

Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus: Fakultas Teknik dan Komputer Universitas Harapan Medan)

Rismayanti

Program Studi Teknik Informatika Universitas Harapan Medan Sumatera Utara
risma.stth@gmail.com

Abstract

University databases such as technology and computer department at Harapan university of Medan keeps academic, administrative and students' data in large volume. If the data is mined properly then it has been known the pattern or knowledge to make a decision. By having the data of Informatics Department graduated students, the information that obtained is the decision tree with the rules to predict the studying period of the informatics Department students at Harapan University of Medan. The data used for designing the decision tree consists of 6 predictor attributes (Gender, Region, Specialization, School, Parrents' occupation and GPA) and 2 target attributes (Passed "On Time" and "Late"). Data analysis used C4.5 algorithm, it used 30 samples taken randomly from alumni data. The research results show the decision tree model with the rules of the students who pass "On Time" is students who have a high GPA and students who have a medium GPA with the focus robotics.

Keywords : Data Mining, C4.5 Algorithm, Decision Tree.

1. PENDAHULUAN

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan masa studi secara tepat waktu. Tingginya persentasi mahasiswa yang tidak mampu menyelesaikan masa studinya secara tepat waktu maka akan menyebabkan berbagai masalah yang berkaitan dalam kegagalan studi mahasiswa seperti meningkatnya status mahasiswa non aktif di lingkungan perguruan tinggi. Program sarjana (S1) fakultas teknologi dan komputer di Universitas Harapan Medan adalah program pendidikan akademik dalam bidang teknik yang meliputi teknik informatika, sistem informasi, teknik sipil, teknik industri, teknik mesin dan teknik elektro. Dan untuk menyelesaikan program akademik, mahasiswa dijadwalkan 8 semester untuk menyelesaikan masa studi secara tepat waktu. Setiap perguruan tinggi perlu melakukan prediksi untuk memperoleh informasi yang tepat serta akurat tentang mahasiswa yang mampu menyelesaikan masa studi secara tepat waktu sehingga dapat mencegah secara dini kegagalan akademik mahasiswa.

Database perguruan tinggi menyimpan data akademik, administrasi dan biodata mahasiswa. Data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan. Serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan *Data Mining* [3]. *Data Mining* memecahkan masalah dengan menganalisis data yang telah ada dalam *database* sehingga menghasilkan keputusan. *Data Mining* merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut [6]. *Data Mining* juga diartikan sebagai sebuah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu [2]. *Data Mining* di sisi lain adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar[8].

Data Mining sendiri memiliki beberapa teknik salah satunya klasifikasi. Klasifikasi merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan [5]. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian

digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya[5]. Teknik klasifikasi terbagi menjadi beberapa teknik yang diantaranya adalah decision tree.

Decision tree adalah struktur *flowchart* yang mempunyai *tree* (pohon), dimana setiap simpul *internal* menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas [9]. Alur pada *decision tree* ditelusuri dari simpul ke akar ke simpul daun yang memegang prediksi kelas. *Decision tree* adalah salah satu metode yang digunakan untuk pengklasifikasian dan prediksi karena memiliki kemudahan dalam interpretasi hasil [4]. *Decision tree* dan algoritma C4.5 merupakan dua model yang tidak terpisahkan, oleh karena itu untuk membangun sebuah *decision tree*, dibutuhkan algoritma C4.5 [1]. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan C4.5 adalah bisa mengatasi *missing value*, *continue data* dan *pruning*. Algoritma C4.5 mempunyai *input* berupa *training samples* dan *samples*, *training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya, sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data [7].

Berdasarkan pembahasan di atas, peneliti bermaksud untuk melakukan analisis algoritma C4.5 untuk Memprediksi Masa Studi Mahasiswa program studi Teknik Informatika di Universitas Harapan Medan. Hasil analisis dan implementasi nantinya akan diperoleh keputusan yang tepat dan akurat dalam memprediksi masa studi mahasiswa berdasarkan analisis algoritma C4.5 yaitu mahasiswa yang lulus “tepat waktu” dan “tidak tepat waktu”.

2. METODE PENELITIAN

2.1 Analisis Pengumpulan Data

Dalam penelitian ini, dilakukan pengumpulan data dengan mengambil objek penelitian pada data mahasiswa alumni prodi teknik informatika di Fakultas Teknologi dan Komputer Universitas Harapan. Data yang diperoleh dalam penelitian ini merupakan data yang diperoleh secara angket yang diisi oleh mahasiswa alumni dan data yang diperoleh dari pihak bagian informasi di prodi teknik informatika.

