

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

Ade Putra

Fakultas Vokasi, Program Studi Komputerasi Akuntansi
Universitas Bina Darma
Email: ade.putra@binadarma.ac.id

ABSTRAK

Data mining merupakan salah satu pengetahuan yang bergerak di bidang penggalian dan pengkajian data, dimana data mining mampu memberikan solusi dalam 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 yaitu 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 algoritma *Decision Tree* atau *C4.5*., Pada algoritma ini, hasil penilaian yang dipakai untuk menentukan *Node* sebagai kunci dalam menilai kelayakan mahasiswa yang *Drop Out* dilihat dari nilai *Entropi* dan *Gain* pada masing – masing *attribute*., Adapun *attribute* yang digunakan untuk penilaian *Entropi* dan *Gain* pada penelitian ini adalah Indeks Prestasi Kumulatif (IPK), Jumlah SKS yang telah ditempuh, Semester dan Status perkuliahan mahasiswa angkatan 2013. Pada penelitian ini *attribute* SKS ditetapkan sebagai node 1 dengan nilai *Gain* terbesar yaitu 0.3276, yang kemudian di ikuti oleh *attribute* 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 a knowledge which is engaged in the extraction and analysis of data, where data mining can provide solutions in solving problems, especially those faced by the study program Information Systems Faculty of Computer Science, to ensure that students of Information Systems Faculty of Computer Science University of Bina Darma can be graduated 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, In this algorithm, the results of assessments are used to determine the node as a key to assess the feasibility of students that can be drop out views on the value of Entropy and Gain on each attribute, the attributes used for the assessment of Entropy and Gain in this study is the grade point average (GPA), number of credits that have been taken, and the Status of students' Semester force in 2013. In this study, SKS attribute set as node 1 with the largest Gain value is 0.3276, then followed by Semester attribute as node 1.1 with a Gain value of 0.0874

Keywords: *knowledge discovery in database*, *clasification* and algoritma *decesion tree* or *C4.5*.

1. PENDAHULUAN

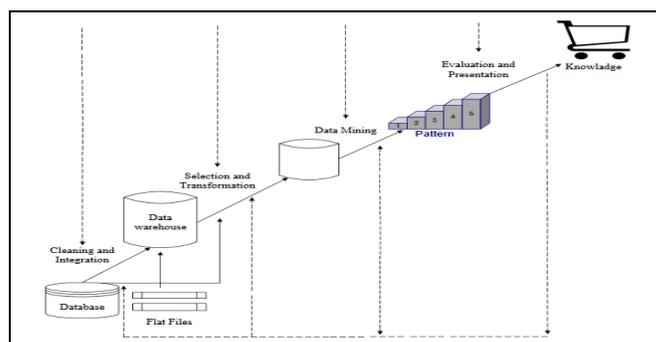
Kemajuan akan ilmu dan pengetahuan pada saat ini sudah tidak dapat terbendung lagi, hal ini merupakan sesuatu yang sangat dibutuhkan guna menjamin kelangsungan hidup manusia dimana peranan akan ilmu dan pengetahuan sangat melekat pada kehidupan sehari – hari. Setiap bidang baik di lingkungan pemerintahan 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 mengestrak 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*.

2. METODOLOGI

2.1 *Knowledge Discovery in Database* (KDD)

KDD merupakan suatu pola kegiatan *non -trivial* (tidak biasa) guna untuk mencari serta mengidentifikasi 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]. Adapun tahapan dari KDD dapat dilihat pada Gambar 1 dibawah:



Gambar 1. Tahapan KDD [3]

1. *Data Selection*
 - a) 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.
 - b) 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 taau 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*
 - a) *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.
 - b) Proses *cleaning* meliputi kegiatan antara lain membuang data duplikat serta memeriksa data yang tidak konsisten atau relevan, kemudian memperbaiki kesalahan pada data tersebut.
 - c) Dilakukan proses *enrichment* / memperkaya data set dengan data lain atau informasi lain (eksternal)
3. *Transformation*
 - a) Pencarian pola – pola pada data serta hubungan antar fitur yang terdapat pada data / keterkaitan data.
 - b) Keterkaitan *attribute* atau fitur yang dipakai untuk menampilkan data bergantung kepada hasil yang ingin di peroleh.

