

SOLUSI PREDIKSI MAHASISWA *DROP OUT* PADA PROGRAM STUDI SISTEM INFORMASI FAKULTAS ILMU KOMPUTER UNIVERSITAS BINA DARMA

Ade Putra

Fakultas Vokasi, Program Studi Komputerisasi Akuntansi

Universitas Bina Darma

Email: ade.putra@binadarma.ac.id

ABSTRAK

Data mining merupakan salah satu bidang ilmu yang bergerak di bidang penggalian dan pengkajian data, dimana data mining dapat memberikan solusi – solusi yang mampu memberikan pemecahan permasalahan khususnya yang di hadapi oleh program studi Sistem Informasi Fakultas Ilmu Komputer guna menjamin agar mahasiswa Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma dapat lulus dengan tepat waktu . adapun tahapan yang digunakan menggunakan konsep *Knowledge Discovery in Database* (KDD) yang terdiri dari *Selection, Pre – Processing, Transformation, Data Mining dan Interpretation / Evaluation*. Pada penelitian ini digunakan metode *Clasificassion* dengan menggunakan algoritma *Decision Tree* atau *C4.5.*, dimana pada algoritma ini melihat hasil penilaian – penilaian untuk nilai *Entropi* dan *Gain* pada masing – masing atribut, adapun atribut yang menjadi poros penilaian *Entropi* dan *Gain* pada penelitian ini adalah Indeks Prestasi Kumulatif (IPK), Jumlah SKS yang telah ditempuh, Semester dan Status perkuliahan mahasiswa angkatan 2013. Hasil dari penelitian ini adalah didapatnya node 1 yaitu atribut SKS dengan nilai *Gain* terbesar yaitu 0.3276 yang kemudian di ikuti oleh atribut Semester sebagai node 1.1 dengan nilai *Gain* sebesar 0.0874

Kata kunci: *Knowledge Discovery in Database, Clasification dan Algoritma Decesion Tree atau C4.5*

ABSTRACT

Data mining is one of the disciplines engaged in extracting and assessment of data, data mining can provide a solution that is capable of providing solving problems, especially that the study program Information Systems Faculty of Computer Science faced to ensure that all students of Information Systems Faculty Computer Science Universitas Bina Darma can graduate on time. As for the stages used in this research is using the concept of Knowledge Discovery in Databases (KDD), which consists of Selection, Pre - Processing, Transformation, Data Mining and Interpretation / Evaluation. In this study used Clasificassion method and C4.5 or Decision Tree algorithm, where the algorithm see the results of the assessment for the value of Entropy and Gain on each attributes, the attributes that used to assessment Gain and Entropi in this study are Grade Point Average (GPA), number of credits that have been taken, and the students' semester Status that accepted in 2013. The results of this study for node 1 is SKS with the largest Gain value is 0.3276 then followed by a semester as a node 1.1 that has a Gain value is 0.0874.

Keywords: *Knowledge Discovery in Database, Clasification and Algoritma Decesion Tree or C4.5*

1. PENDAHULUAN.

Kemajuan akan ilmu dan pengetahuan pada saat ini sudah tidak dapat terbandung lagi, hal ini merupakan sesuatu yang sangat dibutuhkan guna menjamin kelangsungan hidup manusia dimana peranan akan ilmu dan pengetahuan sangat melekat pada kehidupan sehari – hari. Setiap bidang baik di lingkungan pemerintahan maupun swasta telah banyak menggunakan kemajuan ilmu dan pengetahuan baik yang tercipta dalam bentuk perangkat keras maupun perangkat lunak. Dalam penggunaan perangkat – perangkat ini dapat tercipta suatu sumber data yang sangat besar sesuai dengan masa pakai dari perangkat yang bersangkutan. Data – data yang tercipta dapat dimanfaatkan untuk hal – hal dimasa yang akan datang baik yang bersifat prediksi, klasifikasi maupun dalam mengasosiasikan

sesuatu, hal ini menjadi salah satu faktor pendorong terciptanya suatu bidang ilmu Data Mining. [3] Data Mining merupakan suatu disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data.

