σ	1

T	B	D	f
1	0	1	S
2	0	1	S
3	1	1	P
4	1	1	P
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	1	0	S
10	1	1	P
		Σ	3

T	B	E	f
1	0	1	S
2	0	0	S
3	1	0	S
4	1	1	P
5	1	0	S
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	1	0	S
		Σ	1

T	B	F	f
1	0	0	S
2	0	1	S
3	1	1	P
4	1	0	S
5	1	0	S
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	1	0	S
		Σ	3

T	B	G	f
1	0	0	S
2	0	0	S
3	1	1	P
4	1	1	P
5	1	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	1	0	S
		Σ	3

T	C	D	f
1	1	1	P
2	0	1	S
3	0	1	S
4	0	1	S
5	1	0	S
6	0	0	S
7	0	1	S
8	1	0	S
9	0	0	S
10	0	1	S
		Σ	1

T	C	E	f
1	1	1	P
2	0	0	S
3	0	0	S
4	0	1	S
5	1	0	S
6	0	0	S
7	0	0	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	2

T	C	F	f
1	1	0	S
2	0	1	S
3	0	1	S
4	0	0	S
5	1	0	S
6	0	1	S
7	0	0	S
8	1	0	S
9	0	1	S
10	0	0	S
		Σ	0

T	C	G	f
1	1	0	S
2	0	0	S
3	0	1	S
4	0	1	S
5	1	0	S
6	0	1	S
7	0	1	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	1

T	D	E	f
1	1	1	P
2	1	0	S
3	1	0	S
4	1	1	P
5	0	0	S
6	0	0	S
7	1	0	S
8	0	1	S
9	0	0	S
10	1	0	S
		Σ	2

T	D	F	f
1	1	0	S
2	1	1	P
3	1	1	P
4	1	0	S
5	0	0	S
6	0	1	S
7	1	0	S
8	0	0	S
9	0	1	S
10	1	0	S
		Σ	2

T	D	G	f
1	1	0	S
2	1	0	S
3	1	1	P
4	1	1	P
5	0	0	S
6	0	1	S
7	1	1	P
8	0	1	S
9	0	0	S
10	1	0	S
		Σ	3

T	E	F	f
1	1	0	S
2	0	1	S
3	0	1	S
4	1	0	S
5	0	0	S
6	0	1	S
7	0	0	S
8	1	0	S
9	0	1	S
10	0	0	S
		Σ	0

T	E	G	f
1	1	0	S
2	0	0	S
3	0	1	S
4	1	1	P
5	0	0	S
6	0	1	S
7	0	1	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	2

T	F	G	f
1	0	0	S
2	1	0	S
3	1	1	P
4	0	1	S
5	0	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	0	0	S
		Σ	2

Dari tabel-tabel 2 unsur di atas, P artinya item-item yang dijual bersamaan, sedangkan S berarti tidak ada item yang dijual bersamaan atau tidak terjadi transaksi. Σ melambangkan jumlah Frekuensi item set. Jumlah frekuensi item set harus lebih besar atau sama dengan jumlah Frekuensi item set ($\Sigma \geq \Phi$). Dari tabel diatas, maka didapat:

$$F2 = \{\{A,B\}, \{A,F\}, \{B,D\}, \{B,F\}, \{B,G\}, \{D,G\}\}$$

Kombinasi dari itemset dalam F2, dapat kita gabungkan menjadi calon 3-itemset. Itemset-itemset yang dapat digabungkan adalah itemset-itemset yang memiliki kesamaan dalam k-1

item pertama. Misalnya {A,B} dan {A,F} memiliki itemset k-1 pertama yg sama, yaitu A, maka dapat digabungkan menjadi 3- itemset baru yaitu {A, B, F}.

Untuk k = 3 (3 unsur), himpunan yang mungkin terbentuk adalah: {A, B, F}, {B, D, F}, {B, D,G}, {B, F, G}

T	A	B	F	f
1	0	0	0	S
2	1	0	1	S
3	0	1	1	S
4	0	1	0	S
5	1	1	0	S
6	1	1	1	P
7	0	0	0	S
8	0	0	0	S
9	1	1	1	P
10	0	1	0	S
			Σ	2

T	B	D	F	f
1	0	1	0	S
2	0	1	1	S
3	1	1	1	P
4	1	1	0	S
5	1	0	0	S
6	1	0	1	S
7	0	1	0	S
8	0	0	0	S
9	1	0	1	S
10	1	1	0	S
			Σ	1

T	B	D	G	f
1	0	1	0	S
2	0	1	0	S
3	1	1	1	P
4	1	1	1	P
5	1	0	0	S
6	1	0	1	S
7	0	1	1	S
8	0	0	1	S
9	1	0	0	S
10	1	1	0	S
			Σ	2

T	B	F	G	f
1	0	0	0	S
2	0	1	0	S
3	1	1	1	P
4	1	0	1	S
5	1	0	0	S
6	1	1	1	P
7	0	0	1	S
8	0	0	1	S
9	1	1	0	S
10	1	0	0	S
			Σ	2

Dari tabel-tabel di atas, didapat $F_3 = \{ \}$, karena tidak ada $\Sigma \geq \Phi$ sehingga F_4, F_5, F_6 dan F_7 juga merupakan himpunan kosong.

- Rule yang dipakai adalah if x then y, dimana x adalah antecedent dan y adalah consequent. Berdasarkan rule tersebut, maka dibutuhkan 2 buah item yang mana salah satunya sebagai antecedent dan sisanya sebagai consequent. Dari langkah 4 didapat 1 buah F_k yaitu F_2 . F_1 tidak disertakan karena hanya terdiri dari 1 item saja. Untuk antecedent boleh lebih dari 1 unsur, sedangkan untuk consequent terdiri dari 1 unsur.
- Tentukan (ss-s) sebagai antecedent dan s sebagai consequent dari F_k yang telah di dapat berdasarkan rule pada langkah 5.

