

# Prediksi Kelulusan Mahasiswa Fakultas Teknologi Informasi Umby Menggunakan Metode Decision Tree Penerapan Algoritma C4.5

Vidya Anggraini Nurislamiy<sup>1</sup>, Anief Fauzan Rozi<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta  
Yogyakarta, Indonesia

e-mail: <sup>1</sup>vidyaanherman@gmail.com, <sup>2</sup>anief@mercubuana-yogya.ac.id

Diajukan: 04 September 2020; Direvisi: 14 April 2021; Diterima: 04 Mei 2021

## Abstrak

*Kelulusan merupakan salah satu unsur penting bagi pihak fakultas dalam penilaian akreditasi. Sehingga jika mahasiswa lulus tepat waktu maka akan membantu penilaian akreditasi terhadap penilaian suatu fakultas hingga perguruan tinggi. Kelulusan tepat waktu sendiri merupakan salah satu tolak ukur hasil kinerja akademik mahasiswa. Pada Fakultas Teknologi Informasi (FTI) Universitas Mercu Buana Yogyakarta masih banyak ditemui permasalahan mengenai mahasiswa yang menyelesaikan masa studi lebih dari waktu yang ditetapkan, hal ini tentu saja merugikan pihak fakultas yang akan membuat akreditasi fakultas kurang maksimal, serta terlalu banyaknya mahasiswa aktif yang membuat kegiatan belajar mengajar kurang ideal. Berdasarkan hal tersebut, maka dirasa diperlukan untuk melakukan analisis dan klasifikasi pola-pola kriteria kelulusan mahasiswa tepat waktu, untuk melakukan hal tersebut maka dirasa metode paling cocok yang digunakan ialah data mining Algoritma C4.5. berdasarkan dari perhitungan Algoritma C4.5 dengan atribut input berupa Indeks Prestasi Semester 5-7, Program Studi, serta sks tempuh hingga semester 7 maka menghasilkan akurasi sebesar 82,8897% dengan IPS semester 7 menjadi root tree dan SKS kumulatif hingga semester 7 menjadi child node.*

**Kata kunci :** Algoritma C4.5, Data Mining, Kelulusan, FTI UMBY

## Abstract

*Graduation is the important element of accreditation assessment for the faculty. The students who are graduate appositely would help assess the accreditation of a faculty and the college assessment. Graduation appositely itself is one of measure student academic performance results. Faculty of Information Technology Universitas Mercu Buana Yogyakarta still have some problems encountered regarding students who complete their study period more than the stipulated time, this is affected to the faculties which make the faculty accreditation is less optimal, and a huge number of active students who make teaching and learning activities is less than ideal. According to this point, it is necessary to carry out the analysis and classification patterns of student graduation criteria appositely, the most suitable method to use is data mining Algorithm C4.5. According to the calculation of the C4.5 Algorithm with input attributes in the form of 5 to 7 Achievement Index Semester, Study Program, and credits until 7th semester, it produces an accuracy of 82,8897% with 5th Achievement Index Semester as the root tree and the credits until 7th semester Semester becomes a child node.*

**Keywords :** C4.5 Algorithm, Data Mining, Graduation, FTI UMBY

## 1. Pendahuluan

penting bagi pihak fakultas dalam penilaian akreditasi. Sehingga jika mahasiswa lulus tepat waktu maka akan membantu penilaian akreditasi terhadap penilaian suatu fakultas hingga perguruan tinggi. Kelulusan tepat waktu sendiri merupakan salah satu tolak ukur hasil kinerja akademik mahasiswa. Ketentuan masa studi sendiri sudah diatur dalam ketetapan Peraturan Menteri dan Pendidikan Kebudayaan Indonesia yang menjelaskan bahwa kompetensi lulusan bagi mahasiswa S1 dapat menyelesaikan beban wajib minimal 144 SKS dengan masa studi 4 tahun. [1]

Pada Fakultas Teknologi Informasi (FTI) Universitas Mercu Buana Yogyakarta masih banyak ditemui permasalahan mengenai mahasiswa yang menyelesaikan masa studi lebih dari waktu yang ditetapkan, hal ini tentu saja merugikan pihak fakultas yang akan membuat akreditasi fakultas kurang maksimal, serta terlalu banyaknya mahasiswa aktif yang membuat kegiatan belajar mengajar kurang ideal.

