

Implementasi Algoritma C4.5 dalam Memprediksi Masa Studi Mahasiswa STMik Dumai.

Asparizal
Jurusan Teknik Informatika,
STMik Dumai
asparizal73@gmail.com

Putri Yunita
Jurusan Sistem Informasi,
STMik Dumai
Yu-pi19@yahoo.com

Zhiyaul Ihsan
Jurusan Teknik Informatika,
STMik Dumai
zhiyaulihsan7@gmail.com

Abstrak

Masa studi adalah rentang waktu penyelesaian dari seluruh mata kuliah yang diwajibkan bagi mahasiswa sebelum lulus jenjang pendidikan pada sebuah perguruan tinggi. Program Studi (S1) Teknik Informatika memiliki beban studi 152 total sks dan Program Studi (S1) Sistem Informasi 150 total sks yang wajib diselesaikan dalam 8 sampai dengan 14 semester. Namun kenyataannya masa studi mahasiswa STMik Dumai sangat lama. Masih sedikit mahasiswa yang tamat/lulus tepat 8 semester. Hal ini disebabkan oleh rendahnya indeks prestasi (IP) semester mahasiswa sehingga menyebabkan tidak bisa mengambil seluruh mata kuliah yang ditawarkan semester berikutnya. Mata kuliah prasyarat yang belum lulus. Belum terpenuhi total sks untuk bisa mengambil mata kuliah skripsi. Merancang dan mengimplementasikan sistem yang dapat memprediksi masa studi mahasiswa merupakan suatu alternatif yang sangat penting untuk mengatasi masalah-masalah berkaitan dengan masa studi di STMik Dumai. Jika masa studi mahasiswa dapat diprediksi lebih dini, pihak Program Studi dapat memberikan saran/rekomendasi sehingga mahasiswa dapat lulus tepat 8 semester.

Kata Kunci : Masa Studi, Data Mining, Algoritma C4.5, Visual Basic 6.0, MySQL

1. Pendahuluan

Masa studi adalah rentang waktu mahasiswa menjalani masa kuliah. (Meinanda, Annisa, Muhandri,

& Suryadi, 2009) Menurut BAN-PT (BAN-PT.,2011) kualitas sebuah perguruan tinggi diukur berdasarkan tujuh standar dan butir-butir elemen penilaian/parameter, salah satunya adalah mahasiswa dan lulusan. Program sarjana merupakan jalur pendidikan akademik yang mempunyai beban studi kumulatif minimal 144 sks dan maksimal 160 sks dengan lama studi kumulatif antara 8 semester sampai dengan 14 semester setelah sekolah lanjutan tingkat atas. Khusus mengenai evaluasi standar dan lulusan, komponen yang dinilai adalah rata-rata masa studi dan IPK. (Hafidh, 2015). Program Sarjana (S1) Program Studi Teknik Informatika STMik Dumai memiliki beban studi berjumlah 152 sks dan Program Studi Sistem Informasi berjumlah 150 sks berdasarkan kurikulum 2009. Pada kenyataannya tingkat kelulusan di STMik Dumai masih relatif rendah, hal ini terbukti dari angkatan 2010 hanya 13 orang yang diwisuda perdana pada tahun 2014 dari 51 orang mahasiswa terdaftar. Bahkan masih ada mahasiswa angkatan 2009 yang diwisuda sebanyak 10 orang mahasiswa dan 2 orang mahasiswa angkatan 2008 pada tahun yang sama.(Dumai, 2014) Hal ini disebabkan oleh beberapa masalah diantaranya: a) rendahnya indeks prestasi (IP) semester yang didapat oleh mahasiswa sehingga menyebabkan sedikitnya matakuliah semester berikutnya yang bisa diambil. b). Adanya mata kuliah prasyarat, sehingga menjadi penghambat bagi mahasiswa untuk mengambil matakuliah berikutnya sekiranya matakuliah prasyarat tersebut belum lulus. c) belum terpenuhi batas 120 total sks untuk bisa mengambil mata kuliah skripsi. d) pelaksanaan bimbingan akademis dari dosen penasehat akademis (PA) belum efektif kepada mahasiswa karena hanya dilaksanakan 1 kali diawal setiap semester. Semua

kendala diatas berpengaruh kepada mahasiswa baik secara langsung maupun tidak langsung yang menyebabkan banyak mahasiswa yang menempuh masa studi lebih dari empat tahun atau 8 semester. Oleh sebab itu bagaimana merancang dan mengimplementasikan sebuah sistem yang dapat memprediksi masa studi mahasiswa merupakan suatu alternatif yang sangat penting. Jika masa studi mahasiswa dapat diketahui lebih dini, pihak Program Studi dapat memberikan saran/rekomendasi sehingga mahasiswa dapat lulus tepat 8 semester. (Hafidh, 2015)

Penerapan algoritma C4.5 menjadi salah satu solusi untuk memprediksi masa studi mahasiswa. Algoritma C4.5 biasa dipakai untuk pengolahan *data mining*. *Data mining* adalah proses pencarian informasi terhadap data yang sudah ada. *Data mining* memiliki beberapa teknik, salah satunya adalah teknik klasifikasi. Klasifikasi ini adalah proses untuk menemukan model atau fungsi yang menggambarkan kelas atau konsep dari suatu data, sedangkan *decision tree* merupakan salah satu metode dari teknik klasifikasi. Salah satu algoritma yang mengimplementasikan klasifikasi dengan metode *decision tree* yaitu algoritma C4.5.