Pada F_2 didapat himpunan $F_2 = \{ \{A,B\}, \{A,F\}, \{B,D\}, \{B,F\}, \{B,G\}, \{D,G\} \}$

Maka dapat disusun:

- Untuk {A,B}:
 - Jika (ss-s) = A, Jika s = B, Maka \rightarrow If buy A then buy B
 - Jika (ss-s) = B, Jika s = A, Maka \rightarrow If buy B then buy A
- Untuk {A,F}:
 - Jika (ss-s) = A, Jika s = F, Maka \rightarrow If buy A then buy F
 - Jika (ss-s) = F, Jika s = A, Maka \rightarrow If buy F then buy A
- Untuk {B,D}:
 - Jika (ss-s) = B, Jika s = D, Maka If buy B then buy D
 - Jika (ss-s) = D, Jika s = B, Maka If buy D then buy B
- Untuk {B,F}:
 - Jika (ss-s) = B, Jika s = F, Maka If buy B then buy F
 - Jika (ss-s) = F, Jika s = B, Maka If buy F then buy B
- Untuk {B,G}:
 - Jika (ss-s) = B, Jika s = G, Maka If buy B then buy G

- Jika (ss-s) = G, Jika s = B, Maka If buy G then buy B
 - Untuk {D,G} :
 - Jika (ss-s) = D, Jika s = G, Maka If buy D then buy G
 - Jika (ss-s) = G, Jika s = D, Maka If buy G then buy D
7. Dari langkah 6, kita mendapatkan 12 rule yang dapat digunakan yaitu
- If buy A then buy B
 - If buy B then buy A
 - If buy A then buy F
 - If buy F then buy A
 - If buy B then buy D
 - If buy D then buy B
 - If buy B then buy F
 - If buy F then buy B
 - If buy B then buy G
 - If buy G then buy B
 - If buy D then buy G
 - If buy G then buy D
8. Dari langkah 7 dibuat tabel kandidat asosiasi rule untuk 1 antecedent

If antecedent then consequent	Support	Confidence
If buy A then buy B		
If buy B then buy A		
If buy A then buy F		
If buy F then buy A		
If buy B then buy D		
If buy D then buy B		
If buy B then buy F		
If buy F then buy B		
If buy B then buy G		
If buy G then buy B		
If buy D then buy G		
If buy G then buy D		

9. Hitung support dan confidence.

$$SUPPORT = \frac{\Sigma \text{item.yang.dibeli.sekaligus}}{\Sigma \text{jumlah.seluruh.transaksi}} \times 100\%$$

$$CONFIDENCE = \frac{\Sigma \text{item.yang.dibeli.sekaligus}}{\Sigma \text{jumlah.transaksi.pada.bagian.antecedent}} \times 100\%$$

Untuk Σ item yang dibeli sekaligus pada If buy A then buy B, ada 3 transaksi. Jumlah seluruh transaksi adalah 10 transaksi, sehingga support-nya adalah:

$$SUPPORT = \frac{3}{10} \times 100\% = 33,33\%$$

Untuk Σ item yang dibeli sekaligus pada If buy A then buy B, ada 3 transaksi, sedangkan jumlah transaksi yang membeli A adalah 4 transaksi, sehingga confidence-nya adalah:

$$CONFIDENCE = \frac{3}{4} \times 100\% = 75\%$$

Sehingga di dapat tabel sbb:

If antecedent then consequent	Support	Confidence
If buy A then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy A	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy A then buy F	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy F then buy A	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy D	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy D then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy B then buy F	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy F then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy G	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy G then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/5) \times 100\% = 60\%$
If buy D then buy G	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy G then buy D	$(3/10) \times 100\% = 33,33\%$	$(3/5) \times 100\% = 60\%$

10. Setelah di dapat support dan confidence untuk masing-masing kandidat, lakukan perkalian antara support dan confidence, dimana confidence-nya diambil 70% ke atas, sehingga di dapat tabel sbb:

If antecedent then consequent	Support	Confidence	Support x confidence
If buy A then buy B	33,33%	75%	0.249975
If buy A then buy F	33,33%	75%	0.249975
If buy F then buy A	33,33%	75%	0.249975
If buy F then buy B	33,33%	75%	0.249975

11. Setelah didapat hasil perkalian antara support dan confidence, pilihlah yang hasil perkaliannya paling besar. Hasil paling besar dari perkalian perkalian tersebut merupakan rule yang dipakai pada saat menjual. Karena hasil perkalian dari ke-4 penjualan diatas bernilai sama, maka semuanya bisa dijadikan rule.

- Jika membeli A maka akan membeli B dengan support 33,33% dan confidence 75%
- Jika membeli A maka akan membeli F dengan support 33,33% dan confidence 75%

- Jika membeli F maka akan membeli A dengan support 33,33% dan confidence 75%
- Jika membeli F maka akan membeli B dengan support 33,33% dan confidence 75%

Review :

Algoritma A Priori diimplementasikan untuk menentukan kemungkinan item-item barang yang di beli oleh pelanggan. dengan metode tersebut penentuan pola pembelian dapat dilakukan dengan melihat hasil dari kecenderungan konsumen membeli item barang berdasarkan kombinasi 2 atau 3 itemset.

Pengetahuan baru yang dapat diperoleh berdasarkan hasil perhitungan algoritma apriori dan sistem yang dibangun dapat dilakukan pengaturan tata letak item barang secara berdekatan untuk memudahkan keberadaan item barang.

Sehingga dengan penerapan implementasi Algoritma A Priori dapat berguna untuk meningkatkan penjualan barang tersebut dengan mengatur 2 atau 3 itemset barang yang sering dibeli oleh konsumen secara bersamaan.

Nama : Bhijanta Wyasa WM
NIM : 202420019
Kelas : MTI 23
Mata Kuliah : ADVANCED DATABASE (MTIK112)
Dosen : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

Tugas 02

Silahkan cari satu paper yang membahas tentang association rule dan buat ringkasan dan pembahasan tentang domain masalah yang dibahas dalam paper dan berikan pendapat anda (review paper). Format bebas.

Implementasi *Association Rule Mining* Untuk Menentukan Menu Paket Makanan Dengan Algoritma *FIN* Menggunakan *Nodesets* (Studi Kasus R.M. Lesehan Nova Sragen)

Riski Nova Saputra¹, Muhammad Tanzil Furqon², Indriati³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹nova.tbk@gmail.com, ²m.tanzil.furqon@ub.ac.id, ³indriati.tif@ub.ac.id

Rumah Makan Nova menyediakan buku menu sebagai media bagi pengunjung untuk memilih item menu pesanan. Terdapat dua macam jenis menu yaitu menu utama dan menu paket. Menu utama berisi daftar seluruh item makanan yang berjumlah 156 item dan item minuman yang berjumlah 44 item. Sedangkan menu paket berisi daftar item paket terdiri dari 30 item paket makanan dan minuman.