- c) Merupakan proses perubahan/penggabungan (transformasi) pada data set yang telah dipilih.
- 4. Data mining
 - a) Melakukan proses Pemilihan tugas serta pemilihan hasil dari proses KDD misalnya klasifikasi, *regresi*, *clustering*, dll.
 - b) Pemilihan metode algoritma data mining yang akan dipakai.
 - c) 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*
 - a) Penerjemahan pola / *pattern* yang terjadi / pola yang dihasilkan dari proses data mining.
 - b) 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 terdiri dari beberapa *decision node* yang 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* merupakan data – data 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 *attribute* yang akan terpilih dengan cara menghitung nilai *gain* dari setiap *attribute* kemudian nilai *gain attribute* yang tertinggi akan menjadi akar pertama atau node pertama. Sebelum menghitung nilai *gain* dari *attribute* hitung dahulu nilai *entropy* Untuk menghitung nilai *entropy* digunakan Rumus 1 dibawah..

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

Keterangan :

S= Himpunan kasus n = jumlah partisi S

Pi = proporsi Si terhadap S

Kemudian hitung nilai *gain* menggunakan Rumus 2 dibawah :

$$\text{Gain}(S, A) = S - \sum_{i=1}^n \frac{|S_i|}{|S|} * S_i \quad (2)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi *attribute* 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 nilai N mendapat kelas yang sama.
 - b. Tidak ada *attribute* di dalam *record* yang diproses lagi
 - c. Tidak ada *record* yang menjadi cabang yang kosong

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

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 *attribute* Nim, Nama, Program Studi, Status, IPK, Sks dan Semester yang di peroleh dari Unit Pelaksana Teknis – Sistem Informasi Manajemen (UPT-SIM). Adapun data Set dapat dilihat pada Gambar 2 dibawah :

No	CourseName	Statusdesk	IPK	Sks	Semester
1	1 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
2	2 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
3	3 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
4	4 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
5	5 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
6	6 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
7	7 Sistem Informasi	DROP OUT	Sangat Memuaskan	Sangat Baik	Tidak Baik
8	8 Sistem Informasi	DROP OUT	Memuaskan	Tidak Baik	Baik
9	9 Sistem Informasi	TIDAK DROP OUT	Sangat Memuaskan	Baik	Baik
10	10 Sistem Informasi	TIDAK DROP OUT	Sangat Memuaskan	Sangat Baik	Baik
11	11 Sistem Informasi	TIDAK DROP OUT	Sangat Memuaskan	Baik	Baik

Gambar 2. Data Set Selection

4) *Data Transformasi*

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

a. Label Ipk

Pada *attribute* ipk dilakukan proses pengelompokan pada tingkatan skala yang dapat dilihat pada Tabel 1 di bawah sebagai berikut:

Tabel 1. Label *attribute* IPK

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

b. Label SKS

Pada *attribute* 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 *attribute* SKS

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

- c. Label Semester
Adapun label yang akan diberikan pada *attribute* semester dapat dilihat pada Tabel 3 berikut :

Tabel 3. Label *attribute* semester

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

- d. Label *Status Desk*
Pada *attribute status desk* dilakukan klasifikasi sebagai berikut: AKTIF, keluar, TANPA KABAR, STOP OUT, ALUMNI. *Attribute StatusDesk* ini akan dijadikan *class* yang merupakan parameter dalam proses pembentukan pohon keputusan *Decission 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 *attribute statusdesk*