1.2. Rumusan Masalah

Secara garis besar masalah yang harus dipahami pada penelitian ini adalah sebagai berikut:

1. Bagaimana menggunakan *data mining* untuk memprediksi masa studi mahasiswa dengan metode algoritma C4.5?
2. Bagaimana output yang dihasilkan oleh sistem menjadi bahan pertimbangan bagi Program Studi untuk memberi saran atau rekomendasi sehingga mahasiswa dapat lulus tepat waktu 8 semester.

2. Landasan Teori

2.1. Pengertian Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *mechine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Luthfi, 2009)

Istilah *data mining* memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Istilah *knowledge discovery* atau penemuan pengetahuan tepat digunakan karena tujuan utama dari *data mining* memang untuk mendapatkan pengetahuan

yang masih tersembunyi di dalam bongkahan data. Istilah *pattern recognition* atau pengenalan pola pun tepat untuk digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang mungkin juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi. (Sani Susanto & Dedy Suryadi, 2010).

2.1.1. Proses Data Mining. Secara skematis, Goronescu (2011) membagi langkah pelaksanaan *data mining* dalam tiga aktivitas, yaitu: (Prabowo., Rahmadya., 2013)

1. Eksplorasi data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan ciri, dan lain-lain.
2. Membuat Model dan Pengujian Validitas Model, merupakan pemeliharaan terhadap model-model yang sudah dikembangkan yang cocok dengan kasus yang dihadapi. Dengan kata lain, dilakukan pemilihan model secara kompetitif.
3. Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada. Tahap ini merupakan tahap yang menentukan apakah model yang telah dibangun dapat menjawab permasalahan yang dihadapi.

2.1.2. Pemodelan Data Mining. Pemodelan adalah penggunaan prinsip atau teknik-teknik tertentu dalam suatu rancangan sistem. Kian kompleksnya masalah-masalah yang dijumpai saat ini membuat proses pemodelan menjadi makin kompleks juga. Oleh karena itu, diperlukan pengetahuan yang juga lebih spesifik dan mendetail. Sebagai bahan pertimbangan, Gounescu (2011),

menyarankan tahapan-tahapan sebagai berikut: (Prabowo., Rahmadya., 2013)

1. Identifikasi
Ini merupakan tahapan pertama dalam pemodelan *data mining* dari suatu permasalahan yang ada di lapangan. Dalam pengidentifikasian suatu masalah, dijumpai dua pendekatan yang saling bertolak belakang. Pendekatan pertama adalah pendekatan yang mengutamakan pengetahuan terdahulu dari suatu kasus. Dalam hal ini pengetahuan *a priori* menjadi andalan utama para pendukung ini. Pendekatan kedua adalah pengidentifikasian yang murni berdasarkan data yang ada. Se jauh mungkin dihindari dugaan awal terhadap suatu kondisi.
2. Estimasi dan Pencocokan
Setelah tahap seleksi selesai, tahap berikutnya adalah membuat formulasi numerik terhadap

suatu model. Tahapan ini dikenal dengan nama tahapan pencocokan model dengan data. Sedangkan konversi dari model menjadi angka numerik disebut dengan istilah estimasi.

3. Pengujian

Pengujian merupakan tahap akhir sebelum sistem diimplementasikan. Sistem yang telah diuji terhadap data lain yang belum pernah dimiliki dan bukan data yang dipakai untuk membentuk model itu. Keberhasilan suatu pengujian bergantung dari *output* yang dihasilkan oleh suatu sistem yang diuji, apakah sesuai dengan kenyataan yang ada atau tidak.

4. Penerapan Praktis

Sistem yang dirancang adalah ditujukan untuk menyelesaikan permasalahan-permasalahan yang ada di lapangan.

5. Iterasi

Iterasi mengharuskan perancang untuk selalu berfikir kembali terhadap model yang dibuatnya. Dengan adanya perulangan-perulangan diharapkan diperoleh model yang tangguh dan cocok dengan situasi dan kondisi yang terjadi saat implementasi.

2.2. Pohon Keputusan

Pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam data mining dan merupakan salah satu metode yang populer dalam menentukan keputusan studi kasus. Hal ini karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi besar. (Prabowo., Rahmadya., 2013)

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. (Luthfi, 2009).

Masalah pertama pada pembuatan pohon keputusan adalah variabel manakah yang menjadi akar dari pohon tersebut. Akar di sini adalah pemisah pertama dari pohon keputusan. Di kenal istilah *Bayesian Score* yang menilai suatu variabel atau dalam pohon keputusan terkenal dengan sebutan Entropi. Entropi dihitung dengan rumusan sebagai berikut:

Entropi (p_1, p_2, \dots, p_n) = $p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots p_n \log_2 p_n$ di mana p_1, p_2, \dots, p_n adalah probabilitas kondisi pada atribut yang diprediksi, dan: $p_1 + p_2 + \dots + p_n = 1$.