Permasalahan yang muncul :

1. Disebabkan pilihan menu yang banyak, setiap anggota rombongan dalam memilih menu memerlukan waktu yang lama, pelayanan menjadi lama karena pelanggan terlalu banyak list menu.
2. Nota pesanan yang dihasilkan oleh tipe rombongan ini memiliki variasi item pesanan yang tinggi. Sehingga apabila variasi item pesanan tinggi maka akan membutuhkan waktu yang lebih lama dalam proses pembuatan setiap item pesanan.
3. Permasalahan lain muncul disebabkan strategi yang diterapkan rumah makan melalui menu paket tidak sesuai dengan keinginan pengunjung rombongan.

Menggunakan metode Association Rule Mining untuk menemukan kombinasi item makanan dan item minuman. Association Rule Mining melakukan mining terhadap data transaksi

Nama : Bhijanta Wyasa WM
NIM : 202420019
Kelas : MTI 23
Mata Kuliah : ADVANCED DATABASE (MTIK112)
Dosen : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

penjualan yang tercatat pada struk pembayaran, yang dicetak oleh mesin kasir pada kurun waktu tertentu. Alogaritma yang digunakan untuk Association Rule Mining menggunakan Alogaritma Fin. Algoritma FIN adalah algoritma frequent item set yang menggunakan struktur data novel berupa nodesets. Merupakan algoritma pengembangan dari algoritma PPV dan Prepost, dengan penyempurnaan pada konsumsi memori yang lebih rendah serta performa eksekusi komputasi yang cepat.

Algoritma FIN adalah algoritma frequent itemset yang menggunakan struktur data novel berupa nodesets. Merupakan algoritma pengembangan dari algoritma PPV dan Prepost, dengan penyempurnaan pada konsumsi memori yang lebih rendah serta performa eksekusi komputasi yang cepat (Deng, 2014). Langkah-langkah algoritma FIN sebagai berikut :

- a. Menghitung nilai support untuk setiap item pada daftar menu.
- b. Mengurutkan item berdasarkan nilai support terbesar sampai terkecil.
- c. Mengeliminasi seluruh item yang memiliki nilai support kurang dari minimum support.
- d. Mengeliminasi item-item pada setiap transaksi yang memiliki nilai support kurang dari minimum support.
- e. Mengurutkan item-item yang tersisa untuk setiap transaksi sesuai dengan urutan L.
- f. Memasukkan item-item pada setiap transaksi pada POC-Tree.
- g. Memindai POC-tree untuk menyusun frequent 1-itemset.
- h. Memindai POC-tree untuk menyusun frequent 2-itemset yang berdasar pada frequent 1-itemset.
- i. Memindai POC-tree untuk menyusun frequent 3-itemset yang berdasar pada frequent 2-itemset.
- j. Menyusun kombinasi menu paket makanan dan minuman berdasarkan frequent 3-itemset.

Algoritma FIN diimplementasikan untuk menentukan menu paket yang representatif dengan pilihan konsumen, pada sistem pembuat menu paket otomatis. Perhitungan support digunakan untuk mengevaluasi hasil variasi menu paket yang dihasilkan oleh sistem. Nilai minimum support sebagai eliminator terhadap item menu yang tidak infrequent. Nilai minimum support menentukan bentuk POC-Tree dalam menghasilkan

Nama : Bhijanta Wyasa WM
NIM : 202420019
Kelas : MTI 23
Mata Kuliah : ADVANCED DATABASE (MTIK112)
Dosen : Tri Basuki Kurniawan , S.Kom., M.Eng. Ph.D

menu paket. Jumlah variasi menu paket berbanding terbalik dengan nilai minimum support. Serta jumlah variasi menu paket berbanding lurus dengan ukuran POC-Tree. Berdasarkan hasil pengujian menunjukkan pada nilai minimum support = 11, telah menghasilkan variasi menu paket yang proporsional serta representatif dengan pilihan konsumen. Pada nilai minimum support = 11 menghasilkan 6 variasi menu paket. Apabila pihak rumah makan menginginkan jumlah variasi menu paket yang lebih banyak, maka dapat digunakan nilai minimum support = 10 yang menghasilkan 24 variasi menu paket. Pada jurnal ini merekomendasikan hasil variasi menu paket untuk nilai minimum support = 11 sebagai pilihan pertama, dan nilai minimum support = 10 sebagai pilihan kedua kepada pihak rumah makan.

PENDAPAT MENGENAI PAPER

Pada jurnal ini dibahas mengenai bagaimana memanfaatkan clustering data menggunakan algoritma fin, yang bertujuan untuk mendapatkan komposisi menu yang sesuai dengan pelanggan. Jika saya melihat jurnal ini hanya sebatas mengenai bagaimana pelayanan pelanggan cepat dengan memanfaatkan algoritma fin. Alangkah baiknya akan dapat diperkaya dengan memperhatikan juga komposisi mesin produksi dengan mekanisme SCM (Supply Chain Management) . Sehingga nilai minimum support yang dihasilkan pada penelitian ini dapat mengakomodir kesiapan dari fungsi back office dan menjaga pelayanan pelanggan juga.