2.2 Analisis Data

Untuk menentukan masa studi mahasiswa, pada penelitian ini akan dilakukan analisis menggunakan *Algoritma C4.5*, sehingga perlu dilakukan proses analisa data terlebih dahulu sebelum data yang digunakan tersebut di *mining*. Data yang akan digunakan dan dianalisa, yaitu data mahasiswa alumni prodi teknik informatika di Fakultas Teknologi dan Komputer Universitas Harapan Medan. Dalam menganalisa data mahasiswa alumni tersebut, maka ada atribut yang digunakan untuk menentukan masa studi mahasiswa yaitu; NPM, Nama, Jenis Kelamin, Asal Daerah, Peminatan, Asal Sekolah, Pekerjaan Orang Tua, IPK dan Tahun Lulus. Beberapa dari sembilan atribut tersebut akan dijadikan atribut prediktor atau atribut *input* untuk menghasilkan atribut target, di mana atribut target tersebut menjadi *class output* untuk menentukan masa studi yaitu dibedakan menjadi 2 *class* (Tepat Waktu dan Tidak Tepat Waktu). Berikut ini adalah keterangan data mahasiswa yang akan digunakan dalam menentukan masa studi mahasiswa terlihat pada tabel 1.

Tabel 1 Keterangan Data Mahasiswa Alumni

No	AtributPrediktor/Input	Keterangan
1.	NPM	Nomor pokok mahasiswa.
2.	Nama	Nama mahasiswa.
3.	Jenis Kelamin	Laki-laki/Perempuan.
4.	Asal Daerah	Asal domisili daerah mahasiswa(Medan/Luar Medan).
5.	Peminatan	Merupakan kelas peminatan yang ada di jurusan teknik informatika yaitu robotika, multimedia dan jaringan.
6.	Asal Sekolah	Asal sekolah dari mahasiswa yaitu SMA dan SMK
7.	Pekerjaan Orang Tua	Pekerjaan orang tua siswa, terbagi 3 yaitu Pegawai Swasta, PNS dan lainnya(guru,petani,polisi dst).
8.	IPK	Merupakan indeks prestasi kumulatif mahasiswa.
9.	Tahun Lulus	Merupakan tahun kelulusan mahasiswa yang menentukan mahasiswa lulus tepat waktu dan tidak tepat waktu.

Berdasarkan data mahasiswa alumni, kemudian dilakukan pemilihan atribut dan sebagian dari data dalam atribut yang ada akan ditransformasikan untuk memudahkan proses *mining* dalam menentukan masa studi. Adapun data mahasiswa alumni yang digunakan untuk menentukan masa studi mahasiswa yaitu 30 samples data mahasiswa alumni prodi teknik informatika yang diambil secara acak dari data angket yang diisi oleh mahasiswa alumni (telah dijelaskan pada analisis pengumpulan data), terlihat pada tabel 2.