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

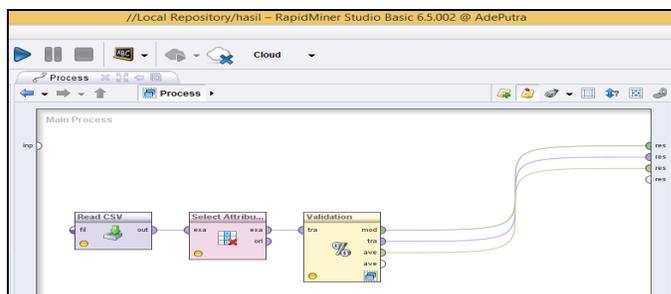
Setelah selesai seluruh tahapan persiapan data set, adapun hasil dari proses tersebut dapat dilihat pada Gambar 3 dibawah yang kemudian disimpan dalam format CSV sehingga dapat di proses oleh *Rapidminer*:

No	UserCode	FullName	CourseName	Statusdesk	IPK	SKS	Semester
1	13142094	RIYO AJI SAPUTRA RIO	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
2	13142027	WASTU SATRYA ERIYANTO	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
3	13142080	MUHAMMAD ALQADRI SEGARA	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
4	13142114	HARDI MASTION	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
5	13142136	M.AFRIADI SETIAWAN	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
6	13142292P	YAYAN MARDANA	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Tidak Baik
7	13142264P	DINA KARLINA A A	Teknik Informatika	DROP OUT	Sangat Memuaskan	Sangat Baik	Tidak Baik
8	13142239	ARVIJUANDI G PRATAMA	Teknik Informatika	DROP OUT	Memuaskan	Tidak Baik	Baik
9	13142219	M FIKRI HADZIQ	Teknik Informatika	TIDAK DROP OUT	Sangat Memuaskan	Baik	Baik
10	13142005	SULISTIYONO	Teknik Informatika	TIDAK DROP OUT	Sangat Memuaskan	Sangat Baik	Baik
11	13142315	RAHMI HAYATI	Teknik Informatika	TIDAK DROP OUT	Sangat Memuaskan	Baik	Baik

Gambar 3. Data Set Siap Olah

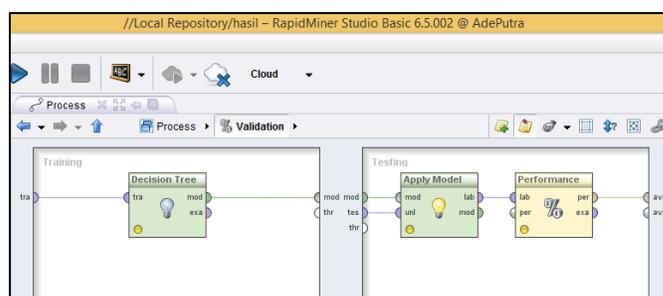
3.2 Model *Rapidminer*

Adapun tampilan dari struktur model data mining yang di gunakan terdiri dari beberapa objek / *operator* antara lain *read file CSV*, *selected attribute* 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 :



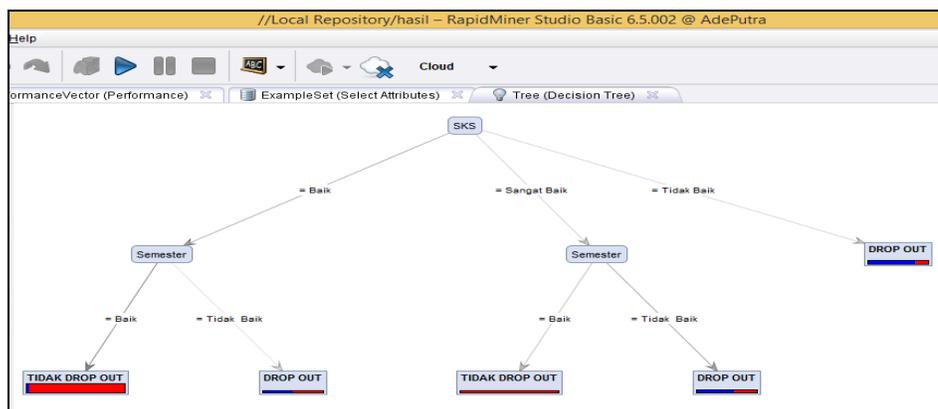
Gambar 4. Rapidminer Model Proses *Classification Decision Tree*