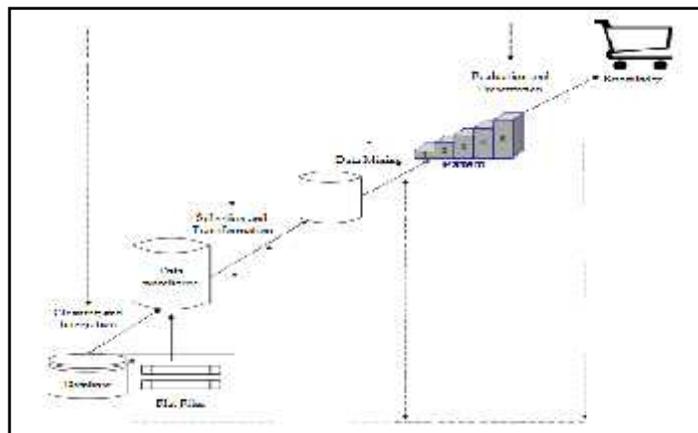
Data mining sering juga disebut dengan *Knowledge Discovery in Databases* (KDD) sebagai suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set yang berukuran besar. Keluaran dari proses data mining ini dapat dipakai untuk memperbaiki pengambilan keputusan – keputusan di masa yang akan datang [4] guna meningkatkan produktivitas perkuliahan dan kualitas lulusan, Program Studi Sistem Informasi Universitas Bina Darma melakukan kajian – kajian terhadap data – data lulusan untuk melakukan peringatan dini akan kendala – kendala yang di hadapi mahasiswa dalam melakukan perkuliahan agar lulus tepat waktu, hal ini di lakukan untuk menghindari adanya mahasiswa yang mengalami *drop out* yang di lihat dari sisi SKS perkuliahan, IPK dan Jumlah Semester yang telah di lalui. Dengan mengetahui kendala akademik dari mahasiswa khususnya pada Program Studi Sistem Informasi diharapkan dapat sesegera mungkin melakukan pembimbingan dengan mahasiswa – mahasiswa yang berkemungkinan mengalami *drop out*. Ada pun tahapan yang di lakukan sesuai dengan proses *Knowledge Discovery in Database* (KDD) antara lain :

1. *Data Selection.*
2. *Preprocessing*
3. *Transformation Data*
4. *Data Mining*
5. *Interpretation / Evaluation.*

2. METODOLOGI

2.1. *Knowledge Discovery in Database* (KDD)

Knowledge Discovery In Database (KDD) merupakan suatu keseluruhan dari proses yang bersifat non -trivial (tidak biasa) guna untuk mencari serta mengidentifikasi pola – pola (*pattern*) yang terjadi pada sekumpulan data, dimana pola yang ditemukan bersifat sah (yang sebenarnya) baru dapat bermanfaat dan dapat dimengerti. Adapun KDD berhubungan dengan hal – hal yang bersifat teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola- pola sejumlah kumpulan data [2].



Gambar 1. Tahapan KDD (Han, et all, 2006)

1. *Data Selection*

- Membuat kumpulan atau himpunan dari data target, pemilihan kumpulan data dengan pola tertentu , atau memfokuskan pada variable data atau sampel, dimana proses kegiatan aktivitas penemuan (*discovery*) dilakukan.
- Pemilihan / seleksi data – data dari sekumpulan data yang memiliki keterkaitan hubungan secara operasional perlu dilakukan terlebih dahulu yang kemudian baru dapat melanjutkan ke tahap penggalian informasi yang ada ke dalam proses KDD. Kumpulan tau himpunan data hasil dari proses seleksi yang akan digunakan pada proses data mining disimpan dalam suatu tempat / berkas yang terpisah dari kumpulan basis data operasional yang menjadi sumber data.