Nama : Cornelia Tri Wahyuni
Mata Kuliah : Advanced Database
NIM : 202420044

Tugas 2

Judul Jurnal	:	Implementasi <i>Association Rules</i> dengan Algoritma Apriori pada Dataset Kemiskinan
Peneliti	:	inda Aprianti, Khairul Anwar Hafizd, M. Redhy Rizani
Publikasi	:	Limits, Journal Of Mathematics and Its Applications
Tahun	:	2017
Pokok Penelitian/Pokok Bahasan	:	<p>Kemiskinan masih menjadi permasalahan yang memerlukan perhatian lebih dari pemerintah. Kalimantan Selatan adalah salah satu Provinsi di Indonesia yang mengalami permasalahan dalam penanggulangan kemiskinan. Penurunan persentase kemiskinan di Kalimantan Selatan dari tahun 2000 sampai dengan 2013 adalah sebesar 8.29%. Persentase kemiskinan di Kabupaten Tanah Laut sendiri telah menurun 5.71% dari tahun 2000-2013 [1].</p> <p>Sekarang ini, kemiskinan tidak lagi dipandang sebagai permasalahan rendahnya pendapatan seseorang, melainkan disebabkan oleh beragamnya faktor penyebab kemiskinan [2, 3, 4]. Faktor pengangguran dan Produk Domestik Regional Bruto (PDRB) termasuk yang diteliti terhadap kemiskinan pada penelitian Wijayanto (2010), Rusdati dan Sebayang (2013), tetapi memberikan hasil yang berbeda.</p> <p>Hal ini menunjukkan bahwa permasalahan kemiskinan sangat kompleks dan perbedaan karakteristik wilayah, serta teknik yang digunakan juga berpengaruh untuk mengetahui indikator penyebab kemiskinan dan perumusan kebijakan untuk pengentasan kemiskinan. Penelitian tentang kemiskinan sebelumnya menggunakan data numerik, tetapi dalam penelitian ini data kemiskinan ditransformasi menjadi data kategorikal. <i>Task</i> data mining yang digunakan untuk mencari pola hubungan antar atribut adalah <i>Association Rules</i> dengan Algoritma Apriori. Penggunaan Algoritma Apriori dinilai efektif untuk menemukan pola hubungan antara beberapa atribut dengan tingkat kecelakaan lalu lintas, hubungan lingkungan fisik dengan terjadinya kebakaran hutan, serta pola pergerakan harga saham antar sektor dan antar perusahaan [5, 6, 7]. Berdasarkan penelitian terdahulu, maka pada penelitian ini <i>Association Rules</i> dengan Algoritma Apriori diimplementasikan pada dataset kemiskinan.</p>
Metode Penelitian	:	<p>Jurnal ini menggunakan metode dengan tiga tahap yaitu Pengumpulan data, <i>Preprocessing</i> data dan <i>Association Rules</i> dengan Algoritma Apriori. Yang pertama pengumpulan data, dataset yang digunakan merupakan data sekunder dari BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan Tahun 2009-2014.</p> <p>Adapun dataset yang berjumlah 46 atribut ini terdiri dari:</p> <ol style="list-style-type: none"> 1. Jumlah penduduk miskin (A1).

2. Pengeluaran rata-rata perkapita untuk kelompok makanan (A2).
3. Pengeluaran rata-rata perkapita untuk kelompok non makanan (A3).
4. Produk Domestik Regional Bruto (PDRB) yang terbagi menjadi Atas Dasar Harga Berlaku (ADHB) dan Atas Dasar Harga Konstan (ADHK) pada setiap sektornya sejumlah 14 sektor, yaitu PDRB Sektor Bangunan; PDRB Sektor Industri Pengolahan; PDRB Sektor Jasa-Jasa; PDRB Sektor Keuangan; PDRB Sektor Persewaan dan Jasa Perusahaan; PDRB Sektor Listrik dan Gas; PDRB Sektor Air Bersih; PDRB Sektor Pengangkutan; PDRB Sektor Komunikasi; PDRB Sektor Perdagangan, Hotel dan Restoran; PDRB Sektor Pertambangan dan Penggalian; PDRB Sektor Pertanian dan Tanaman Pangan; PDRB Sektor Perkebunan; PDRB Sektor Peternakan; PDRB Sektor Kehutanan; serta PDRB Sektor Perikanan (A4 sampai dengan A35).
5. Inflasi (A36).
6. Luas panen tanaman padi dan palawija yang terdiri dari 7 kelompok, yaitu padi; jagung; kedelai; kacang tanah; kacang hijau; ubi kayu; dan ubi jalar (A37 sampai dengan A43).
7. Jumlah industri manufaktur besar dan sedang (A44).
8. Jumlah pengangguran (A45).
9. Jumlah penduduk (A46).

Semua atribut pada dataset merupakan data numerik. Namun data inflasi dan luas panen tanaman berupa data bulanan, sedangkan data yang lain berupa data tahunan. Sehingga data inflasi dan luas panen tanaman diubah menjadi data tahunan dengan cara menghitung rata-rata dari masing-masing data tersebut.

Selanjutnya *Preprocessing* data Tahapan *preprocessing* data dibagi menjadi 2 tahap, yaitu *cleaning* data dan transformasi data kategorikal.

Cleaning data untuk menghilangkan data yang tidak konsisten. Atribut ADHB (A6) dan ADHK (A22) Sektor Jasa-jasa untuk Tahun 2009-2010 dan Tahun 2011-2014 mempunyai pengumpulan data jasa-jasa yang berbeda sehingga kedua atribut harus dihilangkan dari dataset. Selain kedua atribut tersebut, atribut jumlah industri besar dan sedang (A44) juga perlu dihilangkan dari dataset karena data untuk tahun 2011-2014 tidak diketahui. Sehingga dataset sekarang berjumlah 43 atribut.

Proses transformasi data dimulai dengan melakukan perhitungan selisih dari data yang telah terurut dan mencari persentase data. Hasil persentase data kemudian dibagi menjadi 3 bagian yaitu kenaikan, tetap, dan penurunan. Untuk pergerakan yang tidak berubah diberikan kategori Tetap. Sedangkan untuk menentukan kategori kenaikan dan penurunan, data difuzzifikasi menggunakan fungsi keanggotaan segitiga tipe bahu untuk nilai persentase kenaikan dan penurunan. Setiap fungsi keanggotaan kenaikan dan penurunan dibagi menjadi 3

kategori, yaitu Rendah, Sedang, dan Tinggi. Hal ini menyebabkan setiap atribut numerik mempunyai kemungkinan berubah menjadi 7 data kategorikal, sehingga dataset yang awalnya berukuran 43x5 sekarang menjadi dataset berukuran 301x5. Selanjutnya yaitu *Association Rules* dengan Algoritma Apriori Langkah-langkah pada tahap ini adalah sebagai berikut.

1. Menentukan *minimum support* dan *minimum confidence*.
2. Menghitung *k- itemset*, dengan *k* adalah jumlah kombinasi *itemset*.
3. Menentukan *frequent(k) itemset*, yaitu *itemset* yang memenuhi *minimum support*.
4. Menentukan *rules*.

Hasil : Dataset hasil *preprocessing* yang menghasilkan 43 atribut numerik disajikan pada Tabel 1.