Tabel 2. Samples Data Mahasiswa Alumni Prodi Teknik Informatika

NO	NPM	NAMA	JENIS KELAMIN	ASAL SEKOLAH	ASAL DAERAH	PEMINATAN	PEKERJAAN ORANG TUA	IPK	TAHUN LULUS
1.	12351016	Alfi Sulthani Putra	Laki-Laki	SMA	Medan	Jaringan	Pegawai Swasta	3.00	2017
2.	05351076	Sumba di Yuma dila Nst	Laki-Laki	SMA	P. Sidempuan	Multimedia	Pegawai Swasta	3.00	2017
3.	11350453	Anef Sukma Hidayat	Laki-Laki	SMA	Medan	Multimedia	Pegawai Swasta	3.00	2017
4.	11350042	Rizky Nova Hasian	Laki-Laki	SMK	Tap. Selatan	Multimedia	Petani	3.00	2017
5.	12350250	Zulham Prabowo Hadi	Laki-Laki	SMK	Siantar	Multimedia	PNS	3.25	2017
6.	12351241	Andy Armansyah	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	3.00	2017
7.	12350503	Novita Sari Sanusi Hrp	Perempuan	SMA	Medan	Jaringan	Pegawai Swasta	3.00	2017
8.	11351245	Intan Pumama Sari	Perempuan	SMA	Medan	Multimedia	Pegawai Swasta	3.01	2016
9.	11351112	Supri Saftani	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	3.34	2015
10.	12351285	Akhmad Afandi DD	Laki-Laki	SMK	Medan	Multimedia	PNS	3.10	2017
11.	09451498	Ahmad Rivai	Laki-Laki	SMK	Medan	Multimedia	PNS	3.20	2017
12.	12350324	Nurul Sakinah	Perempuan	SMK	Galang	Multimedia	Pegawai Swasta	3.62	2016
13.	12350379	Ilham Pahlevi	Laki-Laki	SMA	Lubuk Pakam	Multimedia	Wirausaha	3.51	2016
14.	12350248	Siti Nur Zaenab	Perempuan	SMK	Medan	Jaringan	Pegawai Swasta	3.80	2016
15.	12351433	Purnomo	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	3.59	2016
16.	10351193	Adlis Legan	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	3.18	2016
17.	11351436	Ardian Iskandar	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	3.27	2016
18.	11351405	Hasyufa Syakim	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	3.21	2016
19.	11350085	Haryan Bima Abdi	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	3.03	2016
20.	12350108	Ofdy Pramana	Laki-Laki	SMA	Medan	Multimedia	Wirausaha	3.55	2016
21.	12351095	Ricky Mardohar S.	Laki-Laki	SMA	Medan	Multimedia	Wirausaha	3.52	2016
22.	12351210	M. Ade Setiawan	Laki-Laki	SMA	Medan	Multimedia	Wirausaha	5.58	2016
23.	12350244	Bayu Fitra Ramadhan	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	3.41	2016
24.	11351104	M. Yogi Syahputra	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	3.39	2015
25.	13350124	Dinda Novita	Perempuan	SMK	Medan	Robotika	Pegawai Swasta	3.64	2017
26.	13350123	Sylvira Arvianti	Perempuan	SMA	Medan	Multimedia	Pegawai Swasta	3.70	2017
27.	11350238	Hendra Gusniawan	Laki-Laki	SMK	Tebing Tinggi	Multimedia	PNS	3.00	2016
28.	11351399	Ahmad Nugroho Nst	Laki-Laki	SMK	Medan	Multimedia	PNS	3.00	2016
29.	11351105	Dika Tamizi	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	3.34	2015
30.	11351204	Damar Wiradana	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	3.25	2015

2.3 Cleaning Data

Pada penelitian ini, data yang digunakan adalah data alumni di prodi teknik informatika. Berdasarkan proses *cleaning*, maka data yang digunakan sebagai atribut model *decision tree* yang dirancang menggunakan algoritma *C4.5* adalah Jenis Kelamin, Asal Daerah, Peminatan, Asal Sekolah, Pekerjaan Orang Tua dan IPK. Keenam atribut tersebut digunakan sebagai atribut prediktor atau *input*. Sedangkan untuk atribut kelas atau *output* dapat ditentukan berdasarkan NPM dan tahun lulus mahasiswa alumni tersebut.

2.4 Transformasi Data

Setelah proses *cleaning* dan itegrasi data, maka langkah selanjutnya adalah melakukan proses transformasi data. Berdasarkan data alumni tersebut, data yang akan ditransformasi adalah Jenis Kelamin, Asal Daerah, Peminatan, Asal Sekolah, Pekerjaan Orang Tua dan IPK. Data tersebut merupakan data yang akan dijadikan atribut *input* untuk merancang model *decision tree* menggunakan algoritma *C4.5*. Proses transformasi dilakukan dengan cara membuat klasifikasi pada masing-masing atribut *input*, seperti terlihat pada Tabel 3.

Tabel 3. Klasifikasi pada Atribut *Input*

No.	Atribut	Klas
1.	IPK	< 3,00 (Rendah)
		3,00 <= IPK < 3,50 (Sedang)
		>= 3,50 (Tinggi)
2.	Jenis Kelamin	Laki-Laki
		Perempuan
3.	Asal Daerah	Medan
		Luar Medan
4.	Asal Sekolah	SMA
		SMK
5.	Pekerjaan Orang Tua	Pegawai Swasta(PS)
		PNS
		Lainnya
6.	Peminatan	Jaringan
		Multimedia
		Robotika

Pada tabel 3 menjelaskan bahwa keenam atribut yang digunakan sebagai *input* sudah memiliki klasifikasi masing-masing sehingga memudahkan dalam proses *mining* untuk menentukan masa studi mahasiswa berdasarkan keenam atribut tersebut.