2. *Pre – processing / Cleaning*
 - *Pre – Processing* merupakan operasi yang paling dasar dilakukan seperti menghilangkan *noise* dilakukan. Barulah proses data mining dapat dilaksanakan, perlu dilakukan terlebih dahulu proses *cleaning* pada kumpulan data set yang menjadi data set awal pada proses KDD.
 - Proses *cleaning* meliputi kegiatan antara lain membuang data duplikat serta memeriksa data yang tidak konsisten atau relevan, kemudian memperbaiki kesalahan pada data tersebut.
 - Dilakukan proses *enrichment* / memperkaya data set dengan data lain atau informasi lain (eksternal)
3. *Transformation*
 - Pencarian pola – pola pada data serta hubungan antar fitur yang terdapat pada data / keterkaitan data.
 - Keterkaitan atribut atau fitur yang dipakai untuk menampilkan data bergantung kepada hasil yang ingin di peroleh.
 - Merupakan proses perubahan/penggabungan (transformasi) pada data set yang telah dipilih.
4. Data mining
 - Melakukan proses Pemilihan tugas serta pemilihan hasil dari proses KDD misalnya klasifikasi, regresi, clustering, dll.
 - Pemilihan metode algoritma data mining yang akan dipakai.
 - Melakukan proses Data mining yaitu proses mencari pola - pola atau informasi yang menarik dalam data set yang terpilih dengan menggunakan teknik serta metode tertentu.
5. *Interpretation/ Evaluation*
 - Penerjemahan pola / *pattern* yang terjadi / pola yang dihasilkan dari proses data mining.
 - Pola / *pattern* informasi yang dihasilkan dari proses data mining akan ditampilkan dalam format / bentuk yang mudah dimengerti oleh pihak yang berkepentingan

2.2. Algoritma *Decision Tree* / C4.5

Decision Tree adalah teknik model prediksi yang digunakan untuk melakukan proses klasifikasi, *clustering*, serta proses prediksi. *Decision Tree* menggunakan teknik “membagi dan menaklukkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah [1].

Decision Tree dapat digunakan untuk melakukan proses klasifikasi terhadap sekumpulan objek atau *record*. Teknik ini terdiri dari kumpulan *decision node*, dihubungkan oleh cabang, bergerak ke bawah dari *root node* sampai berakhir di *leaf node* [5].

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data *histori* yang pernah terjadi sebelumnya atau disebut data masa lalu serta telah mengalami pengelompokan ke dalam kelas tertentu.
2. Menghitung akar dari pohon Akar akan diambil dari atribut yang akan terpilih dengan cara menghitung nilai *gain* dari setiap atribut kemudian nilai *gain atribut* yang tertinggi akan menjadi akar pertama atau node pertama. Sebelum menghitung nilai *gain* dari atribut hitung dahulu nilai *entropy* Untuk menghitung nilai *entropy* digunakan rumus.

$$Entropy(S) = \sum_{i=1}^r (p_i * \log_2 p_i)$$

Keterangan :

S= Himpunan kasus n = jumlah partisi S

Pi = proporsi Si terhadap S

Kemudian hitung nilai *gain* menggunakan rumus :

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

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

|Si| = Proporsi Si terhadap S

|S| = jumlah kasus dalam S

3. kemudian untuk tahap selanjutnya mengulangi proses 2 dan 3 hingga semua record terproses. Proses pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang diproses lagi
 - c. Tidak ada *record* yang menjadi cabang yang kosong

Sehingga akan diperoleh nilai *Gain* dari atribut yang paling tertinggi. *Gain* adalah salah satu atribut tiap *node* pada *tree*. atribut dengan nilai *Gain* tertinggi akan dipilih sebagai *test* atribut untuk *node* berikutnya.