Tabel 1. Dataset Hasil *Preprocessing* Data

No	Nama Atribut	Atribut	Tahun					
			2009	2010	2011	2012	2013	2014
1	Penduduk Miskin	A1	14039	13654	14104	14698	15218	13856
2	Pengeluaran RataRata per Kapita untuk Kelompok Makanan	A2	308434	345529	364140	424891	474028	474164
3	Pengeluaran RataRata per Kapita untuk Kelompok Non Makanan	A3	264873	242825	326409	383895	370033	369209
«	«	«	«	«	«	«	«	«
41	Luas Panen Tanaman Ubi Jalar (Ha)	A43	151	174	180	203	205	228
42	Jumlah Pengangguran	A45	7992	5991	7962	6519	4589	4775
43	Jumlah Penduduk	A46	274526	297814	303190	308510	131725	319098

Dataset pada Tabel 1 kemudian ditransformasi menjadi data kategorikal. Apabila hasil selisih data menunjukkan kenaikan maka digunakan fungsi keanggotaan *fuzzy* yang disajikan pada Persamaan 4, 5, dan 6.

Fungsi keanggotaan untuk kategori Naik Rendah (NR):

$$\mu_{NR}(x) = \begin{cases} 1 & , x \leq 33.33 \\ \frac{40-x}{6.67} & , 33.33 < x < 40 \\ 0 & , x \geq 40 \end{cases} \quad (4)$$

Fungsi keanggotaan untuk kategori Naik Sedang (NS):

$$\mu_{NS}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-33.33}{16.67} & , 33.33 \leq x < 50 \\ \frac{66.67-x}{16.67} & , 50 < x < 66.67 \\ 0 & , x \geq 66.67 \end{cases} \quad (5)$$

Fungsi keanggotaan untuk kategori Naik Tinggi (NT):

Implementasi

$$\mu_{NT}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-60}{6.67} & , 60 < x < 66.67 \\ 1 & , x \geq 66.67 \end{cases} \quad (6)$$

Sedangkan apabila hasil selisih data menunjukkan penurunan maka digunakan Persamaan 7, 8, dan 9.

Fungsi keanggotaan untuk kategori Turun Rendah (TR):

$$\mu_{TR}(x) = \begin{cases} 1 & , x \geq -33.33 \\ \frac{x+40}{6.67} & , -40 < x < -33.33 \\ 0 & , x \leq -40 \end{cases} \quad (7)$$

Fungsi keanggotaan untuk kategori Turun Sedang (TS):

$$\mu_{TS}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{-x-33.33}{16.67} & , -50 \leq x < -33.33 \\ \frac{x+66.67}{16.67} & , -66.67 < x < -50 \\ 0 & , x \leq -66.67 \end{cases} \quad (8)$$

Fungsi keanggotaan untuk kategori Turun Tinggi (TT):

$$\mu_{TT}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{x+60}{6.67} & , -66.67 < x < -60 \\ 1 & , x \leq -66.67 \end{cases} \quad (9)$$

Hasil transformasi data disajikan pada Tabel 2.

Tabel 2. Dataset Hasil Transformasi Data

No	Atribut	Tahun					
		2009	2010	2011	2012	2013	2014
1	A1	TR_A1	NR_A1	NR_A1	NR_A1	TR_A1	TR_A1
2	A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2
3	A3	TR_A3	NR_A3	NR_A3	TR_A3	TR_A3	TR_A3
«	«	«	«	«	«	«	«
41	A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43
42	A45	TR_A45	NR_A45	TR_A45	TR_A45	NR_A45	TR_A45
43	A46	NR_A46	NR_A46	NR_A46	TS_A46	NT_A46	NR_A46

Pada Tabel 2, dataset telah berupa atribut kategori: Turun Rendah (TR), Turun Sedang (TS), Turun Tinggi (TT), Tetap (T), Naik Rendah (NR), Naik Sedang (NS), dan Naik Tinggi (NT). Dataset ini kemudian dibentuk menjadi *itemset* yang disajikan pada Tabel 3.

Tabel 3. *Itemset*

No	Item	waktu (t)					Jumlah
		t1	t2	t3	t4	t5	
1	NR_A1	0	1	1	1	0	3
2	NS_A1	0	0	0	0	0	0
3	NT_A1	0	0	0	0	0	0
«	«	«	«	«	«	«	«
299	TR_A46	0	0	0	0	0	0
300	TS_A46	0	0	0	1	0	1
301	TT_A46	0	0	0	0	0	0

Dengan *minimum support* = 30% dan *minimum confidence* = 80%, penerapan *association rules* dengan algoritma Apriori menghasilkan *1itemsets*, *2-itemsets*, *3-itemsets*, *4-itemsets*, *5-itemsets*, *6-itemsets*, *7itemsets*, dan *8-itemsets* seperti yang disajikan pada Gambar 2.

```

Associator output
Apriori
=====
Minimum support: 0.95 (5 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8
Size of set of large itemsets L(2): 28
Size of set of large itemsets L(3): 56
Size of set of large itemsets L(4): 70
Size of set of large itemsets L(5): 56
Size of set of large itemsets L(6): 28
Size of set of large itemsets L(7): 8
Size of set of large itemsets L(8): 1

Best rules found:

1. A4=NR_A4 5 ==> A2=NR_A2 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. A2=NR_A2 5 ==> A4=NR_A4 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

```

Gambar 2. Hasil Penerapan *Association Rules* dengan Algoritma Apriori Menggunakan Weka

Setelah mendapatkan *itemsets*, maka selanjutnya diperoleh 4614 *rules* yang menunjukkan hubungan antar indikator (Tabel 4).

Tabel 4. *Rules* yang diperoleh

No	Rule
1	A4=NR_A4 ==> A2=NR_A2
2	A2=NR_A2 ==> A4=NR_A4
3	A5=NR_A5 ==> A2=NR_A2
«	«
4612	A5=NR_A5 A23=NR_A23 A25=NR_A25 A33=NR_A33 A43=NR_A43 ==> A29=NR_A29
4613	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A43=NR_A43 ==> A33=NR_A33
4614	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A33=NR_A33 ==> A43=NR_A43

Pada Tabel 4 untuk *rule* 1 diperoleh pola hubungan antara indikator 4, yaitu ADHB Sektor Bangunan dengan indikator 2, yaitu Pengeluaran RataRata per Kapita untuk Kelompok Makanan yang menunjukkan bahwa apabila ADHB Sektor Bangunan naik rendah maka Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan juga akan mengalami naik rendah. Pola hubungan yang diperoleh dari *rule* 1 ini dapat membantu pemerintah dalam pengambilan kebijakan yang berkaitan dengan ADHB Sektor Bangunan dan Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan dengan memperhatikan pola hubungan kedua atribut ini. Begitu juga dengan *rule* 2 sampai dengan *rule* 4614.