Enam atribut prediktor tersebut akan di analisa menggunakan algoritma *C4.5* untuk menentukan tepat waktu atau tidak tepat waktunya masa studi mahasiswa, sesuai dengan atribut target dari data yaitu pada tabel 4.

Tabel 4. Atribut Target

Atribut Target	Keterangan
Tepat Waktu	Masa Studi <= 4 Tahun
Tidak Tepat Waktu	Masa Studi > 4 Tahun

Dari tabel 4 di atas dapat dijelaskan bahwa terdapat 2 klasifikasi dari atribut target yang diinginkan yaitu Tepat Waktu dan Tidak Tepat Waktu sebagai penentu masa studi mahasiswa.

Data yang akan digunakan untuk merancang *decision tree* terdiri dari 6 atribut *input* dan 1 atribut target. Untuk analisis data menggunakan algoritma *C4.5*, akan digunakan 30 sample data yang diambil secara acak dari data alumni, setelah itu 30 sample data alumni tersebut ditransformasi sehingga menghasilkan keputusan (lulus tepat waktu dan tidak tepat waktu) seperti terlihat pada tabel 5.

Tabel 5. Transformasi Data Mahasiswa Alumni

NO	NPM	JENIS KELAMIN	ASAL SEKOLAH	ASAL DAERAH	PEMINATAN	PEKERJAAN ORANG TUA	IPK	MASA STUDI
1.	12351016	Laki-Laki	SMA	Medan	Jaringan	Pegawai Swasta	Sedang	Tidak Tepat Waktu
2.	05351076	Laki-Laki	SMA	Luar Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
3.	11350453	Laki-Laki	SMA	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
4.	11350042	Laki-Laki	SMK	Luar Medan	Multimedia	Lainnya	Sedang	Tidak Tepat Waktu
5.	12350250	Laki-Laki	SMK	Siantar	Multimedia	PNS	Sedang	Tidak Tepat Waktu
6.	12351241	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
7.	12350503	Perempuan	SMA	Medan	Jaringan	Pegawai Swasta	Sedang	Tidak Tepat Waktu
8.	11351245	Perempuan	SMA	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
9.	11351112	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	Sedang	Tepat Waktu
10.	12351285	Laki-Laki	SMK	Medan	Multimedia	PNS	Sedang	Tidak Tepat Waktu
11.	09451498	Laki-Laki	SMK	Medan	Multimedia	PNS	Sedang	Tidak Tepat Waktu
12.	12350324	Perempuan	SMK	Luar Medan	Multimedia	Pegawai Swasta	Tinggi	Tepat Waktu
13.	12350379	Laki-Laki	SMA	Luar Medan	Multimedia	Lainnya	Tinggi	Tepat Waktu
14.	12350248	Perempuan	SMK	Medan	Jaringan	Pegawai Swasta	Tinggi	Tepat Waktu
15.	12351433	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	Tinggi	Tepat Waktu
16.	10351193	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
17.	11351436	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
18.	11351405	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
19.	11350085	Laki-Laki	SMK	Medan	Multimedia	Pegawai Swasta	Sedang	Tidak Tepat Waktu
20.	12350108	Laki-Laki	SMA	Medan	Multimedia	Lainnya	Tinggi	Tepat Waktu
21.	12351095	Laki-Laki	SMA	Medan	Multimedia	Lainnya	Tinggi	Tepat Waktu
22.	12351210	Laki-Laki	SMA	Medan	Multimedia	Lainnya	Tinggi	Tepat Waktu
23.	12350244	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	Sedang	Tepat Waktu
24.	11351104	Laki-Laki	SMK	Medan	Robotika	Pegawai Swasta	Sedang	Tepat Waktu
25.	13350124	Perempuan	SMK	Medan	Robotika	Pegawai Swasta	Tinggi	Tepat Waktu
26.	13350123	Perempuan	SMA	Medan	Multimedia	Pegawai Swasta	Tinggi	Tepat Waktu
27.	11350238	Laki-Laki	SMK	Luar Medan	Multimedia	PNS	Sedang	Tidak Tepat Waktu
28.	11351399	Laki-Laki	SMK	Medan	Multimedia	PNS	Sedang	Tidak Tepat Waktu
29.	11351105	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	Sedang	Tepat Waktu
30.	11351204	Laki-Laki	SMK	Medan	Jaringan	Pegawai Swasta	Sedang	Tepat Waktu