3. Hasil dan Pembahasan

3.1. Hasil

1. Data Cleaning (Pembersihan Data)

Pembersihan data sumber yakni membuang data yang tidak konsisten dan *noise / redundancy* data. Adapun kesalahan pada yang dimaksud, sample data yang tidak konsisten tersebut dapat di lihat pada gambar dibawah

2. Data Integration

Tahap *integrasi* data merupakan tahap penggabungan data dari berbagai sumber yang terkait dengan kebutuhan penelitian. *Data set* yang digunakan merupakan data mahasiswa program studi Sistem Informasi Fakultas Ilmu Komputer berasal dari satu sumber yaitu Unit Pelaksana Teknis – Sistem Informasi Manajemen (UPT-SIM), dan dari data yang ada tersebut dapat digunakan untuk pengolahan data mining dengan teknik *Decission Tree* dengan menggunakan algoritma C4.5.

3. Data Selection.

Proses seleksi data dilakukan dengan cara melihat kecenderungan data serta kesesuaian data dengan topic atau judul penelitian yang akan diteliti oleh penulis, ada pun data yang dipakai terdiri dari atribut Nim, Nama, Program Studi, Status, IPK, Sks dan Semester yang di peroleh dari Unit Pelaksana Teknis – Sistem Informasi Manajemen (UPT-SIM).

No	CourseName	Status	IPK	SRS	Semester
1	1 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
2	2 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
3	3 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
4	4 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
5	5 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
6	6 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Tidak Baik
7	7 Sistem Informasi	DROP OUT	Sangat Memburuk	Sangat Baik	Tidak Baik
8	8 Sistem Informasi	DROP OUT	Memburuk	Tidak Baik	Baik
10	9 Sistem Informasi	TIDAK DROP OUT	Sangat Memburuk	Baik	Baik
11	10 Sistem Informasi	TIDAK DROP OUT	Sangat Memburuk	Sangat Baik	Baik
12	11 Sistem Informasi	TIDAK DROP OUT	Sangat Memburuk	Baik	Baik

Gambar 2. Data Set Selection

4. Data Transformasi

Pada tahapan ini atribut Ipk, Sks, Status dan Semester akan diberi label mengikuti kondisi data data yang pada atribut tersebut sebagai berikut :

1. Label Ipk.

Pada atribut ipk dilakukan proses pengelompokan pada tingkatan skala yang dapat dilihat pada tabel 1 di bawah sebagai berikut:

Tabel 1. Label Atribut IPK

No	Keterangan	Skala Nilai
1	Memuaskan	2,25 - 2,75
2	Sangat Memuaskan	2,76 - 3,50
3	Cumlaude	3,51 - 4,00

2. Label SKS

Pada atribut sks akan dilakukan proses label dengan melihat jumlah sks yang telah di tempuh oleh mahasiswa, ada pun skala jumlah SKS untuk proses label dapat di lihat pada tabel 2 dibawah :

Tabel 2. Label Atribut Sks

No	Keterangan	Skala SKS
1	Sangat Baik	100 - 130 sks
2	Baik	50 - 99 sks
3	Tidak Baik	1-40 ks

3. Label Semester

Adapun label yang akan diberikan pada atribut semester dapat dilihat pada tabel 3 berikut :

Tabel 3. Label Atribut Semester

No	Label	Keterangan
1	Baik	Mahasiswa dengan nim tanpa akhiran P
2	Tidak Baik	Mahasiswa dengan nim akhiran P

4. Label *StatusDesk*

Pada atribut *statusdesk* dilakukan klasifikasi sebagai berikut: AKTIF, keluar, TANPA KABAR, STOP OUT, ALUMNI. Atribut *StatusDesk* ini akan dijadikan *class* yang merupakan parameter dalam proses pembentukan pohon keputusan *Decision Tree* dengan menggunakan algoritma C4.5 yang merupakan tolak ukur untuk menentukan mahasiswa dengan status *Drop Out* atau Tidak *Drop Out*. Adapun proses label dapat dilihat pada tabel 4 dibawah.