2.3. Algoritma C4.5

Algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Pada akhir tahun 1970 sampai awal tahun 1980, J. Ross Quinlan, seorang peneliti di bidang *machine learning*, membuat sebuah algoritma *decision tree* yang dikenal dengan ID3 (*Iterative Dichotomiser*). Quinlan kemudian membuat algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan pengembangan dari algoritma ID3. (Prabowo., Rahmadya., 2013). Algoritma C4.5 merupakan struktur pohon di mana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Prabowo., Rahmadya., 2013) yaitu:

1. Menyiapkan *data training*. *Data training* biasanya diambil dari histori yang pernah terjadi sebelumnya dan sesudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i$$
 Keterangan :
 S = Himpunan kasus
 n = Jumlah partisi S
 pi = Proporsi Si terhadap S
3. Kemudian hitung nilai gain menggunakan rumus:

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

Keterangan:

S = Himpunan kasus

A = Fitur

n = Jumlah partisi atribut A

|Si| = Proporsi Si terhadap S

|S| = Jumlah kasus dalam S

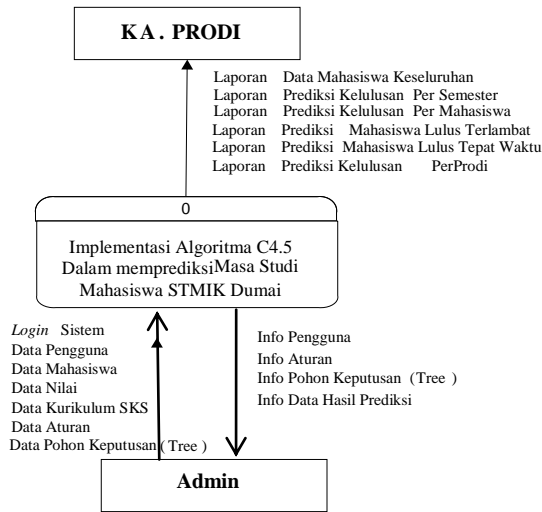
4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.

- b. Tidak ada atribut di dalam record yang dipartisi lagi.
- c. Tidak ada record di dalam cabang yang kosong.

3. Perancangan Sistem

3.1. Konteks Diagram

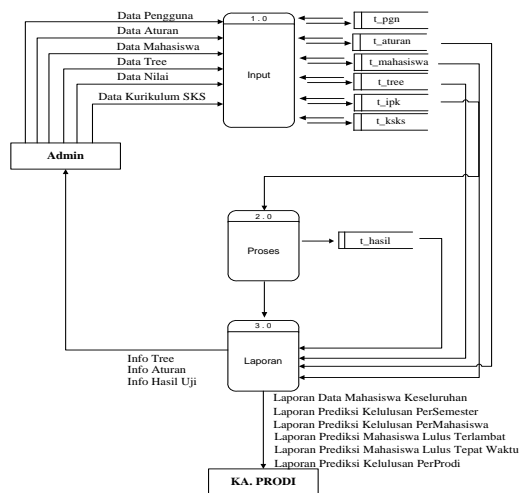
Desain sistem informasi terpadu diawali dengan sebuah diagram konteks sebagai berikut. (Nataniel & Hatta, 2009)



Gambar 1. Konteks diagram

3.2. DFD (Data Flow Diagram)

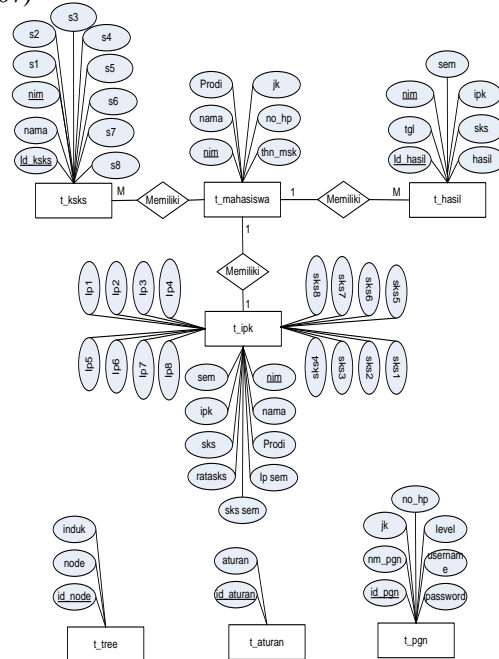
DFD ialah diagram yang menggunakan notasi untuk menggambarkan arus dari data sistem. (Indrajani, 2011)