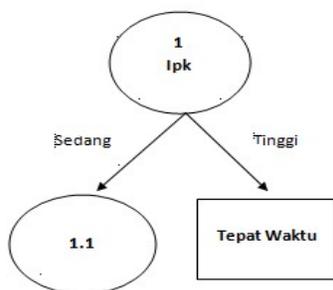
2.5 Merancang Decision Tree dengan Algoritma C4.5

Berdasarkan tabel 5, maka dilakukan proses perhitungan *entropy*(1) dan *gain*(2) untuk menentukan akar (*root*) dari pohon keputusan dalam membantu menentukan masa studi mahasiswa. Di bawah ini adalah hasil perhitungan *entropy* dan *gain* pada node 1, terlihat pada tabel 6.

Tabel 6. Hasil Perhitungan Gain dan Entrophy pada Node 1.

Node	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
1	Total		30	16	14	0,996792	
	IPK						0,442491
		Rendah	0	0	0	0	
		Sedang	21	16	5	0,791858	
		Tinggi	9	0	9	0	
	Jenis Kel						0,029237
		Laki-Laki	24	14	10	0,979869	
		Perempuan	6	2	4	0,918296	
	Asal Daerah						0,044099
		Medan	25	12	13	0,998846	
		Luar Medan	5	4	1	0,721928	
	Peminatan						0,225605
		Robotika	4	0	4	0	
		Multimedia	20	14	6	0,881291	
		Jaringan	6	2	4	0,918296	
	Asal Sekolah						0,001609
		SMA	10	5	5	1	
		SMK	20	11	9	0,992774	
	Pekerjaan ORT						0,209804
		Peg. Swasta	20	10	10	1	
	PNS	5	5	0	0		
	Lainnya	5	1	4	0,721928		

Dari hasil tabel 6. dapat diketahui bahwa atribut dengan *gain* tertinggi adalah IPK, yaitu 0,442491. Dengan demikian IPK dapat dijadikan *node* akar. Dari hasil perhitungan nilai *entropy* dan *gain* pada *node* 1, maka dapat digambarkan *decision tree* sementara seperti terlihat pada gambar 1.



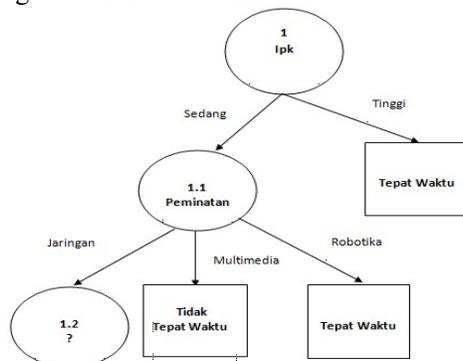
Gambar 1. Decision Tree Node 1 Penentuan Masa Studi Mahasiswa

Ada 3 nilai atribut dari IPK yaitu Rendah, Sedang dan Tinggi. Akan tetapi berdasarkan 30 sample data yang dianalisis maka ada 2 klasifikasi yang digunakan pada atribut IPK yaitu Sedang dan Tinggi, hal ini karena tidak adanya IPK yang termasuk klasifikasi Rendah pada 30 sample data yang dianalisis menggunakan Algoritma C4.5. Untuk IPK yang bernilai Tinggi sudah mengklasifikasikan nilai menjadi 1 keputusan yaitu lulus ‘Tepat Waktu’, Sedangkan untuk IPK yang bernilai Sedang belum mengklasifikasikan nilai menjadi satu keputusan, sehingga perlu dilakukan proses perhitungan *entropy* dan *gain* lebih lanjut, terlihat pada tabel 7.