Berdasarkan hasil penerapan *association rules* dengan algoritma Apriori pada dataset kemiskinan menggunakan *minimum support* 30% dan *minimum confidence* 80% diperoleh 4614 *rules* hubungan antar indikator. Namun, banyaknya jumlah *rules* yang dihasilkan mengakibatkan banyaknya pengetahuan mengenai pola hubungan antar indikator. Pengetahuan yang terlalu luas ini berakibat pada kesulitan pengambilan keputusan untuk menentukan pola hubungan antar indikator mana yang lebih berpengaruh terhadap kemiskinan.

Kelebihan

: Dari pola hubungan yang ada dapat membantu pemerintah dalam pengambilan kebijakan yang berkaitan dengan ADHB Sektor Bangunan dan Pengeluaran Rata-

		Rata per Kapita untuk Kelompok Makanan dengan memperhatikan pola hubungan antar atribut yang ada.
Kekurangan	:	Jumlah <i>rules</i> yang banyak mengakibatkan banyaknya pengetahuan mengenai pola hubungan antar indikator. Pengetahuan yang terlalu luas ini berakibat pada kesulitan pengambilan keputusan untuk menentukan pola hubungan antar indikator mana yang lebih berpengaruh terhadap kemiskinan.
Perbandingan dengan metode usulan/penelitian usulan	:	Perbandingan metode <i>algoritma apriori</i> yang diterapkan pada paper ini yaitu <i>cleaning</i> data dan transformasi data kategorikal saja, berbeda dengan penelitian usulan yang akan menerapkan <i>Principal Component Analysis</i> sebagai salah satu <i>preprocessing</i> data pada penerapan Algoritma Apriori untuk menghasilkan pola hubungan indikator yang lebih spesifik.

Nama : Cynthia Anisa Agatha

NIM : 202420022

Paper yang saya ambil sebagai contoh pada tugas ini berjudul *Association Rule Mining for Selecting Proper Students to Take Part in Proper Discipline Competition: A Case Study of Zhejiang University of Finance and Economics*.

Domain masalah dalam *paper* ini yaitu mencari. Rangkain *rule* akan membantu dalam meningkatkan kemungkinan bagi mahasiswa dalam mendapat penghargaan di kompetisi persahabatan.

Dunia pendidikan menjadi salah satu dari beberapa contoh yang mengundang perhatian bagi para peneliti dan pengajar tentunya. Selain itu juga, adanya perkebang dalam bidang *data mining*, dapat memberikan sebuah metode baru dalam mengekstraksi informasi yang berguna dari data di bidang pendidikan yang cukup rumit. Guna meningkatkan peluang para mahasiswa dapat diberi penghargaan dalam kompetisi persahabatan ini yaitu dengan memilih mahasiswa yang cocok untuk ikut berpartisipasi dalam kompetisi ini.

Pada *paper* ini data yang dikumpulkan sebanyak 164 Sarjana yang akan digunakan sebagai studi kasus. Kemudian untuk semua mahasiswa berasal dari jurusan Teknik Informatika Universitas Zhejiang Ekonomi dan *Finace*. Dan algoritma apriori dengan *group strategy* yang digunakan untuk menemukan hubungan diantara mata kuliah mahasiswa dan penghargaan kompetisi. Pada *paper* ini menggunakan bahasa pemrograman *R*.

Langkah-langkah yang digunakan dalam *paper* ini:

1. *Data preprocessing*
2. *Transformation*
3. *Association rule mining*

- **Data Preprocessing dan Transformation**

Tahapan yang dilakukan penulis yaitu melakukan *Data Preprocessing* dan *Transformation*. *Data Preprocessing* adalah suatu proses/langkah yang dilakukan untuk membuat

data mentah menjadi data yang berkualitas(input yang baik untuk *data mining tools*). Guna melindungi kerahasiaan data dari mahasiswa itu sendiri. Penulis menghilangkan nomor kartu mahasiswa dan nama mahasiswa itu sendiri. Kedua, untuk meningkatkan tingkat akurasi dari *association rule mining*, penulis juga menghilangkan beberapa mata kuliah yang mana mahasiswa jarang ikuti, seperti mata kuliah jurusan kedua untuk teknik informatika, lalu menghapus beberapa mata kuliah yang seluruh mahasiswanya mendapat nilai serupa, seperti Pelatihan Militer, Proyek Kelulusan, dan sebagainya, serta menghapus informasi mahasiswa tersebut yang memiliki sedikit atribut.

Jika mahasiswa yang mendapat nilai bahasa Inggris yang tinggi untuk ujian masuk perguruan tinggi dapat dibebaskan dari ujian bahasa Inggris (2), jadi penulis mengisi nilai NA bahasa Inggris (2) dengan nilai skor maksimal mata kuliah ini. Kemudian, kami penulis nilai NA mata kuliah lain dengan skor rata-rata yang sesuai dengan mata kuliah, dan mengubah catatan apabila ada yang meninggalkan ujian, membatalkan ujian dan mencontek selama ujian dengan memberi nilai 0. Kemudian, penulis menggabungkan beberapa mata kuliah serupa menjadi satu mata kuliah dan menetapkan nilainya dengan rata-rata yang sesuai skor. Misalnya, penulis menggabungkan bahasa Inggris (2), Inggris (3), dan Inggris (4) menjadi perkuliahan inggris. Untuk membuat data sesuai dengan tipe data *association rule mining* di pemrograman menggunakan bahasa *R*, penulis membagi skor menjadi empat kelas: apabila Gagal (skor <60), Lulus (skor >= 60 dan skor <80), Baik (skor >= 80 dan skor <90), dan Sangat Baik (skor >= 90). Setelah data preprocessing dan transformasi dilakukan, diperoleh data yang mengandung informasi dari 159 mahasiswa dengan 25 nilai mata kuliah dan penghargaan kompetisi ketertiban yang diperoleh.

- **Association Rule Mining**

Paper ini menggunakan bahasa pemrograman *R* 3.4.0 dan algoritma apriori dengan strategi pengelompokkan untuk *association rule mining* lebih lanjut. Langkah pertama, kita hitung frekuensi dari setiap *itemset* setelah dari menemukan *frequent 1-itemsets*, kemudian dibagi *itemset* menjadi beberapa kelompok berdasarkan kemunculan dan menamai kelompok sesuai dengan kemunculan frekuensi *item*.