Gambar 2. Data flow diagram

3.3. Entity Relation Diagram (ERD)

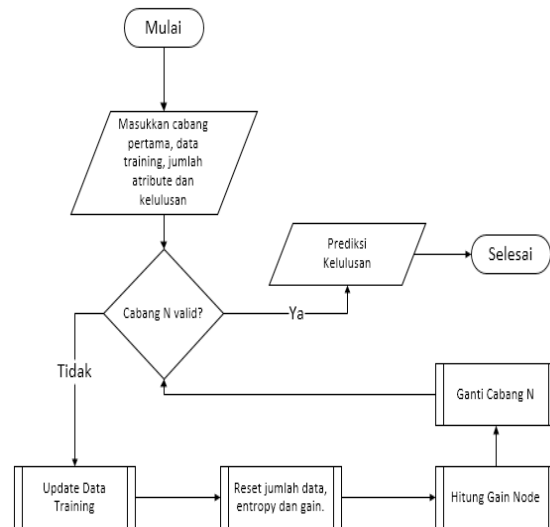
ERD berisi komponen-komponen himpunan entitas dan himpunan relasi yang dilengkapi dengan atribut-atribut yang merepresentasikan seluruh fakta. (Fattah, 2007)



Gambar 3. Entity relation diagram (ERD)

3.4. Rancangan Algoritma

3.4.1. Flowchart Algoritma C4.5.



Gambar 4. Flowchart algoritma C4.5

2. Rata-rata SKS

Rata-rata SKS dapat diklasifikasikan dalam 2 kelas, yaitu untuk rata-rata “SKS ≤ 18” diklasifikasikan menjadi kelas “Kurang” dan untuk rata-rata “SKS > 18” yang diklasifikasikan menjadi kelas “Cukup”.

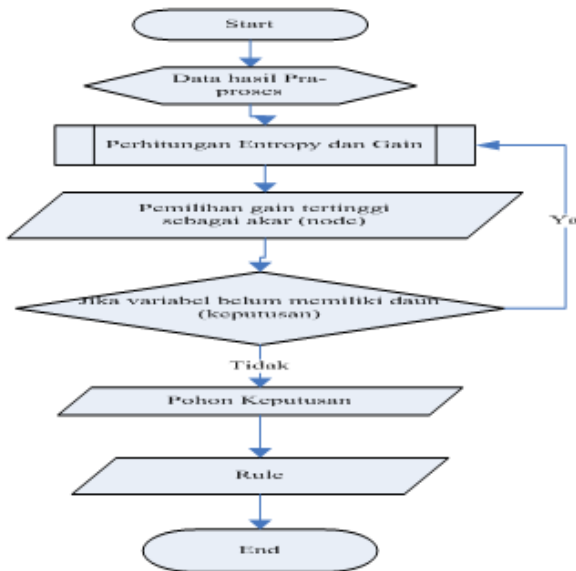
Tabel 4. Tabel klasifikasi rata-rata SKS

IPK	Klasifikasi
SKS ≤ 18	Kurang
SKS > 18	Cukup

3.4.6. Perhitungan Algoritma C4.5. Data hasil transformasi selanjutnya dianalisa untuk menghasilkan sebuah pohon keputusan dengan menggunakan algoritma C4.5, secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Perhitungan Entropy dan Gain
2. Pemilihan Gain tertinggi sebagai akar (Node)
3. Ulangi proses perhitungan Entropy dan Gain untuk mencari cabang sampai semua kasus pada cabang memiliki kelas yang sama yaitu pada saat semua variabel telah menjadi bagian dari pohon keputusan atau masing –masing variabel telah memiliki daun atau keputusan.
4. Membuat Rule berdasarkan pohon keputusan.

Tahapan algoritma C4.5 dapat gambarkan dalam flowchart yang terlihat pada gambar 6.



Gambar 6. Flowchart tahapan algoritma C4.5

3.4.7. Perhitungan Entropy dan Gain. Baris Total kolom Entropy dapat dihitung dengan persamaan 2 sebagai berikut :

$$\begin{aligned}
 Entropy(Total) &= \left(-\frac{6}{10} \cdot \log_2 \left(\frac{6}{10}\right)\right) + \left(-\frac{4}{10} \cdot \log_2 \left(\frac{4}{10}\right)\right) \\
 &= 0,970950594 \\
 Entropy(Prodi TI) &= \left(-\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right)\right) \\
 &= 1 \\
 Entropy(Prodi SI) &= \left(-\frac{4}{6} \cdot \log_2 \left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} \cdot \log_2 \left(\frac{2}{6}\right)\right) \\
 &= 0,918295834 \\
 Entropy(IPK Sangat Kurang) &= \left(-\frac{2}{2} \cdot \log_2 \left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} \cdot \log_2 \left(\frac{0}{2}\right)\right) \\
 &= 0 \\
 Entropy(IPK Kurang) &= \left(-\frac{3}{3} \cdot \log_2 \left(\frac{3}{3}\right)\right) + \left(-\frac{0}{3} \cdot \log_2 \left(\frac{0}{3}\right)\right) \\
 &= 0 \\
 Entropy(IPK Cukup) &= \left(-\frac{1}{3} \cdot \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \cdot \log_2 \left(\frac{2}{3}\right)\right) \\
 &= 0,918295834 \\
 Entropy(IPK Tinggi) &= \left(-\frac{2}{2} \cdot \log_2 \left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} \cdot \log_2 \left(\frac{0}{2}\right)\right) \\
 &= 0 \\
 Entropy(SKS Kurang) &= \left(-\frac{5}{5} \cdot \log_2 \left(\frac{5}{5}\right)\right) + \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right) \\
 &= 0 \\
 Entropy(SKS Cukup) &= \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} \cdot \log_2 \left(\frac{4}{5}\right)\right) \\
 &= 0,721928059
 \end{aligned}$$