Berdasarkan hal tersebut, maka dirasa diperlukan untuk melakukan analisis dan klasifikasi pola-pola kriteria kelulusan mahasiswa tepat waktu, untuk melakukan hal tersebut maka dirasa metode paling cocok yang digunakan ialah data mining Algoritma C4.5, Algoritma C4.5 sendiri merupakan algoritma yang menggunakan klasifikasi pohon keputusan. Digunakanannya metode ini dalam penelitian ini untuk memanfaatkan data mahasiswa untuk menentukan pola-pola yang ada antara sejumlah atribut *input* dengan sebuah atribut target. Data mahasiswa yang

akan menjadi atribut *input* antara lain adalah IPS semester 5-7, jumlah SKS yang ditempuh saat semester 7, dan prodi. Atribut status diubah menjadi atribut *output* yang akan digunakan untuk mengklasifikasikan data mahasiswa yang terdiri dari “lulus tepat waktu” dan “lulus tidak tepat waktu”. Hasil dari analisis ini nanti akan digunakan pihak fakultas untuk bahan evaluasi memperbaiki sistem pembelajaran dan menangani calon-calon mahasiswa yang akan lulus tidak tepat waktu. Sehingga fakultas dapat menghasilkan lulusan yang berkualitas. Dalam Penelitian yang berjudul “*Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5*” peneliti melakukan analisis kelulusan mahasiswa tepat waktu dengan analisis data mahasiswa Program Studi teknik Informatika Universitas Muhammadiyah Ponorogo angkatan 2012/2013 Penggunaan atribut penelitian ini diantaranya adalah NIM, jenis sekolah asal, asal daerah, pekerjaan orang tua, dan kelas. Atribut-atribut inilah yang akan menjadi bahan analisis ketepatan waktu studi mahasiswa. Dari pengukuran kinerja algoritma yang telah dilakukan, dapat disimpulkan untuk kasus ini algoritma C4.5 memiliki nilai akurasi sebesar 82%. Dalam hal ini berarti penggunaan algoritma masih ditemukan tingkat kesalahan/*error* sebesar 18%. [2].

Dalam penelitian berjudul “*Prediksi Kelulusan Mahasiswa dengan Metode Algoritma C45*” Peneliti melakukan penelitian memprediksi kelulusan mahasiswa sebagai bahan untuk melakukan kebijakan akademik, sehingga dapat menekan tingkat putus kuliah, *drop out* dan mengulang pada mahasiswa di universitas singaperbangsa karawang. Peneliti mengambil dari data historis dan biodata awal penerimaan mahasiswa, dari data tersebutlah untuk melakukan prediksi menggunakan algoritma C45, peneliti menggunakan data dari mahasiswa angkatan 2008 Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang dari hasil analisis ternyata bahwa untuk kelulusan yang dilihat dari *gender* ternyata *gender* perempuan lebih besar persentasenya dibandingkan dengan *gender* laki-laki, kemudian nilai *gain* yang didapat dari usia pada mahasiswa yang bekerja lebih kepada mahasiswa dengan kategori usia 1, dan nilai *gain* yang didapat dari usia pada mahasiswa yang tidak bekerja lebih kepada mahasiswa dengan kategori usia 3 [3]. Dalam Penelitian berjudul “*Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : Stmik Royal Kisaran)*” Penelitian ini bertujuan untuk mengetahui persoalan dalam memprediksi kelulusan tepat waktu pada perguruan tinggi. Metode yang peneliti gunakan adalah metode Klasifikasi dengan Algoritma C.45 *tools* yang digunakan di dalam pengolahan data adalah Rapidminer 5. Penyebab masalah yang sering terjadi kelulusan tidak tepat waktu pada lokasi penelitian peneliti adalah rendahnya ipk, tidak penuhnya sks, etika yang tidak baik, dan juga tidak adanya prestasi mahasiswa yang dicapai. data yang dikumpulkan yang akan dianalisis dari data kelulusan mahasiswa yang diberikan perguruan tinggi STMIK ROYAL. Dari hasil pengujian pada *tools* Rapidminer 5 menghasilkan sebuah akurasi dengan tingkat kinerja Algoritma C.45 yaitu 92,60% +/-1.60%. jadi dapat diketahui bahwa kelulusan mahasiswa tepat waktu diperguruan tinggi STMIK ROYAL Kisaran yang didapatkan dari data 35 mahasiswa aktif, mendapatkan tingkat kelulusan tepat waktu berjumlah 21 mahasiswa dari 35 mahasiswa [4]