Oleh karena itu, penulis dapat menemukan *frequent k-itemsets* berdasarkan beberapa kelompok, yang mana nilai labelnya lebih besar daripada *k*, bukan seluruh *database*.

- **Hasil dan Analisa**

Dari gambar 1 dibawah ini empat plot aturan asosiasi mahasiswa yang diberikan pada penghargaan kompetisi ketertiban. Setiap titik dalam gambar ini mewakili satu aturan asosiasi, dan warnanya mewakili nilai daya angkat. Sumbu X mewakili nilai dari *support*, dan sumbu Y melambangkan nilai dari *confidence*.

Berdasarkan definisi dari *support* dan *confidence*, *rule* dengan nilai *support* atau *confidence* yang terlalu kecil dianggap tidak memiliki arti, dikarenakan hal tersebut memiliki kemungkinan tinggi yang dapat memicu kecelakaan. Dari tes yang penulis lakukan dari memasukkan *support* dan *confidence* sebesar 0.007 dan 0.08.

- **Kesimpulan**

Pada artikel ini, penulis bertujuan untuk menemukan hubungan antara nilai mata kuliah dan penghargaan kompetisi ketertiban. Aturan tersebut akan membantu meningkatkan kemungkinan bagi siswa untuk diberikan dalam kompetisi disiplin. Kami mengekstrak informasi berguna dari 164 mahasiswa sarjana yang terdaftar di jurusan teknik informatika Universitas Zhejiang Ekonomi dan Finance untuk *Association Rule Mining*. Data mentah telah diproses sebelumnya dan diubah menjadi format yang sesuai untuk algoritma Apriori dengan strategi pengelompokan dalam Pemrograman R. Kemudian penulis memperoleh data yang berisi 159 mahasiswa dengan 25 nilai matakuliah dan penghargaan kompetisi ketertiban yang sesuai.

Akhirnya, sesuai dengan hasil yang diperoleh dari *Association Rule Mining*, yaitu dengan ditunjukkan pada Tabel 2-5, sebagian, kami menemukan bahwa C # Development, Internet WebKursus Desain, Berorientasi Objek, Struktur Data (C #), dan Pemrograman Dasar berpengaruh besar dalam peluang mahasiswa untuk mendapatkan penghargaan dalam kompetisi ketertiban. Namun, masih ada beberapa batasan dalam *paper* ini. Misalnya, karena Teknik Informatika adalah jurusan yang relatif baru di Universitas Zhejiang Ekonomi dan Finance, penulis hanya mengumpulkan informasi dari 164 mahasiswa. Kumpulan data yang terbatas juga dapat mempengaruhi akurasi *association rule mining*. Pada artikel ini, penulis hanya menganalisis beberapa *rule* perwakilan, tetapi *rule* yang tersisa akan dianalisis dalam pekerjaan mereka selanjutnya. Selain itu, penulis juga dapat menggabungkan teknologi *data mining* lainnya dengan

association rule mining untuk menemukan hubungan yang bermakna dengan akurasi dan efisiensi yang lebih tinggi.

Dari *Paper* ini dapat dikatakan *association rule mining* berperan penting dalam bidang *data mining*. Dikarenakan, proses *association rule mining* dapat memberikan kemudahan *user* dalam menunjukkan nilai asosiatif yang memiliki keterkaitan antara variabel didalam *database* yang besar dan tercipta sebuah pola yang dapat menentukan domain permasalahan sebuah sistem dari pola yang telah didapat maka dapat dibuat sebuah dasar keputusan dalam mendapatkan hasil dari sistem yang diinginkan.

Nama : Efrik Kartono Ahsa
NIM : 202420030
MK : Advanced Database (Review Paper)

Judul : Association Rule Mining Untuk Penentuan Rekomendasi Promosi Produk
Jurnal : SIMETRIS, Vol 5 No 2 Nopember 2014
Volume & Halaman : Vol 5 No 2 Nopember 2014
Tahun : 2014
Penulis : Wiwit Agus Triyanto
Reviewer : Efrik Kartono Ahsa (202420030)
Tanggal : 06 Oktober 2020

A. PENDAHULUAN

Rekomendasi Promosi Produk merupakan model aplikasi dari hasil observasi terhadap keadaan dan keinginan pelanggan untuk membeli suatu produk. Rekomendasi Promosi Produk memanfaatkan pendapat pelanggan terhadap suatu barang untuk membantu pelanggan dalam memilih produk. Karena itu Rekomendasi Promosi Produk memerlukan model rekomendasi yang tepat agar apa yang direkomendasikan sesuai dengan keinginan pelanggan, serta mempermudah pelanggan mengambil keputusan yang tepat dalam menentukan produk yang akan dibelinya.

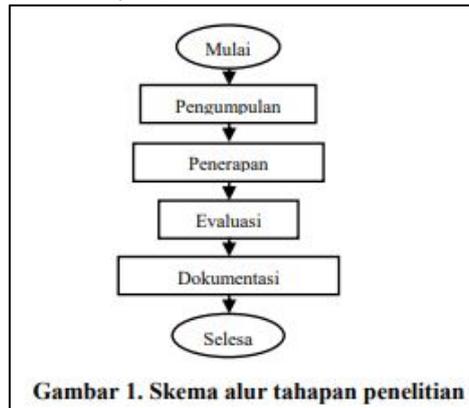
Analisis keranjang pasar (juga dikenal sebagai Association Rule Mining) merupakan salah satu metode data mining yang berfokus pada menemukan pola pembelian dengan mengekstraksi asosiasi atau kejadian dari data transaksional sebuah toko [1]. Selain itu, analisis keranjang pasar adalah cara yang baik untuk memberikan dukungan keputusan ilmiah pada pasar retail melalui hubungan mining association antara barang yang telah dibeli secara bersama-sama [2]. Penemuan dari hubungan ini dapat membantu pedagang untuk mengembangkan strategi penjualan dengan mempertimbangkan barang yang sering dibeli bersama oleh pelanggan [3], hal ini sangat penting karena dapat membantu rekomendasi produk dan promosi produk sehingga strategi pemasaran menjadi lebih tepat. Penentuan pola pembelian barang yang kurang akurat dapat mengakibatkan kebijakan rekomendasi produk dan promosi produk menjadi tidak tepat sasaran.