Tabel 4. Label atribut *statusdesk*

No	Label	Nilai Atribut
1	Tidak <i>Drop Out</i>	Aktif
2	<i>Drop Out</i>	keluar
		Tanpa Kabar
		Alumni
		Stop Out

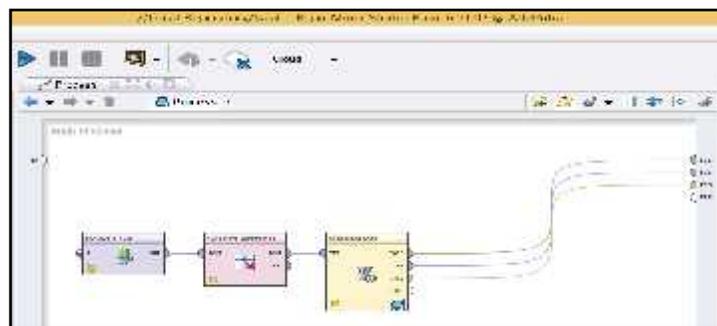
Setelah selesai seluruh tahapan persiapan data set, adapun hasil dari proses tersebut dapat dilihat pada gambar dibawah yang merupakan data set yang disimpan dalam format CSV yang dapat di lihat pada Gambar 3 dibawah sehingga dapat di proses oleh rapidminer:

No	Id	Nama	Alamat	Jenis	Gender
1	111123456	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
2	111123457	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
3	111123458	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
4	111123459	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
5	111123460	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
6	111123461	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
7	111123462	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
8	111123463	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
9	111123464	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
10	111123465	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
11	111123466	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
12	111123467	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
13	111123468	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
14	111123469	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
15	111123470	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
16	111123471	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
17	111123472	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
18	111123473	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
19	111123474	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
20	111123475	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
21	111123476	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
22	111123477	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
23	111123478	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
24	111123479	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita
25	111123480	WALYATI, WATI	Perumahan Bumi Sakti	Perumahan	Wanita

Gambar 3. Data Set Siap Olah

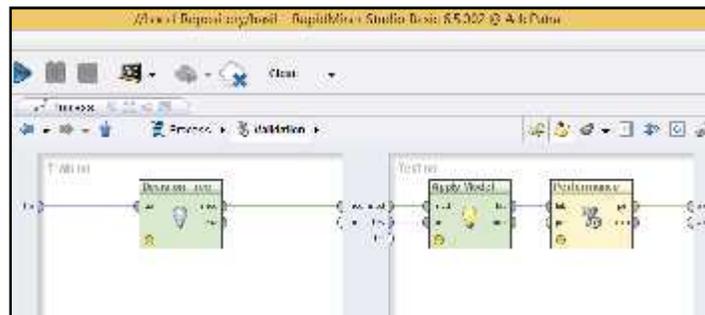
3.2. Model Rapidminer

Adapun tampilan dari struktur model data mining yang di gunakan terdiri dari beberapa objek / operator antara lain *read file CSV*, *selected atribut* dan *Cros Validation* yang tampil pada proses utama untuk menjalankan proses data mining. Bentuk dari model Rapidminer dapat di lihat pada gambar 4 sebagai berikut :



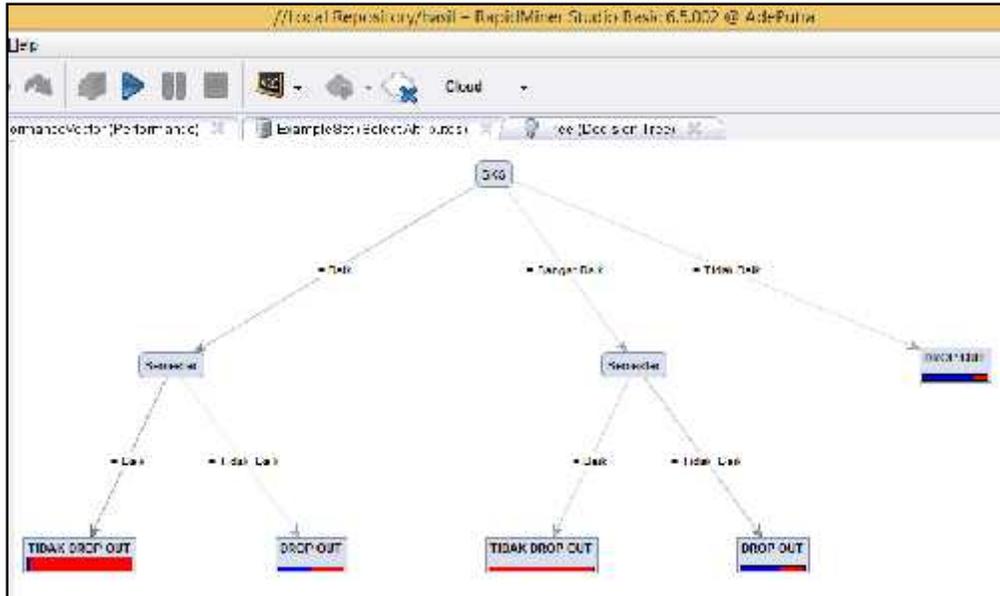
Gambar 4. Rapidminer Model Proses Clasification Decision Tree