Tabel 7. Hasil Perhitungan *Gain* dan *Entropy* pada Node 1.1

Node	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
1.1	IPK-Sedang		21	16	5	0,791858	0,039572
		Jenis Kelamin					
		Laki-Laki	19	14	5		
		Perempuan	2	2	0		
	Asal Daerah						0,084350
		Medan	17	12	5	0,873981	
		Luar Medan	4	4	0	0	
	Peminatan						0,601382
		Robotika	3	0	3	0	
		Multimedia	14	14	0	0	
		Jaringan	4	2	2	1	
	Asal Sekolah						0,109163
		SMA	5	5	0	0	
		SMK	16	11	5	0,896038	
	Pekerjaan ORT						0,135933
		Peg. Swasta	15	10	5	0,918296	
		PNS	5	5	0		
		Lainnya	1	1	0		

Dari hasil tabel 7 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah Peminatan yaitu 0,601382. Dengan demikian Peminatan dapat menjadi *node* cabang dari IPK yang bernilai “Sedang”. Dari hasil perhitungan nilai *entropy* dan *gain* pada *node* 1.1, maka dapat digambarkan *decision tree* sementara seperti terlihat pada gambar 2 di bawah ini.



Gambar 2. Decision Tree Node 1.1 Penentuan Masa Studi Mahasiswa

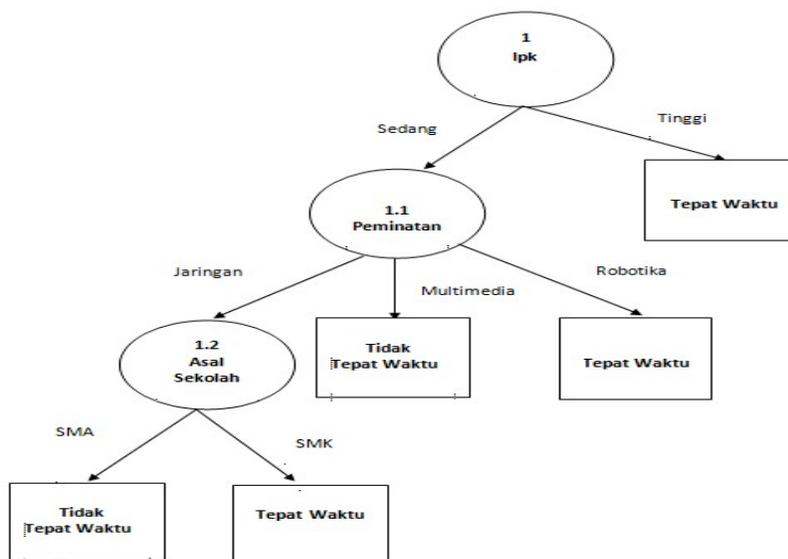
Untuk atribut Peminatan memiliki 3 nilai atribut yaitu Robotika, Multimedia dan Jaringan. Nilai atribut “Robotika” sudah mengklasifikasikan kasus menjadi 1 keputusan yaitu lulus “Tepat Waktu”, dan nilai atribut “Multimedia” juga sudah mengklasifikasikan kasus menjadi 1 keputusan yaitu lulus

“Tidak Tepat Waktu”, sedangkan nilai atribut “Jaringan” belum mengklasifikasikan kasus menjadi 1 keputusan sehingga perlu dilakukan proses perhitungan *entropy* dan *gain* lebih lanjut yaitu node 1.2.

Tabel 8. Hasil Perhitungan *Gain* dan *Entropy* pada Node 1.2

Node	Atribut	Kelas	Jumlah Kasus	Tepat Waktu	Tidak Tepat Waktu	Entropy	Gain
1.1	IPK-Sedang & Peminatan Jaringan	-	4	2	2	1	
	Jenis Kelamin						0,311278
		Laki-Laki	3	1	2	0,918296	
		Perempuan	1	1	0	0	
	Asal Daerah						0
		Medan	4	2	2	1	
		Luar Medan	0	0	0	0	
	Asal Sekolah						1
		SMA	2	2	0	0	
		SMK	2	0	2	0	
	Pekerjaan ORT						0
		Peg. Swasta	4	2	2	1	
		PNS	0	0	0	0	
		Lainnya	0	0	0	0	

