Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5

Lydia Yohana Lumban Gaol¹, M. Safii², Dedi Suhendro³

¹STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

^{2,3}AMIK Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹lydiayohana.lgaol@gmail.com, ²m.safii@amiktunasbangsa.ac.id,

³dedi.su@amiktunasbangsa.ac.id

Abstract

Graduation is an important element in an accreditation assessment process of an institution or university. It is important to find out information about the predictions of student graduation in the Information Systems Study Program at STIKOM Tunas Bangsa Pematangsiantar, so that students who cannot graduate on time can be identified earlier. The application of data mining can be used to predict student graduation. Method that often used to predict student graduation is classification method. This research using C4.5 Algorithm. Data that used as training data are from alumni of Information Systems Study Program at STIKOM Tunas Bangsa and the final level student data of Information Systems Study Program at STIKOM Tunas Bangsa used as testing data. Through this research, it is expected that the results can provide information on predictions of student graduation on time and as a suggestion for the university in making good decisions for improvement in the future.

Keywords : Prediction, Graduation, Data mining, C4.5 Algorithm

Abstrak

Kelulusan merupakan salah satu elemen penting dalam proses penilaian akreditasi suatu institusi atau perguruan tinggi. Penting untuk mengetahui informasi mengenai prediksi kelulusan mahasiswa di Program Studi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar, sehingga sejak dini dapat diidentifikasi mahasiswa yang tidak dapat lulus tepat waktu. Penerapan data mining dapat digunakan untuk melakukan prediksi terhadap kelulusan mahasiswa. Metode yang sering digunakan untuk melakukan prediksi kelulusan mahasiswa adalah metode klasifikasi. Dalam penelitian ini Algoritma yang digunakan Algoritma C4.5. Adapun data yang digunakan sebagai data training adalah alumni prodi Sistem Informasi STIKOM Tunas Bangsa dan data mahasiswa tingkat akhir prodi Sistem Informasi STIKOM Tunas Bangsa digunakan sebagai data testing. Dengan adanya penelitian ini, diharapkan hasilnya dapat memberikan informasi mengenai prediksi kelulusan tepat waktu mahasiswa dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan demi perbaikan di masa yang akan datang.

Kata Kunci : Prediksi, Kelulusan, Data mining, Algoritma C4.5

1. Pendahuluan

STIKOM Tunas Bangsa merupakan salah satu perguruan tinggi IT yang terdapat di kota Pematangsiantar. STIKOM Tunas Bangsa sudah berdiri sejak tahun 2014. Adapun Program Studi yang berada dibawah naungan STIKOM Tunas Bangsa Pematangsiantar antara lain, Program Studi Teknik Informatika dan Sistem Informasi. Setiap 5 tahun sekali, STIKOM Tunas Bangsa Pematangsiantar selalu melakukan pembaharuan akreditasi kampus. Kelulusan merupakan salah satu elemen penting dalam proses penilaian akreditasi suatu institusi atau perguruan tinggi. Jika mahasiswa dapat lulus tepat waktu maka akan membantu proses penilaian akreditasi

suatu perguruan tinggi [1]. Keuntungan bagi mahasiswa yang lulus tepat waktu adalah mereka dapat segera bekerja setelah lulus tanpa perlu terikat pada dunia Pendidikan. Yang menjadi kendala adalah waktu kelulusan mahasiswa tidak selalu dapat dideteksi sejak dini. Hal tersebut berdampak terjadinya keterlambatan lulusan mahasiswa. Pentingnya untuk mengetahui informasi mengenai prediksi kelulusan mahasiswa Program Studi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar, sehingga sejak dini dapat diidentifikasi mahasiswa yang akan gagal atau tidak dapat lulus tepat waktu. Dengan mengetahui hasil prediksi tersebut, sehingga dapat diberikan beberapa alternatif pilihan untuk melakukan peningkatan kemampuan melalui bimbingan yang sifatnya khusus.