Dalam penelitian berjudul “*Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya*” Penelitian ini dilakukan karena masih banyak ditemui permasalahan mengenai kinerja akademik mahasiswa dalam hal masa studi. Di program studi Si Filkom Universitas Brawijaya masih banyak mahasiswa yang menyelesaikan studi melebihi batas masa studi. Berdasarkan data yang terdapat pada *website*. Peneliti melakukan penelitian terhadap data mahasiswa dengan menggunakan pendekatan data *mining Naïve bayes* dengan membuat pola – pola kriteria kelulusan yang diimplementasikan pada *dashboard* sistem. Pada penelitian ini algoritma *data mining* klasifikasi yang digunakan adalah algoritma Naive Bayes. Algoritma ini dipilih karena algoritma Naive. Pengumpulan data pada penelitian ini diperoleh dari akademik FILKOM. Data yang diperoleh sebanyak 1352 *records* dengan asal data mahasiswa angkatan 2011 - 2016. Atribut yang dimiliki dari setiap data diantaranya adalah ID Mhs, Jenis kelamin, Jalur Masuk, IP. Untuk memprediksi kelulusan dapat dilakukan dengan memanfaatkan Weka CLI sebagai API yang memproses data dengan data *mining*. Selain itu hasil pengujian *black-box* dalam pengembangan sistem menunjukkan hasil yang *valid* untuk semua *fitur*, dan hasil pengujian *usability* menggunakan *System Usability Scale* menghasilkan nilai sebesar 57.5 dimana nilai tersebut termasuk kedalam *adjective rating good* namun tingkat *acceptability* masih tergolong *Low*. [5]. Dalam penelitian berjudul “*Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network*” Penelitian ini dilakukan dikarenakan kelulusan mahasiswa tepat waktu merupakan salah satu penilaian dalam proses akreditasi perguruan tinggi. Dalam penelitian ini fokus pada bagaimana prediksi kelulusan mahasiswa tepat waktu menggunakan algoritma *decision tree* dan *artificial neural network*. Penelitian ini. Data yang akan digunakan data induk dan data akademik mahasiswa angkatan 2006 - 2010 yang sudah dinyatakan lulus. Data ini memiliki atribut NIM, jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orang tua, IP semester 1-4, IPK semester 1-4 dan keterangan lulus. Berdasarkan hasil pengujian metode *decision tree* memiliki akurasi sebesar 74,51% dan *artificial neural network* sebesar 79,74%, Metode *artificial neural network* memiliki akurasi lebih tinggi jika dibandingkan dengan *decision tree* karena data yang digunakan dalam penelitian ini berupa data label [6]. Data *mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data [7]. Pohon keputusan merupakan salah satu metode klasifikasi yang kuat dan terkenal. Metode pohon keputusan mengubah fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan, aturan

tersebut dapat dengan mudah untuk diinterpretasi oleh manusia. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel *input* dengan sebuah variabel target [8] [9]. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa *database* seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target. [10]

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, yaitu :

1. Mempersiapkan data nasabah. Data nasabah diambil dari *record* yang ada sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*.

Tahapan Algoritma *Decision Tree* C4.5 :

- a) Menyiapkan data latih
- b) Menentukan akar dari pohon
- c) Hitung nilai *Entropy* :

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

- S : Himpunan Kasus
- n : Jumlah partisi S
- $p_i$  : Proporsi dari  $S_i$  terhadap S

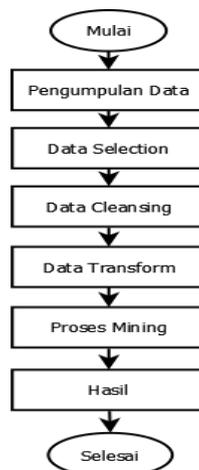
- d) Menghitung nilai *gain* :

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

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- $|S_i|$  : Jumlah kasus pada partisi ke i
- $|S|$  : Jumlah kasus dalam S

## 2. Metode Penelitian

Secara garis besar proses jalan penelitian ini dibagi menjadi enam tahapan yaitu : (1) Pengumpulan Data, (2) Data Selection, (3) Data Cleansing, (4) Data Transform, (5) Proses Mining, (6) Hasil. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1 Jalan Penelitian

**a. Pengumpulan Data**

Pada tahap ini akan dilakukan pengumpulan data yang berasal dari data historis mahasiswa yang disimpan oleh ICT Universitas Mercu Buana Yogyakarta data yang digunakan ialah data Mahasiswa Fakultas Teknologi Informasi Universitas Mercu Buana Yogyakarta Angkatan 2012-2015. Pengumpulan data dilakukan dengan metode dokumen primer.