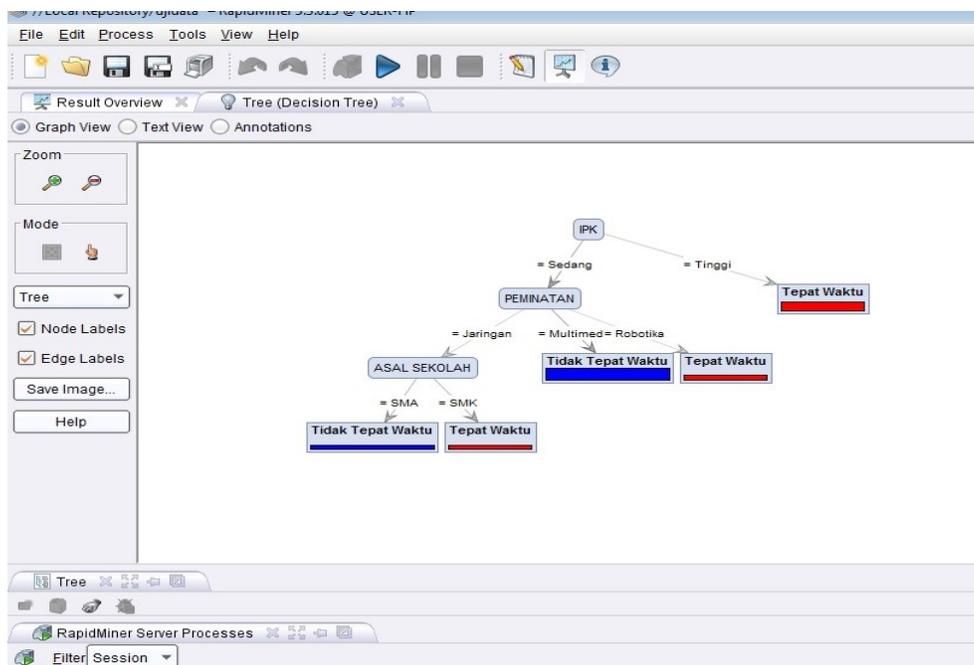
Dari hasil tabel 8 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah atribut “Asal Sekolah” yaitu 1. Dengan demikian atribut “Asal Sekolah” dapat menjadi *node* cabang dari IPK bernilai “Sedang” dan Peminatan bernilai “Jaringan”. Untuk atribut Asal Sekolah memiliki 2 nilai atribut yaitu SMA dan SMK. Nilai atribut “SMA” sudah mengklasifikasikan kasus menjadi 1 keputusan yaitu lulus “Tidak Tepat Waktu”, dan nilai atribut “SMK” juga sudah mengklasifikasikan kasus menjadi 1 keputusan yaitu lulus “Tepat Waktu”. Karena kedua nilai dari atribut “Asal Sekolah” tersebut sudah mengklasifikasikan kasus menjadi 1 maka berdasarkan hasil perhitungan nilai *entropy* dan *gain* pada *node* 1.2, dapat digambarkan *decision tree* akhir dalam menentukan masa studi mahasiswa prodi teknik informatika di Universitas Harapan Medan seperti terlihat pada gambar 3 di bawah ini.



Gambar 3. *Decision Tree* Node 1.2 Penentuan Masa Studi Mahasiswa

3. HASIL DAN PEMBAHASAN

Implementasi dilakukan menggunakan salah satu *software Data Mining* yaitu *Rapid Miner 5.3.015*. Semua atribut indikator *input* dan atribut tujuan disimpan dalam format *xlsx*, kemudian di *import* ke *software Rapid Miner 5.3.015* dan menghasilkan *decision tree* sebagai berikut:



Gambar 4. *Decision Tree* Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika di Fakultas Teknik dan Komputer Universitas Harapan Medan

Berdasarkan pengujian yang telah dilakukan, maka dapat disimpulkan bahwa hasil analisis penentuan masa studi mahasiswa menggunakan metode algoritma C4.5 adalah sebagai berikut:

1. Jika IPK mahasiswa bernilai “Tinggi” maka masa studi mahasiswa lulus “Tepat Waktu”.
2. Jika IPK mahasiswa bernilai “Sedang” dan Peminatan adalah “Robotika” maka masa studi mahasiswa lulus “Tepat Waktu”.
3. Jika IPK mahasiswa bernilai “Sedang” dan peminatan adalah “Multimedia” maka masa studi mahasiswa lulus “Tidak Tepat Waktu”.
4. Jika IPK mahasiswa bernilai “Sedang” dan peminatan adalah “Jaringan” dan asal sekolah adalah “SMK” maka masa studi mahasiswa lulus “Tepat Waktu”.
5. Jika IPK bernilai “Sedang” dan peminatan adalah “Jaringan” dan asal sekolah adalah “SMA” maka masa studi mahasiswa lulus “Tidak Tepat Waktu”.