Penerapan *Data mining* dapat digunakan untuk melakukan prediksi terhadap kelulusan mahasiswa Program Studi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar. *Data mining* adalah gabungan dari beberapa disiplin ilmu yang menyatukan Teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualiasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar. *Data mining* sendiri dibagi menjadi beberapa kelompok berdasarkan tugas yang dilakukannya, yaitu deskripsi, estimasi, prediksi, klasifikasi dan pengklusteran [2]. Metode yang sering digunakan untuk melakukan prediksi kelulusan mahasiswa adalah metode klasifikasi. Dimana metode klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam *data mining* yaitu menggolongkan data [1]. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi atas informasi yang belum diketahui [2]. Adapun beberapa Algoritma yang dapat digunakan antara lain adalah Algoritma *Decision Tree C4.5*, *Artificial Neural Networks (ANN)*, *K-Nearest Neighbour (KNN)*, *Naïve Bayes*, Algoritma Genetika, *Rough Set*, Metode Berbasis Aturan, *Memory Based Reasoning*, dan *Support Vector Machine*.

Pada penelitian yang ini, metode yang akan digunakan oleh peneliti adalah Algoritma C4.5. Algoritma C4.5 sendiri sudah banyak digunakan untuk melakukan klasifikasi pada data yang memiliki atribut numerik maupun kategorial. Terdapat beberapa penelitian terdahulu yang membahas tentang prediksi kelulusan mahasiswa tepat waktu. Eko Prasetyo Rohmawan (2018) melakukan penelitian, yaitu Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* dan *Artificial Neural Network*, diperoleh nilai akurasi dengan melakukan pengujian dengan metode *Decision Tree* adalah sebesar 74,51% dan pengujian dengan menggunakan *Artificial Neural Network* sebesar 79,74%. Penelitian yang sama juga pernah dilakukan oleh Ratna Puspita Sari Putri (2018), yaitu penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. Pada penelitian ini diperoleh atribut yang paling dominan dalam kelulusan mahasiswa adalah IPK. Dan diperoleh nilai *precision* 63,93%, *recall* 60,73% dan akurasi 60,52%.

Perlu dilakukan evaluasi dengan melakukan perhitungan terhadap kinerja dari pohon keputusan yang telah dibangun sebelumnya. Pengukuran kinerja dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* diperoleh dengan membandingkan hasil prediksi dengan menggunakan aplikasi dengan hasil yang sebenarnya. Untuk mengukur kinerja dari pohon keputusan digunakan *precision*, *recall* dan akurasi. Pada penelitian ini, penulis akan melakukan penelitian terhadap prediksi kelulusan mahasiswa terkhusus pada Program Studi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar dengan menggunakan Algoritma C4.5. Dilakukan pengukuran terhadap pohon keputusan yang terbentuk untuk memperoleh nilai *precision*, *recall* dan akurasinya. Sebagai *data training* yang akan digunakan oleh peneliti adalah data alumni mahasiswa yang berasal dari Program Studi Sistem Informasi STIKOM Tunas Bangsa Pematangsiantar Angkatan 2014 dan 2015.

Sedangkan untuk *data testing* yang akan digunakan adalah data mahasiswa tingkat akhir Angkatan 2016. Dari hasil pengolahan *data training* dan *data testing* dapat diketahui informasi kelulusan mahasiswa yang dapat dijadikan dasar dalam pengambilan keputusan untuk meningkatkan jumlah lulusan di Program Studi Sistem Informasi STIKOM Tunas Bangsa Pematangsiantar.

2. Metodologi Penelitian

2.1. Data Mining

Data mining merupakan suatu proses penggalian atau penambangan informasi untuk menemukan informasi yang tidak ditemukan sebelumnya dari sekumpulan data besar (big data) atau repository database lainnya [3].

Data Mining bukan sekedar terkumpul data saja tetapi mencakup analisis dan prediksi dari informasi yang ingin ditampilkan. Data yang dikumpulkan disimpan dalam database kemudian diproses sehingga dapat dijadikan untuk pengambilan keputusan dalam melihat informasi yang akan di gunakan [4].