Sementara itu, nilai Gain dihitung dengan menggunakan persamaan 1 :

$$\begin{aligned}
 Gain(Total, Prodi) &= 0,970950594 - \left(\left(\frac{4}{10}\right) \cdot 1\right) + \left(\frac{6}{10} \cdot 0,918295834\right) \\
 &= 0,019973094 \\
 Gain(Total, IPK) &= 0,970950594 - \left(\left(\frac{2}{10}\right) \cdot 0\right) + \left(\frac{3}{10}\right) \cdot 0 + \left(\frac{3}{10} \cdot 0,918295834\right) + \left(\frac{2}{10}\right) \cdot 0 \\
 &= 0,695461844 \\
 Gain(Total, SKS) &= 0,970950594 - \left(\left(\frac{5}{10}\right) \cdot 0\right) + \left(\frac{5}{10}\right) \cdot 0,721928059 \\
 &= 0,609986547
 \end{aligned}$$

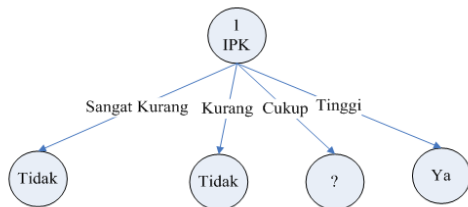
Tabel 5. Perhitungan node 1

		Jumlah Kasus	tidak	ya	entropy	gain
Total		10	6	4	0,970950594	
Prodi						0,019973094
	TI	4	2	2	1	
	SI	6	4	2	0,918295834	
IPK						0,695461844
	Sangat Kurang	2	2	0	0	
	Kurang	3	3	0	0	
	Cukup	3	1	2	0,918295834	
	Tinggi	2	0	2	0	
Rata-rata SKS						0,609986547
	Kurang	5	5	0	0	
	Cukup	5	1	4	0,721928059	

Dari hasil pada tabel 5 dapat diketahui bahwa atribut dengan Gain tertinggi adalah IPK, yaitu sebesar 0.695461844. Dengan demikian, IPK dapat menjadi node akar. Ada 4 nilai atribut dari IPK, yaitu Sangat Kurang, Kurang, Cukup dan Tinggi, dari ke empat nilai atribut, atribut Sangat Kurang, Kurang dan Tinggi sudah mengklasifikasikan kasus menjadi satu yaitu, Sangat Kurang dan Kurang keputusannya Tidak dan

Tinggi keputusannya Ya, sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini. Tetapi untuk IPK Cukup masih perlu dilakukan perhitungan lagi.

Dari hasil tersebut dapat digambarkan pohon keputusan sementara tampak pada gambar 19 di bawah ini :



Gambar 7. Pohon keputusan perhitungan node 1

3.4.8. Perhitungan Cabang Level 1.

Langkah-langkah menghitung cabang level 1, IPK cukup sebagai berikut :

Baris total perhitungan *entropy* dihitung dengan persamaan 2 :

$$\begin{aligned} \text{Entropy (Prodi TI)} &= (-\frac{0}{3} \cdot \log_2(\frac{0}{3})) + (-\frac{1}{3} \cdot \log_2(\frac{1}{3})) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy (Prodi SI)} &= (-\frac{1}{2} \cdot \log_2(\frac{1}{2})) + (-\frac{1}{2} \cdot \log_2(\frac{2}{2})) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{Entropy (SKS Cukup)} &= (-\frac{1}{3} \cdot \log_2(\frac{1}{3})) + (-\frac{0}{3} \cdot \log_2(\frac{0}{3})) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy (SKS Tinggi)} &= (-\frac{0}{2} \cdot \log_2(\frac{0}{2})) + (-\frac{2}{2} \cdot \log_2(\frac{2}{2})) \\ &= 0 \end{aligned}$$

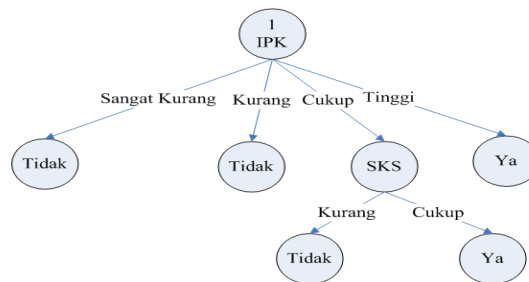
Sementara itu nilai gain dihitung menggunakan persamaan 1 :

$$\begin{aligned} \text{Gain(Total, Prodi)} &= 0.918295834 - ((\frac{1}{3} \cdot 0) + (\frac{2}{3} \cdot 1)) \\ &= 0.251629167 \\ \text{Gain(Total, SKS)} &= 0.918295834 - ((\frac{1}{3} \cdot 0) + (\frac{2}{3} \cdot 0)) \\ &= 0,918295834 \end{aligned}$$