4. KESIMPULAN

Setelah melakukan analisis, perancangan, implementasi beserta pengujian dalam menerapkan algoritma C4.5 untuk menentukan masa studi mahasiswa prodi teknik informatika di Universitas Harapan Medan, maka didapat kesimpulan bahwa:

1. Dari hasil analisa data mahasiswa alumni prodi teknik informatika di Universitas Harapan Medan, maka ada 6 atribut yang digunakan sebagai atribut prediktor yaitu Jenis Kelamin, Asal Sekolah, Asal Daerah, Peminatan, Pekerjaan Orang Tua dan IPK.
2. Hasil *Decision tree* penentuan masa studi mahasiswa menghasilkan keputusan bahwa mahasiswa yang lulus ‘Tepat Waktu’ adalah mahasiswa yang memiliki IPK tinggi.
3. Mahasiswa yang memiliki IPK sedang, berdasarkan hasil decision tree penentuan masa studi mahasiswa menghasilkan keputusan bahwa mahasiswa yang lulus ‘Tepat Waktu’ adalah mahasiswa yang mengambil peminatan robotika dan jaringan dengan asal sekolah SMK.
4. Mahasiswa yang ‘Tidak Lulus Tepat Waktu’ adalah mahasiswa yang memiliki IPK Sedang dengan peminatan multimedia.

UCAPAN TERIMA KASIH

Alhamdulillah, puji syukur peneliti kepada Allah SWT karena nikmat iman, Islam serta ilmu sehingga memberikan kemudahan atas segala urusan, kemudian terima kasih kepada seluruh anggota tim penelitian dan semua pihak yang telah membantu terlaksananya penelitian ini.

BAHAN REFERENSI

- [1] Andriani, Anik, 2013, Sistem Prediksi Penyakit Diabetes Berbasis *Decision Tree*, *Jurnal Bianglala Informatika*, Vol.1, No.1, 1-10.
- [2] Goldie Gunadi, dan Indra Sensue, 2012, Penerapan *Data Mining* Market Basket Analysis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan *Frequent Pattern Growth(FB: Growth)*: Studi Kasus Percetakan Gramedia, *Jurnal Telematika MKom*, Vol.4, No.1, 118-132.
- [3] Hastuti, Khafiizh, 2012, Analisis Komparasi Algoritma Klasifikasi *Data Mining* untuk Prediksi Mahasiswa Non Aktif, *Seminar Nasional Teknologi Informasi & Komunikasi Terapan(Semantik)*, Semarang, 23 Juni 2012.
- [4] Rani Wahyu Ningrat, dan Budi Santosa, 2012, Pemilihan Diet Nutrien bagi Penderita Hipertensi Menggunakan Metode Klasifikasi *Decision Tree*(Studi Kasus: RSUD Syarifah Ambami Rato Ebu Bangkalan), *Jurnal Teknik ITS*, Vol.1, No.1, A536-A539.
- [5] Selvia Lorena Br.Ginting, Wendi Zarman, Ida Hamidah, 2014, Analisis dan Penerapan *Algoritma C4.5* dalam *Data Mining* untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik, *Seminar Nasional Aplikasi Sains dan Teknologi (SNAST)*, Yogyakarta, 15 November 2014.
- [6] Siska Haryati, Aji Sudarsono, Eko Suryana, 2015, Implementasi *Data Mining* untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma *C4.5*(Studi Kasus: Universitas Dehasen Bengkulu), *Jurnal Media Infotama*, No.2, Vol. 11, 130-138.
- [7] Sunjana, 2010, Aplikasi *Data Mining* Data Mahasiswa dengan Metode Klasifikasi *Decision Tree*, *SNATI 2010*, Yogyakarta, 19 Juni 2010.
- [8] Swastina,liliana, 2013, Penerapan Algoritma C4.5 untuk Penentuan Prodi Mahasiswa, *Jurnal GEMA AKTUALITA*, No.1, Vol.2, 93-98.
- [9] Wahyudin, 2009, Metode Iterative Dischotomizer 3 (ID3) untuk Penyeleksian Penerimaan Mahasiswa Baru, *Jurnal Pendidikan Teknologi Informasi dan Komunikasi(PTIK)*, Vol.1, No.2, 5-15.
- [10] Nasution, Muhammad Irwan Padli, 2012, Sistem Informasi Pengontrolan Mutu Produk Pada PT SC Johnson Manufacturing Medan, *Seminar Nasional Informatika 2012 (SNIIf-2012)*