2.2. Algoritma C4.5

Algoritma C4.5 pertama sekali diperkenalkan oleh Quinlan pada tahun 1993. Algoritma C4.5 merupakan pengembangan dari Algoritma ID3 (*Iterative Dichotomiser 3*). Algoritma C4.5 merupakan salah satu algoritma yang digunakan dalam membentuk sebuah pohon keputusan. Dimana metode pohon keputusan dapat mengubah kumpulan data yang sangat besar menjadi sebuah pohon keputusan yang dapat merepresentasikan aturan [5]. Adapun kelebihan dari Algoritma C4.5 dibandingkan dengan algoritma ID3 adalah mampu mengatasi nilai yang hilang (*missing value*), mampu mengatasi data yang bertipe kontinu, dan melakukan pemangkasan pohon (*prunning tree*) [6]. Tahapan-tahapan pembuatan pohon keputusan dengan menggunakan algoritma C4.5 antara lain sebagai berikut [5].

a) Menyiapkan *data training*

Data berasal dari data yang sudah dikumpulkan sebelumnya dan di kelompokkan kedalam kelas tertentu.

b) Menentukan akar dari pohon

Akar di pilih dari atribut dengan cara menghitung nilai *gain* dari semua atribut, dan nilai *gain* yang paling besar akan menjadi akar pertama. Sebelum menghitung nilai *gain*, sebaiknya hitung nilai *entropy* terlebih dahulu. Berikut merupakan rumus yang digunakan dalam menghitung nilai *entropy*:

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

Keterangan :

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

c) Hitung nilai *gain* dengan menggunakan rumus sebagai berikut :

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

$|S_i|$ = Jumlah Kasus pada partisi ke-i

$|S|$ = Jumlah Kasus dalam S

d) Kemudian ulangi langkah kedua sampai semua *record* terpartisi secara sempurna.

e) Proses partisi pohon keputusan akan berhenti jika :

- a. Semua *record* pada simpul N mendapatkan kelas yang sama.
- b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
- c. Tidak ada *record* didalam cabang yang kosong.

2.3. Pohon Keputusan (*Decision Tree*)

Pohon keputusan menggunakan struktur representasi pohon (*tree*) dimana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai atribut dan daun merepresentasikan kelas [7]. Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [5].

3. Hasil dan Pembahasan

Dalam penelitian ini, untuk mengetahui mahasiswa dapat lulus tepat waktu adalah lama studi. Jika Lama studi mahasiswa adalah ≤ 4 tahun, maka dapat diketahui bahwa mahasiswa tersebut lulus tepat waktu (*Yes*). Jika Lama studi mahasiswa adalah > 4 tahun, maka dapat diketahui bahwa mahasiswa tersebut tidak lulus tepat waktu (*No*). Oleh karena itu, kriteria lama studi digunakan sebagai label.

Rumus untuk mencari *entropy* adalah sebagai berikut.

A. Perhitungan mencari *Entropy*

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

maka :

$$Entropy(\text{Total}) = \left(-\frac{\text{sum}(\text{yes})}{\text{total}} * \log_2 \left(\frac{\text{sum}(\text{yes})}{\text{total}} \right) \right) + \left(-\frac{\text{sum}(\text{no})}{\text{total}} * \log_2 \left(\frac{\text{sum}(\text{no})}{\text{total}} \right) \right)$$

$$Entropy(\text{Total}) = \left(-\frac{180}{200} * \log_2 \left(\frac{180}{200} \right) \right) + \left(-\frac{20}{200} * \log_2 \left(\frac{20}{200} \right) \right) = 0.468995594$$

Tabel 1. *Entropy* total kasus

Total Kasus	Lama Studi		<i>Entropy</i>
	≤ 4 Tahun (Ya)	> 4 Tahun (Tidak)	
200	180	20	0.468995594

1. Perhitungan *entropy* IPK

Sebelum menghitung *entropy* untuk IPK diperlukan untuk melakukan pengelompokan IPK yang akan dibagi menjadi 3 kategori yaitu tinggi sangat memuaskan dan rendah. Untuk IPK kurang dari 2,75 akan masuk kategori memuaskan. Untuk IPK yang berada diantara 2,76 – 3,50 akan masuk kategori sangat memuaskan. Dan untuk IPK yang berada diatas 3,50 akan masuk kategori *cumlaude*.