**b. Data Selection**

Pada tahap ini akan dilakukan seleksi data yang dapat dipakai pada pengujian ini, proses ini dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.

**c. Data Cleansing**

Pada tahap ini akan dilakukan penghilangan data yang tidak lengkap atau menggabungkan data dengan data yang memiliki jenis yang sama.

**d. Data Transform**

Pada tahap ini dilakukan agar meningkatkan akurasi dan efisiensi algoritma. Pada tahapan ini dari seluruh data mahasiswa didapatkan data pengelompokan atribut yang digunakan untuk proses data *mining*.

**e. Proses Mining**

Pada tahap ini merupakan proses mencari pola atau informasi yang dibutuhkan dalam data dengan menggunakan Teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode yang digunakan pada penelitian ini ialah C4.5 (*decision tree*). Pada tahap ini akan dilakukan proses *mining* dibantu oleh tools WEKA.

**f. Hasil**

Pada tahap ini akan didapatkan *output* yang dicari yaitu berupa *decision tree*, serta tingkat keakuratan proses klasifikasi data setelah didapatkan *output* yang diinginkan maka akan dilakukan analisis dari hasil *decision tree* tersebut.

**3. Hasil dan Pembahasan**

Data yang akan diolah mempunyai beberapa kriteria yang merupakan syarat dalam pengolahan Data Mining dengan menggunakan teknik algoritma C4.5. Dimulai dengan perhitungan *Entropy* dan *Gain* untuk menentukan akar (*root*) dari pohon keputusan, sampai terbentuk pohon keputusan memprediksi kelulusan mahasiswa tepat waktu. Data yang akan diolah merupakan historis data Mahasiswa Fakultas Teknologi Informasi Angkatan 2012-2015 data diminta melalui ICT Universitas Mercu Buana Yogyakarta dengan jumlah data sebanyak 7.955 data yang terdiri dari 11 atribut yaitu NIM, nama, smuasal, mhstgllahir, jns\_kelamin, prodi, periode, ipk, ips, serta statusmhs.

nim	nama	smuasal	mhstgllahir	jns_kelamin	prodi	periode	sks	ipk	ips	statusmhs
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20121	16	3,19	0	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20123	39	3,05	3,19	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20131	61	3,05	2,96	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20133	79	2,94	3,05	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20141	96	2,79	2,56	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20143	108	2,81	2,12	A
12111001	MUHAMMAD ABDURROHMAN ALHAFIDH	SMKN 2 SEKAYU	05/01/1995	L	Informatika	20151	121	2,82	3,13	A

Adapun atribut yang digunakan dalam prediksi kelulusan tepat waktu adalah atribut nama, prodi, sks, ips, serta status mhs. Setelah tahap pengumpulan data dan seleksi data maka tahap selanjutnya yaitu melakukan proses *cleansing* Pada tahap ini peneliti akan melakukan *cleansing* data duplikasi pada atribut nama dan prodi serta melakukan penghapusan data tidak terpakai seperti pada atribut sks, data yang digunakan pada atribut sks hanyalah sks semester 7, peneliti juga akan melakukan data *cleansing* pada atribut ips yang digunakan hanyalah data ips semester 5-7, serta akan dilakukan pula proses data *cleansing* pada atribut statusmhs selain data pada semester 8. Selanjutnya akan dilakukan transform data sebagai berikut

Tahapan pertama peneliti melakukan perubahan nama atribut, yang akan dijelaskan pada tabel 1 :

Tabel 1 Transform tahap pertama

Atribut Sebelum	Atribut Sesudah
-----------------	-----------------

Sks	Sks Tempuh
Ips	Ips smt 5, ips smt 6, ips smt 7
Statusmhs	Keterangan (klasifikasi)

Tahapan kedua ialah melakukan proses *transform* data ips menjadi sebuah indeks dengan rentang seperti pada tabel 2 sebagai berikut :

Tabel 2 Transform tahap kedua

Rentang	Kategori
3,51-4,00	A
3,00-3,50	B
2,51-2,99	C
2,00-2,50	D
0,00-1,99	E

Tahapan ketiga ialah melakukan *transform* data keterangan dengan ketentuan, seperti pada tabel 3 berikut:

Tabel 3 Transform tahap ketiga

Indeks	Kategori
A, N, C, T	LULUS TIDAK TEPAT WAKTU
L	LULUS TEPAT WAKTU

Tahapan keempat ialah melakukan *transform* data pada atribut prodi dengan ketentuan seperti pada tabel 4 berikut :