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 *Classification 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

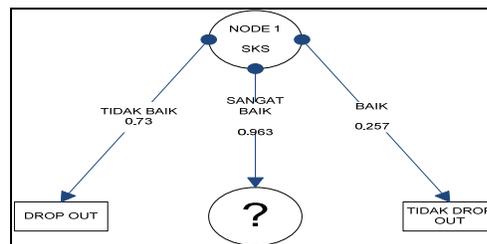
3.3.1 Perhitungan Tabel berdasarkan Node Attribute

Tabel perhitungan *Node* akan dibentuk berdasarkan data yang telah siap yang nantinya akan diuji dengan menggunakan *software* data mining, berikut pada Tabel 5 akan dilakukan proses secara teoritis perhitungan *Entropy* dan *Gain* dari masing – masing *Attribute* untuk menentukan Node 1:

Tabel 5. Perhitungan *entropy* dan *gain* node 1

<i>NODE</i>	<i>Keterangan</i>	<i>Jumlah Kasus</i>	<i>Drop Out</i>	<i>Tidak Drop Out</i>	<i>Entropy</i>	<i>Gain</i>	
1	IPK	<i>Cumlaude</i>	19	5	14	0.8315	0.0742
		Sangat Memuaskan	188	27	161	0.5936	
	SKS	Memuaskan	80	37	43	0.9959	0.3276
		Sangat Baik	53	22	31	0.963	
		Baik	185	8	177	0.257	
	SEMESTER	Tidak Baik	49	39	10	0.73	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*. Adapun pembentukan *Node 1* dapat dilihat pada Gambar 7 dibawah :



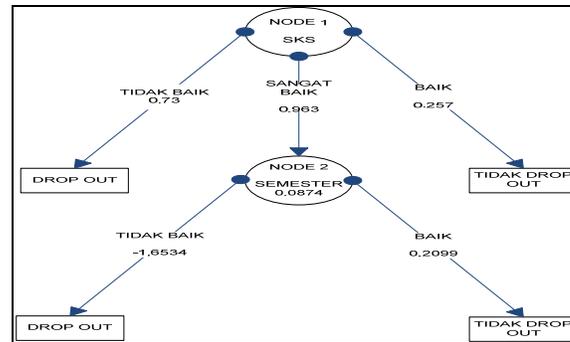
Gambar 7. Pohon Keputusan / *Decision Tree* Node 1

Untuk proses selanjutnya dalam menentukan *Node* berikutnya dapat dilakukan dengan cara yang sama sampai pada *Node* terakhir. Pada Tabel 6 merupakan pembentukan *Node 1.1* :

Tabel 6. Perhitungan *Entropy* dan *Gain* Node 1.1

<i>NODE</i>	<i>Keterangan</i>	<i>Jumlah Kasus</i>	<i>Drop Out</i>	<i>Tidak Drop Out</i>	<i>Entropy</i>	<i>Gain</i>
1.1	SKS	185	8	177	0.257	-
	IPK					
	<i>Cumlaude</i>	11	6	31	0.6395	0.0874
	Sangat Memuaskan	137	2	135	0.2186	
	Memuaskan	37	0	11	0	
SEMESTER	Baik	181	6	175	0.2099	0.0874
	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 dapat dilihat pada Gambar 8 :



Gambar 8. Pohon Keputusan / Decision Tree Node 1.1

4. KESIMPULAN

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

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

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

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

“JIKI 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.uui.ac.id/index.php/Snati/article/view/1628> diakses tanggal 25 Januari 2017.