$$Entropy(\text{IPK, Cumlaude}) = \left(-\frac{90}{93} * \log_2 \left(\frac{90}{93} \right) \right) + \left(-\frac{3}{93} * \log_2 \left(\frac{3}{93} \right) \right) = 0.205592508$$

$$Entropy(\text{IPK, Sangat memuaskan}) = \left(-\frac{86}{93} * \log_2 \left(\frac{86}{93} \right) \right) + \left(-\frac{7}{93} * \log_2 \left(\frac{7}{93} \right) \right) =$$

$$0.385285119$$

$$Entropy(\text{IPK, Memuaskan}) = \left(-\frac{4}{14} * \log_2 \left(\frac{4}{14} \right) \right) + \left(-\frac{10}{14} * \log_2 \left(\frac{10}{14} \right) \right) =$$

$$0.863120569$$

2. Perhitungan *entropy* Status mahasiswa

$$Entropy(\text{Status Mahasiswa, Lulus}) = \left(-\frac{89}{101} * \log_2 \left(\frac{89}{101} \right) \right) + \left(-\frac{12}{101} * \log_2 \left(\frac{12}{101} \right) \right) =$$

$$0.525935984$$

$$Entropy(\text{Status Mahasiswa, Aktif}) = \left(-\frac{91}{99} * \log_2 \left(\frac{91}{99} \right) \right) + \left(-\frac{8}{99} * \log_2 \left(\frac{8}{99} \right) \right) =$$

$$0.405020133$$

3. Perhitungan *entropy* Absensi

Sebelum dilakukan perhitungan *entropy* untuk absensi terlebih dahulu dilakukan pengelompokan absensi. Untuk nilai absensi yang berada diatas 90 akan masuk kategori sangat baik. Untuk absensi yang berada diantara 76- 90 akan masuk ke dalam kategori baik. Untuk absensi yang berada diantara 61-75 akan masuk ke dalam kategori cukup. Untuk absensi yang berada dibawah 60 masuk kedalam kategori kurang.

$$Entropy (\text{Absensi, Sangat baik}) = \left(-\frac{91}{93} * \log_2 \left(\frac{91}{93} \right) \right) + \left(-\frac{2}{93} * \log_2 \left(\frac{2}{93} \right) \right) = 0.149811367$$

$$Entropy (\text{Absensi, Baik}) = \left(-\frac{78}{85} * \log_2 \left(\frac{78}{85} \right) \right) + \left(-\frac{7}{85} * \log_2 \left(\frac{7}{85} \right) \right) = 0.410416142$$

$$Entropy (\text{Absensi, Cukup}) = \left(-\frac{10}{18} * \log_2 \left(\frac{10}{18} \right) \right) + \left(-\frac{8}{18} * \log_2 \left(\frac{8}{18} \right) \right) = 0.99107606$$

$$Entropy (\text{Absensi, Kurang}) = \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4} \right) \right) + \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4} \right) \right) = 0.811218124$$

4. Perhitungan *entropy* Status bekerja

$$Entropy (\text{Status bekerja, Ya}) = \left(-\frac{53}{70} * \log_2 \left(\frac{53}{70} \right) \right) + \left(-\frac{17}{70} * \log_2 \left(\frac{17}{70} \right) \right) = 0.799759411$$

$$Entropy (\text{Status bekerja, Tidak}) = \left(-\frac{127}{130} * \log_2 \left(\frac{127}{130} \right) \right) + \left(-\frac{3}{130} * \log_2 \left(\frac{3}{130} \right) \right) = 0.158384407$$

B. Perhitungan mencari Gain

Rumus untuk menghitung *gain* adalah sebagai berikut

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

1. Perhitungan *gain* IPK

$$Gain (\text{total, IPK}) = Entropy(\text{total}) - \sum_{i=1}^n \frac{|IPK|}{|Total|} * Entropy(IPK)$$

$$Gain (\text{total, IPK}) = 0.468995594 - \left(\frac{93}{200} * 0.205592508 \right) + \left(\frac{93}{200} * 0.385285119 \right) + \left(\frac{14}{200} * 0.863120569 \right) = 0.13381905762$$

2. Perhitungan *gain* Status mahasiswa

$$Gain (\text{total, SM}) = Entropy(\text{total}) - \sum_{i=1}^n \frac{|SM|}{|Total|} * Entropy(SM)$$

$$Gain (\text{total, SM}) = 0.468995594 - \left(\frac{101}{200} * 0.525935984 \right) + \left(\frac{99}{200} * 0.405020133 \right) = 0.00291295624$$

3. Perhitungan *gain* Absensi

$$Gain (\text{total, Absensi}) = Entropy(\text{total}) - \sum_{i=1}^n \frac{|Absensi|}{|Total|} * Entropy(Absensi)$$

$$Gain (\text{total, Absensi}) = 0.468995594 - \left(\frac{93}{200} * 0.149811367 \right) + \left(\frac{85}{200} * 0.410416142 \right) + \left(\frac{18}{200} * 0.99107606 \right) + \left(\frac{4}{200} * 0.811218124 \right) = 0.11948524012$$