Tabel 6. Perhitungan cabang level 1 IPK cukup

		Jumlah Kasus	tidak	ya	entropy	gain
IPK	Cukup	3	1	2	0,918295834	
Prodi						0,251629167
	TI	1	0	1	0	
	SI	2	1	1	1	
Rata-rata SKS						0,918295834
	Kurang	1	1	0	0	
	Cukup	2	0	2	0	

Dari hasil pada tabel 6 di atas, dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah SKS, yaitu sebesar 0.918295834. Dengan demikian SKS dapat menjadi node cabang dari IPK Cukup. Ada dua nilai atribut dari SKS yaitu Cukup dan Tinggi. Dari kedua nilai atribut sudah mengklasifikasikan kasus menjadi satu yaitu, Kurang keputusannya Tidak, sedangkan Cukup keputusannya Ya, sehingga nilai tidak perlu dilakukan perhitungan lagi.



Gambar 8. Pohon keputusan perhitungan cabang level 1

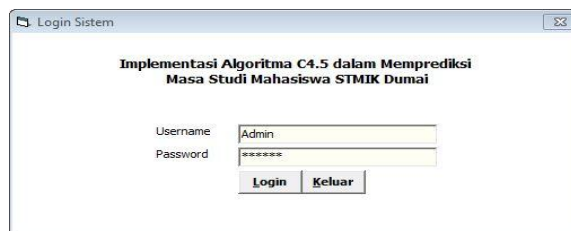
Berdasarkan contoh pembuatan *rule* diatas, *Rule* yang dihasilkan dari pohon keputusan berdasarkan gambar 8 adalah sebagai berikut:

1. IF IPK = Sangat Kurang Then Tidak.
2. IF IPK = Kurang Then Tidak.
3. IF IPK = Cukup And SKS = Kurang Then Tidak
4. IF IPK = Cukup And SKS = Cukup Then Ya
5. IPK = Tinggi Then Ya

4. Implementasi Sistem

4.1. Form Input Login Aplikasi

Untuk login dari aplikasi *user* harus login dengan menginput *user* dan *password* dengan benar, setelah *user* dan *password* benar, selanjutnya tekan tombol *enter* dan aplikasi otomatis masuk kedalam menu utama.



Gambar 9. Form input login aplikasi

4.2. Form Data Pengguna

Pada *form* data pengguna berisi informasi pengelolaan *user* dan *password*, dimana pada sub menu admin ini dapat melakukan penambahan dan penghapusan dan melihat *password* secara keseluruhan.

No	ID	Nama	Jenis Kelamin	No. HP	Level
1	1	Zhiyaul Ihsan	LK	0852XXXXXXX	Admin
2	2	User	LK	0813XXXXXXX	User

Gambar 10. Form data pengguna

4.3. Form Data Aturan

Form data Aturan berguna untuk melakukan proses pengolahan data-data Aturan.

No	Aturan
1	Jika IPK <= 2,4 Maka Lulus Lebih dari 8 Semester
2	Jika IPK = 2,5 - 2,74 Maka Lulus Lebih dari 8 Semester
3	Jika IPK = 2,75 - 2,99 Dan Rata-Rata <= 18 Maka Lulus Lebih da...
4	Jika IPK = 2,75 - 2,99 Dan Rata-Rata > 18 Maka Lulus Tepat 8 Se...
5	Jika IPK >= 3 Maka Lulus Tepat 8 Semester

Gambar 11. Form data aturan

4.4. Form Data Tree

Form data Tree berguna untuk melakukan proses pengolahan data-data Tree.

No	ID No...	Node	Induk
1	1	IPK ?	1
2	11-1	<= 2,4 : Tidak	1
3	11-2	2,5 - 2,74 : Tidak	1
4	11-3	2,75 - 2,99 : Rata-Rata SKS ?	1
5	11-3-1	<= 18 : Tidak	11-3
6	11-3-2	> 18 : Ya	11-3
7	11-4	>= 3 : Ya	1

Gambar 12. Form data tree

4.5 Form Master Data Mahasiswa

Form data mahasiswa berguna untuk melakukan proses pengolahan data-data mahasiswa.

No	NIM	Nama	Jenis Kelamin	No. HP
1	2123022	Endang Opianty	PR	08948934934
2	2123027	Firmandi	LK	08999272387
3	2123030	Hikmatul Hayati	PR	08328749349
4	2123034	Ilham Saputra	LK	08374394384
5	2123036	Ira Maya Wulandari	PR	08442565262
6	2124013	Dwi Guna Putra M	LK	07493493783
7	2124023	M. Arafat	LK	08448347384
8	2124028	Mery Christiany	PR	08347649385
9	2124031	Muhammad Rayhan	LK	09392738627
10	2124045	Sunriadi	LK	08439874983

Gambar 13. Form master data mahasiswa

4.6. Form Data Kurikulum SKS

Form data kurikulum ini merupakan bagian dari *bar* menu master, *form* data ini berguna untuk melakukan proses pengolahan data-data SKS.