Pada objek / operator *Cross Validation* terdapat sub proses yang terdiri dari operator *Decision Tree* sebagai algoritma yang di gunakan pada proses data mining *Classification* yang akan di lakukan, operator *Apply Model* dan operator *Performance* sebagai operator untuk menghasilkan pengolahan data mining berupa *Decision Tree*. Adapun model sub proses dapat di lihat pada gambar 5 sebagai berikut:



Gambar 5. Rapidminer Model Sub Proses Clasification Decision Tree

Ada pun hasil dari model Rapidminer diatas menghasilkan tampilan *Decision Tree* yang dapat dilihat pada gambar 6 sebagai berikut :



Gambar 6. *Decision Tree* Prediksi Mahasiswa *Drop Out* Program Studi Sistem Informasi

3.3. PEMBAHASAN

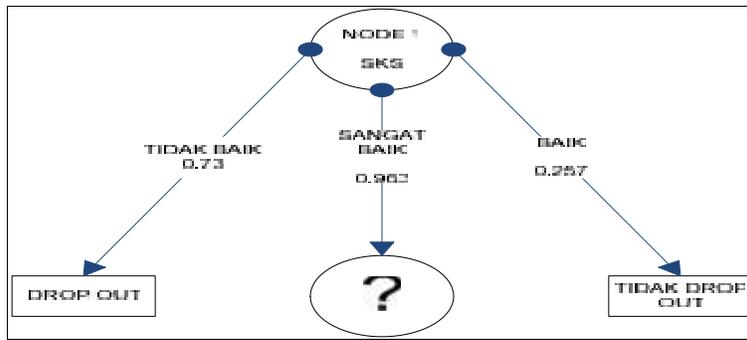
3.3.1. Perhitungan Tabel berdasarkan *Node* Atribut

Tabel perhitungan *node* akan dibentuk berdasarkan data yang telah siap yang nantinya akan diuji dengan menggunakan *software* data mining, berikut ini akan dilakukan proses secara teoritis perhitungan *entropy* dan *Gain* dari masing – masing atribut dan nilai atribut :

Tabel 5. Perhitungan *Entropy* dan *Gain* Node 1

<i>NODE</i>	Keterangan		Jumlah Kasus	<i>Drop Out</i>	Tidak <i>Drop Out</i>	<i>Entropy</i>	<i>Gain</i>
1			287	69	218	0.7957	
	IPK	<i>Cumlaude</i>	19	5	14	0.8315	0.0742
		Sangat Memuaskan	188	27	161	0.5936	
		Memuaskan	80	37	43	0.9959	
	SKS						0.3276
		Sangat Baik	53	22	31	0.963	
		Baik	185	8	177	0.257	
		Tidak Baik	49	39	10	0.73	
	SEMESTER						0.0938
		Baik	245	42	203	0.661	
		Tidak Baik	42	27	15	0.9403	

Pada hasil perhitungan untuk *node 1* dapat dilihat nilai dari *Gain* SKS merupakan nilai tertinggi dibandingkan dengan *Gain* IPK dan Semester sehingga SKS dapat dijadikan *Node 1* dalam pembentukan Pohon Keputusan / *Decession Tree*