4. Perhitungan *gain* Status Bekerja

$$Gain (\text{total, SB}) = Entropy(\text{total}) - \sum_{i=1}^n \frac{|SB|}{|Total|} * Entropy(SB)$$

$$Gain (\text{total, SB}) = 0.701471460 - \left(\frac{70}{200} * 0.799759411 \right) + \left(\frac{130}{200} * 0.158384407 \right) = 0.0861299356$$

Tabel 2. Perhitungan Node 1

Atribut	Kasus (S)	Yes (S1)	No (S2)	Entropy	Gain
Total	200	180	20	0.468995594	

Atribut	Kasus (S)	Yes (S1)	No (S2)	Entropy	Gain
IPK					0.13381905762
Cumlaude	93	90	3	0.205592508	
Sangat Memuaskan	93	86	7	0.385285119	
Memuaskan	14	4	10	0.863120569	
Status Mahasiswa					0.00291295624
Lulus	101	89	12	0.525935984	
Aktif	99	91	8	0.405020133	
Absensi					0.11948524012
Sangat Baik	93	91	2	0.149811367	
Baik	85	78	7	0.410416142	
Cukup	18	10	8	0.991076060	
Kurang	4	1	3	0.811218124	
Status Bekerja					0.08612993560
Ya	70	53	17	0.799759411	
Tidak	130	127	3	0.158384407	

Pada tabel diatas dapat kita lihat bahwa atribut status mahasiswa memiliki *gain* yang paling tinggi yaitu 0.13381905762. Maka atribut IPK akan menjadi *node* akar. Selanjutnya akan dilakukan perhitungan terhadap cara yang sama untuk mendapatkan *node* berikutnya. Pada atribut IPK terdapat tiga partisi yaitu *cumlaude*, sangat memuaskan dan memuaskan. Dari hasil perhitungan diketahui bahwa pada partisi *cumlaude* memiliki nilai *entropy* 0.205592508, partisi sangat memuaskan memiliki nilai *entropy* 0.385285119 dan partisi memuaskan memiliki nilai *entropy* 0.863120569 pada kasus Yes (≤ 4 tahun). Sehingga dapat dipastikan semua status mahasiswa aktif akan menghasilkan label No (> 4 tahun). Dan nilai *entropy* tertinggi ada pada partisi lulus yaitu 0.341510030 yang akan digunakan sebagai *entropy* total untuk mencari *Node* 2, maka pohon keputusan akan dibuat sebagai berikut.

Gambar 1. Pohon keputusan 1

Sedangkan untuk partisi lulus masih terdapat dua keputusan Yes (≤ 4 tahun) dan No (> 4 tahun), sehingga masih perlu dilakukan pencarian *node* berikutnya dengan menggunakan data yang telah direduksi dengan melakukan filter status mahasiswa = lulus. Atribut yang perlu dihitung kembali adalah IPK, absensi, dan status bekerja. Sehingga diperoleh data hasil perhitungan *gain* untuk *node* 2 adalah sebagai berikut.

Tabel 3. Perhitungan *Node* 2

Atribut	Kasus (S)	Yes (S1)	No (S2)	Entropy	Gain
Total	200	180	20	0.468995594	
Status Mahasiswa					0.37458185727
Lulus	46	45	1	0.151096971	
Aktif	47	45	2	0.253878440	
Absensi					0.40872319525
Sangat Baik	64	64	0	0.000000000	
Baik	25	23	2	0.402179190	
Cukup	2	2	0	0.000000000	
Kurang	2	1	1	1.000000000	
Status Bekerja					0.40475321611
Ya	23	20	3	0.558629373	
Tidak	70	70	0	0.000000000	

Diperoleh nilai *gain* tertinggi dari atribut absensi yaitu 0.40475321611. Sehingga atribut status bekerja dapat menjadi *node* berikutnya setelah status bekerja. Maka pohon keputusan dapat dibuat sebagai berikut.