Banyak algoritma yang diusulkan untuk menemukan pola asosiasi [4], seperti Christidist et al. melakukan penelitian pada e-commerce dengan menggunakan latent topic models pada data transaksi histori pasar pelanggan sehingga dapat membantu menentukan rekomendasi produk kepada pelanggan menjadi lebih efektif bahkan pada dataset dan itemset yang besar, Xian Wen et al. menentukan association rule dengan menggunakan matrix database.

B. METODOLOGI PENELITIAN

Metode penelitian dalam penelitian ini menggunakan metode eksperimen dengan tahapan : (1) Pengumpulan data, (2) Penerapan algoritma, (3) Evaluasi hasil, (4) Dokumentasi eksperimen. Adapun skema alur tahapan penelitian ini ditunjukkan pada gambar 1.

Nama : Efrik Kartono Ahsa
NIM : 202420030
MK : Advanced Database (Review Paper)



C. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Teknik pengumpulan data pada penelitian ini dilakukan dengan mengambil dataset analisis keranjang pasar. Dalam dataset ini mendefinisikan transaksi penjualan pada sebuah supermarket. Terdapat 304 atribut yang 303 merupakan item/barang dan salah satunya merupakan ID transaksi penjualan. Terdiri dari 1361 record transaksi. Adapun contoh data dari dataset dapat dilihat pada tabel 1.

Basket ID	Lemons	Standard coffee	Frozen Chicken Wings	98pct. Fat Free Hamburger	Sugar Cookies
C11867	false	false	true	true	false
C5096	false	false	false	true	false
C4295	false	false	false	false	false
C2837	true	false	false	false	true
C2693	true	false	true	false	false
....

Keterangan:

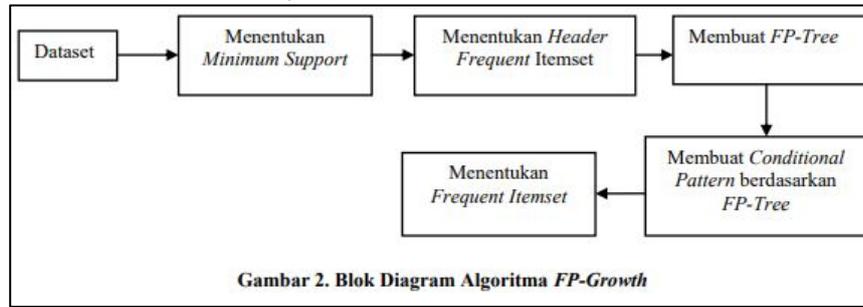
- Basket ID** merupakan ID Transaksi penjualan barang.
- Lemons, Standard coffee, Frozen Chicken Wings, 98pct. Fat Free Hamburger, Sugar Cookies** merupakan barang-barang yang dijual di Supermarket.
- False** menandakan bahwa barang tersebut tidak dibeli oleh pembeli.
- True** menandakan bahwa barang tersebut dibeli oleh pembeli.

2. Penerapan Algoritma

Algoritma yang biasa dipakai dalam mencari frequent itemset untuk penentuan rekomendasi promosi produk antara lain algoritma Apriori dan algoritma FP-Growth. Pada penelitian ini akan dibahas bagaimana pencarian frequent itemset menggunakan algoritma FP-Growth. FP-Tree (Frequent Pattern Tree) digunakan bersamaan dengan algoritma FP-Growth untuk menentukan frequent itemset (data yang paling sering muncul) dari sebuah dataset. Algoritma Apriori memerlukan langkah candidate generation, yaitu dengan melakukan scanning dataset secara berulang-ulang untuk menentukan frequent itemset. Algoritma FP-Growth adalah salah satu cara alternatif untuk menemukan himpunan data yang paling sering muncul (frequent itemset) tanpa menggunakan generasi kandidat [5].

Untuk menentukan frequent itemset pada data transaksi tersebut, dapat ditunjukkan dengan bagan berikut:

Nama : Efrik Kartono Ahsa
 NIM : 202420030
 MK : Advanced Database (Review Paper)



Gambar 2 blok diagram algoritma FP-Growth memperlihatkan langkah-langkah dalam menentukan frequent itemset, dimana pertama kali dilakukan penentuan minimum support dari dataset, setelah itu menentukan header frequent itemset dan membuat FP-Tree, kemudian FP-Tree yang ada digunakan untuk membuat Conditional Pattern. Dari Conditional Pattern kemudian dilakukan penentuan Frequent Itemset.

Association Rule Mining didasarkan pada tiga matrik: Support, Confidence dan Lift. Ketiga matrik tersebut berasal dari catatan transaksi untuk bisnis [6].

- a. Support Matrik pertama ditetapkan untuk analisis keranjang pasar adalah Support, yang merupakan probabilitas dari asosiasi (probabilitas dari dua item yang dibeli bersama-sama).
- b. Confidence Confidence dihasilkan dari seberapa kuat hubungan produk yang sudah dibeli.
- c. Lift Lift Ratio mengukur seberapa penting rule yang telah terbentuk berdasarkan nilai support dan confidence.

3. Hasil Evaluasi

Dari berbagai eksperimen yang dilakukan dalam penentuan rekomendasi produk dengan Algoritma FP-Growth, rule yang dihasilkan dengan minimum support 40% dan minimum confidence 80% adalah 3 buah rule dengan masing-masing nilai Lift > 0,8. Adapun rule secara detail dapat dilihat pada tabel 2.

Tabel 2. Rule yang dihasilkan dengan minimum support 40% dan minimum confidence 80%

No	Premises	Conclusion	Lift
1	2pct. Milk	Peaches	0.875
2	White Bread	Peaches	0.888
3	Eggs	Peaches	0.923

4. Kesimpulan

Dari hasil penelitian ini dapat disimpulkan bahwa penerapan Association Rule dengan menggunakan Algoritma FP-Growth dapat membantu penentuan rekomendasi promosi produk dengan tepat. Penelitian ini menghasilkan 3 rule ketika menggunakan minimum support 40% dan minimum confidence 80%. Diharapkan penelitian mendatang dapat melanjutkan dengan menambah preprocessing ataupun algoritma lain dengan tujuan untuk mendapatkan rekomendasi produk yang lebih baik.