Data Kurikulum SKS

NIM : 212 * 3 Digit Depan NIM
 Prodi : TI

SKS

Semester 1	22	Semester 5	101
Semester 2	41	Semester 6	122
Semester 3	59	Semester 7	136
Semester 4	79	Semester 8	151

Simpan Ubah Hapus Batal

No	NIM	Prodi	Sem 1	Sem 2
1	212	TI	22	41
2	212	SI	21	41
3	213	SI	21	42
4	213	TI	22	41
5	214	SI	21	41
6	214	TI	22	41

Gambar 14. Form data kurikulum SKS

4.7. Form Data Nilai

Form data nilai berguna untuk melakukan proses pengolahan data-data nilai.

Data Nilai

NIM : 2124031
 Nama : Muhammad Rayhan
 Semester : 7
 Prodi : TI

IP Semester

Semester 1	3,22
Semester 2	2,21
Semester 3	3,05
Semester 4	3,55
Semester 5	3,18
Semester 6	3,52
Semester 7	3
Semester 8	

Index Prestasi Kumulatif : 3,10

SKS

Semester 1	22
Semester 2	19
Semester 3	18
Semester 4	20
Semester 5	22
Semester 6	21
Semester 7	12
Semester 8	

Total SKS : 134
 Rata-rata SKS : 19

Simpan Ubah Hapus Batal

Cari Data Nilai Mahasiswa

NIM	Nama	Semester	IPK	Total SKS
2123007	Redi Adriadi	7	2,88	144
2123022	Endang Opia...	7	2,88	140
2123027	Firmandi	7	3,22	133
2123030	Hikmatul Ha...	7	2,58	133
2123034	Jilam Saputra	5	2,16	76
2123036	Ira Maya Wul...	7	2,40	125
2124013	Dwi Guna Put...	7	3,04	139
2124032	M. Amfak	7	2,54	134

Gambar 15. Form data nilai

4.8. Form Lihat Aturan

Form menu Lihat Aturan berguna untuk melihat aturan yang terbentuk dari hasil perhitungan Algoritma C4.5.

Lihat Aturan

No	Aturan
1	Jika IPK <= 2,4 Maka Lulus Lebih dari 8 Semester
2	Jika IPK = 2,5 - 2,74 Maka Lulus Lebih dari 8 Semester
3	Jika IPK = 2,75 - 2,99 Dan Rata-Rata <= 18 Maka Lulus Lebih da...
4	Jika IPK = 2,75 - 2,99 Dan Rata-Rata >18 Maka Lulus Tepat 8 Se...
5	Jika IPK >= 3 Maka Lulus Tepat 8 Semester

Gambar 16. Form lihat aturan

4.9. Form Visualisasi Pohon Keputusan

Form menu visualisasi pohon keputusan berguna untuk melihat pohon keputusan yang terbentuk dari hasil perhitungan Algoritma C4.5.

Visualisasi Pohon Keputusan (Tree)

Pohon Keputusan

```

    graph TD
      A[IPK ?] -- "<= 2,4 : Tidak" --> B[2,5 - 2,74 : Tidak]
      A -- "2,5 - 2,74 : Tidak" --> B
      A -- "2,75 - 2,99 : Rata-Rata SKS ?" --> C[<= 18 : Tidak]
      A -- "2,75 - 2,99 : Rata-Rata SKS ?" --> D[> 18 : Ya]
      C -- "<= 18 : Tidak" --> E[> 18 : Ya]
      D -- "> 18 : Ya" --> E
      E -- ">= 3 : Ya" --> F[>= 3 : Ya]
    
```

Gambar 17. Visualisasi pohon keputusan

4.10 Form Proses Keputusan

Form proses berguna untuk melakukan proses pengolahan data nilai.

Proses Keputusan

Tanggal : 19/05/2016
 NIM : 2123027
 Nama : Firmandi
 Semester : 7
 SKS : 145
 SKS Sebenarnya : 143