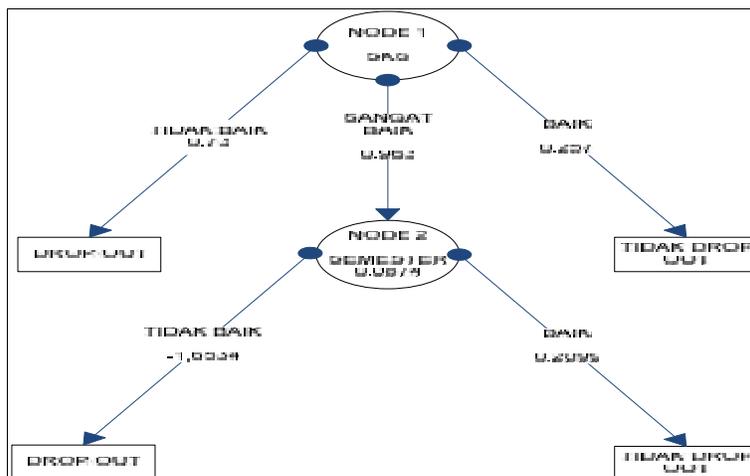
Gambar 7. Pohon Keputusan / Decision tree Node 1

Untuk proses selanjutnya dalam menentukan *node* berikutnya dapat dilakukan dengan cara yang sama sampai pada *node* terakhir.

Tabel 6. Perhitungan Entrophy dan Gain Node 1.1

NODE	Keterangan		Jumlah Kasus	Drop Out	Tidak Drop Out	Entropy	Gain
1.1	SKS		185	8	177	0.257	
	IPK						-0.0328
		Cumlaude	11	6	31	0.6395	
		Sangat Memuaskan	137	2	135	0.2186	
		Memuaskan	37	0	11	0	
	SEMESTER						0.0874
		Baik	181	6	175	0.2099	
		Tidak Baik	4	8	2	-1.6534	

Dari hasil perhitungan node 1.1 dapat dilihat nilai dari *Gain* Semester merupakan nilai *Gain* tertinggi di bandingkan dengan nilai *Gain* IPK, sehingga Semester dapat dijadikan *Node* berikutnya



Gambar 8. Pohon Keputusan / Decision tree Node 1.1

4. KESIMPULAN

Dengan memperhatikan bentuk dari pohon keputusan diatas dapat diketahui bahwa semua atribut sudah masuk kedalam class pada pohon keputusan, setelah pohon keputusan terbentuk dapat ditarik kesimpulan sebagai berikut :

“JIKA Sks = Tidak Baik maka *Class* = “ Drop Out ”

“JIKA Sks = Baik maka *Class* = “ Tidak Drop Out ”

“JIKA Sks = Sangat Baik dan Semester = Baik maka *Class* =” Tidak Drop Out ”

“JIKA Sks = Sangat Baik dan Semester = Tidak Baik maka *Class* = “ Drop Out ”

Sehingga pihak Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma dapat melakukan pembimbingan terhadap mahasiswa yang memiliki jumlah SKS yang Tidak baik dan jumlah Semester yang Tidak Baik sehingga mahasiswa yang bersangkutan dapat lulus dengan tepat waktu.

DAFTAR PUSTAKA

- [1] Dunham, Margareth H. (2003) .*Data Mining Introductory and Advanced Topics*. New Jersey: Prentice Hall
- [2] Fayyad, Usama. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [3] Han, Jiawei dan Kamber, Micheline.(2006), *Data Mining : Concept and Techniques Second Edition*, Morgan Kaufmann Publishers.
- [4] Santosa, Budi (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.
- [5] Yusuf, W. Y. (2007). Perbandingan Performansi Algoritma Decision Tree C5.0, CART dan CHAID : Kasus Prediksi Status Resiko Kredit di Bank X. Seminar SNATI.
<http://journal.uii.ac.id/index.php/Snati/article/view/1628> diakses tanggal 25 Januari 2017