Gambar 2. Pohon keputusan 2

Atribut status bekerja memiliki 2 partisi, yaitu ya dan tidak. Dari hasil perhitungan *gain* untuk *node* 2 diketahui nilai *entropy* dari partisi tidak adalah 0 (nol). Sehingga dapat diketahui bahwa status bekerja tidak, akan menghasilkan label Yes (≤ 4 tahun). Dan nilai *entropy* yang paling tinggi adalah pada partisi status bekerja ya yaitu sebesar 0.558629373. Nilai *entropy* tersebut yang akan digunakan sebagai *entropy* total untuk melakukan perhitungan *node* 3 sebagai berikut.

Tabel 4. Perhitungan *Node* 3

Atribut	Kasus (S)	Yes (S1)	No (S2)	Entropy	Gain
Total	23	20	3	0.558629373	
Status Mahasiswa					0.36731356630
Lulus	14	13	1	0.151096971	
Aktif	9	7	2	0.253878440	
Absensi					0.22686812691
Sangat Baik	5	5	0	0.000000000	
Baik	14	12	2	0.402179190	
Cukup	2	2	0	0.000000000	
Kurang	2	1	1	1.000000000	

Diperoleh nilai *entropy* yang paling tinggi dari atribut absensi yaitu 0.402179190. Sehingga atribut absensi dapat dijadikan *node* ke 3. Atribut absensi memiliki empat partisi yaitu sangat baik, baik, cukup dan kurang. Dari hasil perhitungan *node* 3 diketahui partisi sangat baik dan cukup memiliki nilai *entropy* 0 (nol) berarti dapat dipastikan bahwa partisi absensi sangat baik akan menghasilkan label Yes (≤ 4 tahun). Partisi kurang memiliki nilai *entropy* 1 (satu) berarti dapat dipastikan bahwa partisi absensi sangat baik akan menghasilkan label No (> 4 tahun). Dan partisi baik memiliki nilai *entropy* 0.402179190. Maka hasil akhir pohon keputusan dapat dibuat seperti pada gambar berikut.

Gambar 3. Hasil akhir pohon keputusan

Setelah dilakukan perhitungan dan pengujian pada data dengan menggunakan Algoritma C4.5 maka akan didapatkan pola keputusan akhir.

Gambar 4. Menampilkan *decision tree*

Dari pohon keputusan yang terbentuk dengan menggunakan algoritma C4.5 diperoleh *rule* berupa teks yang terbentuk dari proses penggalan data.

Tree

```

IPK = Cumlaude
|   SB = Tidak: <= 4 tahun {> 4 tahun=0, <= 4 tahun=70}
|   SB = Ya
|   |   Absensi = Baik: <= 4 tahun {> 4 tahun=2, <= 4 tahun=12}
|   |   Absensi = Cukup: <= 4 tahun {> 4 tahun=0, <= 4 tahun=2}
|   |   Absensi = Kurang: > 4 tahun {> 4 tahun=1, <= 4 tahun=1}
|   |   Absensi = Sangat Baik: <= 4 tahun {> 4 tahun=0, <= 4 tahun=5}
IPK = Memuaskan: > 4 tahun {> 4 tahun=10, <= 4 tahun=4}
IPK = Sangat memuaskan: <= 4 tahun {> 4 tahun=7, <= 4 tahun=86}
  
```

Gambar 5. Deskripsi *decision tree*

Gambar diatas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan yang telah terbentuk dengan menggunakan Algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa Algoritma C4.5 baik digunakan dalam proses penggalan data untuk menarik beberapa kesimpulan yang divisualisasikan ke dalam pohon keputusan. Berikut merupakan *rule* yang diperoleh pohon keputusan yang terbentuk.

1. Jika IPK = Cumlaude dan Status Bekerja = Tidak, maka Lama studi ≤ 4 tahun (Lulus tepat waktu).
2. Jika IPK = Cumlaude, Status Bekerja = Ya dan Absensi = Baik, maka Lama studi ≤ 4 tahun (Lulus tepat waktu).
3. Jika IPK = Cumlaude, Status Bekerja = Ya dan Absensi = Cukup, maka Lama studi ≤ 4 tahun (Lulus tepat waktu).
4. Jika IPK = Cumlaude, Status Bekerja = Ya dan Absensi = Kurang, maka Lama studi > 4 tahun (tidak Lulus tepat waktu).
5. Jika IPK = Cumlaude, Status Bekerja = Ya dan Absensi = Sangat Baik, maka Lama studi ≤ 4 tahun (Lulus tepat waktu).
6. Jika IPK = Memuaskan, maka Lama studi > 4 tahun (tidak Lulus tepat waktu).
7. Jika IPK = Sangat Memuaskan, maka Lama studi ≤ 4 tahun (Lulus tepat waktu).