Atribut

Prodi	SI
IPK	3,18
Rata-rata SKS	20

Proses Data Mining Algoritma C4.5

Hasil Prediksi

Lulus Tepat 8 Semester

Simpan Batal

Gambar 18. Form proses keputusan

4.11. Tampilan Hasil Prediksi

No	Tanggal	NIM	Nama	Prodi	Sem	IPK	SKS	Hasil Keputusan
1	23/05/2016	2123034	Iham Saoutra	SI	5	2,16	76	Lulus Lebih Dari 8 Semes.
2	23/05/2016	2134012	M. Anief Taufiq	TI	5	3,73	101	Lulus Tepat 8 Semester
3	23/05/2016	2134024	Samuel Samosir	TI	5	2,59	94	Lulus Lebih Dari 8 Semes.
4	23/05/2016	2134028	Svensen Vannistien	TI	5	3,61	101	Lulus Tepat 8 Semester
5	23/05/2016	2134031	Viola Dwi Putri Kebaar	TI	5	2,89	99	Lulus Tepat 8 Semester
6	23/05/2016	2143013	Muzita Hardo	SI	3	3,14	63	Lulus Tepat 8 Semester
7	23/05/2016	2143015	Prima Jumi Rahayu	SI	3	3,40	63	Lulus Tepat 8 Semester
8	23/05/2016	2143021	Septian Harhemelda	SI	3	3,34	63	Lulus Tepat 8 Semester
9	23/05/2016	2143022	Sri Sulawati	SI	3	3,56	63	Lulus Tepat 8 Semester
10	23/05/2016	2133018	Erida Syanti	SI	5	2,89	107	Lulus Tepat 8 Semester
11	23/05/2016	2133013	Dian Ardi	SI	5	2,31	85	Lulus Lebih Dari 8 Semes.
12	23/05/2016	2124045	Supriadi	TI	7	3,03	139	Lulus Tepat 8 Semester
13	23/05/2016	2123030	Hikmatul Hayati	SI	7	2,58	133	Lulus Lebih Dari 8 Semes.
14	23/05/2016	2123027	Firmandi	SI	7	3,22	133	Lulus Tepat 8 Semester
15	23/05/2016	2123022	Endang Opianity	SI	7	2,88	140	Lulus Tepat 8 Semester
16	23/05/2016	2123007	Redi Adriadi	TI	7	2,88	144	Lulus Tepat 8 Semester
17	23/05/2016	2123036	Ira Maya Wulandari	SI	7	2,40	125	Lulus Lebih Dari 8 Semes.
18	23/05/2016	2124013	Dwi Guna Putra H	TI	7	3,04	139	Lulus Tepat 8 Semester
19	23/05/2016	2124023	M. Arafat	TI	7	2,54	131	Lulus Lebih Dari 8 Semes.
20	23/05/2016	2124028	Mery Christiany	TI	7	3,01	136	Lulus Tepat 8 Semester
21	23/05/2016	2124031	Muhammad Rayhan	TI	7	3,10	134	Lulus Tepat 8 Semester
22	23/05/2016	2144022	Syahmandar	TI	3	2,79	63	Lulus Tepat 8 Semester

Gambar 19. Form Hasil Prediksi

5. Simpulan

Berdasarkan hasil eksperimen, analisis dan evaluasi yang dilakukan, maka dapat disimpulkan bahwa :

1. Jumlah matakuliah yang diambil, jumlah mata kuliah mengulang, pengambilan mata kuliah tertentu/prasyarat (SKS), variabel Indeks Prestasi Kumulatif (IPK) mempengaruhi masa studi.
2. Alur kerja konsep *data mining* menggunakan algoritma C4.5 untuk prediksi masa studi, cukup sederhana, yaitu dimulai dari proses *data selection, data cleaning, data transformation*
3. Secara umum algoritma C4.5 untuk membangun pohon keputusan hanya perlu 2 rumus, yaitu Perhitungan *Entropy* dan *Gain*
4. Dalam melakukan prediksi masa studi, pemakaian teknik Pohon keputusan (*decision tree*) dirasa cukup efektif .
5. *Output* yang dihasilkan oleh sistem menjadi bahan pertimbangan bagi Program Studi untuk

memberi saran atau rekomendasi sehingga mahasiswa dapat lulus tepat waktu 8 semester.

6. Mencegah atau meminimalisir terjadinya mahasiswa dengan masa studi lebih dari delapan semester.
7. Membantu akademis khususnya prodi memberikan informasi dini prediksi masa studi mahasiswa secara efektif kepada, mahasiswa dan orang tua/wali mahasiswa.

6. Referensi

- Dumai, S. (2014). Buku Wisuda IX STMIK Dumai. In 2014. Dumai: STMIK Dumai.
- Fattah, H. Al. (2007). *Analisis & Perancangan Sistem Informasi*. Yogyakarta: Penerbit Andi.
- Hafidh, K. (2015). *MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN METODE JACCARD COEFFICIENT (Studi Kasus: Mahasiswa Program Studi Teknik Informatika Jurusan Teknik Elektro Fakultas Teknik Universitas Tanjungpura)*. Universitas Tanjungpura.
- Indrajani. (2011). *Perancangan Basis Data Dalam All in 1*. Jakarta: PT. Elex Media Komputindo.
- Luthfi, K. & E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Meinanda, M. H., Annisa, M., Muhandri, N., & Suryadi, dan K. (2009). Prediksi masa studi sarjana dengan artificial neural network. *Internetworking Indonesia Journal*, 1(2), 31–35.
- Nataniel, D., & Hatta, H. R. (2009). Perancangan Sistem Informasi Terpadu Pemerintah Daerah Kabupaten Paser, 4(1), 47–54.
- Prabowo., Rahmadya., & H. (2013). *Penerapan Data Mining dengan MATLAB*. Bandung.: Rekayasa Sains.
- Sani Susanto & Dedy Suryadi. (2010). *Pengantar Data Mining*. Yogyakarta: Andi Offset.