Dari proses pengolahan data mahasiswa dengan menggunakan *Software RapidMiner* didapatkan hasil akurasi dengan penerapan Algoritma C4.5 sebesar 90.00%. Untuk lebih jelasnya dapat dilihat pada gambar berikut.

accuracy: 90.00%

	true > 4 tahun	true <= 4 tahun	class precision
pred. > 4 tahun	1	1	50.00%
pred. <= 4 tahun	5	53	91.38%
class recall	16.67%	98.15%

Gambar 6. Nilai akurasi Algoritma C4.5

PerformanceVector

```
PerformanceVector:
accuracy: 90.00%
ConfusionMatrix:
True:  > 4 tahun    <= 4 tahun
> 4 tahun:    1      1
<= 4 tahun:    5     53
precision: 91.38% (positive class: <= 4 tahun)
ConfusionMatrix:
True:  > 4 tahun    <= 4 tahun
> 4 tahun:    1      1
<= 4 tahun:    5     53
recall: 98.15% (positive class: <= 4 tahun)
ConfusionMatrix:
True:  > 4 tahun    <= 4 tahun
> 4 tahun:    1      1
<= 4 tahun:    5     53
AUC (optimistic): 0.883 (positive class: <= 4 tahun)
AUC: 0.728 (positive class: <= 4 tahun)
AUC (pessimistic): 0.574 (positive class: <= 4 tahun)
```

Gambar 7. Nilai *Performance Vektor* Algoritma C4.5

Dari hasil *performance vector* diatas dapat diketahui bahwa penerapan Algoritma C4.5 dalam memprediksi kelulusan dengan menggunakan *software RapidMiner* memperoleh tingkat akurasi sebesar 90.00%. Dari gambar diatas dijelaskan bahwa prediksi lulus ≤ 4 tahun memiliki nilai 53 dan prediksi lulus > 4 tahun memiliki nilai 5 dengan nilai *precision* sebesar 91.38% dan nilai *recall* sebesar 98.15%.

4. Kesimpulan

Berdasarkan pembahasan pada bab - bab sebelumnya, mengenai Prediksi Kelulusan Mahasiswa STIKOM Tunas Bangsa Pematangsiantar Prodi Sitem Informasi dengan Menggunakan Algoritma C4.5, maka dapat diperoleh kesimpulan sebagai berikut:

- Penerapan Algoritma C4.5 pada prediksi kelulusan mahasiswa tingkat akhir prodi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar, diperoleh tingkat akurasi sebesar 90.00% dengan *precision* sebesar 91.38% dan *recall* sebesar 98.15%.
- Berdasarkan hasil penelitian yang telah dilakukan, menunjukkan bahwa atribut IPK (C1) memiliki pengaruh paling tinggi dalam menentukan kelulusan mahasiswa.
- Dengan adanya penerapan *data mining* Algoritma C4.5 dalam melakukan prediksi kelulusan mahasiswa STIKOM Tunas Bangsa prodi Sitem Informasi dapat membantu pihak institusi dalam mengambil keputusan dalam membantu meningkatkan jumlah lulusan mahasiswa prodi Sistem Informasi di STIKOM Tunas Bangsa Pematangsiantar.

Daftar Pustaka

- [1] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network," *J. Ilm. Matrik*, vol. 20, no. 1, pp. 21–30, 2018.
- [2] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [3] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi:

-
- 10.15575/join.v2i1.71.
- [4] A. S. Sunge, "Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 (Studi Kasus : PT Hankook Tire Indonesia)," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 23–24, 2018.
 - [5] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, 2018, doi: 10.31227/osf.io/x6svc.
 - [6] W. Rumaenda, Y. Wilandari, and D. Safitri, "Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner dan Algoritma C4.5 (Studi Kasus UPT Puskesmas Ponjong I, Gunungkidul)," *J. Gaussian*, vol. 5, no. 2, pp. 299–309, 2016.
 - [7] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 13, no. 1, p. 33, 2018, doi: 10.30872/jim.v13i1.629.