Tabel 4 Transform tahap keempat

Indeks	Keterangan
Informatika	TI
Sistem Informasi	SI

Tahapan kelima ialah melakukan *transform* data pada atribut sks dengan ketentuan seperti pada table 5 berikut :

Tabel 5 Transform tahap kelima

Rentang	Keterangan
0-50	RENDAH
51-100	SEDANG
101-150	TINGGI

Tahapan keenam ialah melakukan proses perubahan struktur data,data ips belum menjadi data horizontal. Pada tahap ini akan dilakukan proses tranformasi data IPS menjadi *horizontal* serta penambahan atribut ips smt 5, ips smt 6, ips smt 7. Dari hasil pengolahan data tersebut maka menghasilkan data training sebagai berikut

Nama	Sks Tempuh	Prodi	IPS smt 5	IPS smt 6	IPS smt 7	Keterangan
MUHAMMAD ABDURROHMAN ALHAFIDH	TINGGI	TI	C	D	B	LULUS TIDAK TEPAT WAKTU
FARID MA RUF	SEDANG	TI	C	C	E	LULUS TIDAK TEPAT WAKTU
YATI YASINTA	TINGGI	TI	A	B	A	LULUS TEPAT WAKTU
MUHAMMAD ISLAMSYAH	TINGGI	TI	B	B	B	LULUS TEPAT WAKTU
DIYAN SAWARNI NING TYAS	SEDANG	TI	C	D	D	LULUS TIDAK TEPAT WAKTU
BUDI SANTOSO	TINGGI	TI	B	C	C	LULUS TIDAK TEPAT WAKTU
IRANKA RJI SAPUTRO	TINGGI	TI	B	C	B	LULUS TEPAT WAKTU
SERVIA R.Y	TINGGI	TI	A	C	A	LULUS TEPAT WAKTU
DWI ANDRI YULLANTO	TINGGI	TI	C	C	C	LULUS TIDAK TEPAT WAKTU
SONYA NOVELLA MARGARETA	TINGGI	TI	A	B	A	LULUS TIDAK TEPAT WAKTU
ARIF SETYAWAN	TINGGI	TI	C	C	B	LULUS TIDAK TEPAT WAKTU
ODE HENDIRA JAYA DILAGA	TINGGI	TI	A	B	C	LULUS TIDAK TEPAT WAKTU
HABIBUL HAKIM	TINGGI	TI	B	B	A	LULUS TIDAK TEPAT WAKTU
ADE CHRISTIAN PRIAMBODO	TINGGI	TI	B	A	A	LULUS TEPAT WAKTU
DWI CAHYO NUGROHO	TINGGI	TI	B	A	A	LULUS TEPAT WAKTU

Gambar 2 Data latihan

Tahap pembangunan pohon keputusan dilakukan setelah selesai menghitung nilai *entropy* total, nilai *entropy* dari tiap-tiap atribut, menghitung nilai *gain ratio*,menentukan *gain ratio* tertinggi. Pada tabel proses perhitungan *entropy* dan *gain* seperti pada tabel 6 berikut

Tabel 6 Perhitungan root node

No	Atribut	Niai atribut	S	Y	N	Entropy	Information Gain
TOTAL			263	109	154	0,2346971	
	SKS TEMPUH						0,586627775
		TINGGI	168	99	69	0,663515112	
		SEDANG	89	10	79	-2,29245287	
		RENDAH	6	0	6	0	
	PRODI						0,025086981
		TI	180	68	112	0,082614579	
		SI	83	41	42	0,485022132	
	IPS SMT 5						0,582613287
		A	59	39	20	0,72138771	
		B	103	57	46	0,610532912	
		C	55	10	45	-1,32822379	
		D	15	1	14	-3,29307188	
		E	31	2	29	-3,35396955	
	IPS SMT 6						0,361778368
		A	45	35	10	0,683633378	
		B	97	54	43	0,616104603	
		C	55	18	37	-0,17194044	
		D	30	0	30	0	
		E	33	2	31	-3,46943325	
	IPS SMT 7						1,087617875
		A	62	48	14	0,687253693	
		B	98	50	48	0,524179386	
		C	36	5	31	-1,87112028	
		D	21	0	21	0	
		E	41	1	40	-5,06145297	

Data dikelompokkan berdasarkan atribut dan nilai atribut nya lalu dihitung jumlah keseluruhan, jumlah mahasiswa yang lulus tepat waktu atau tidak, lalu dihitung nilai *entropy* dan *gain* masing-masing atributnya. Baris TOTAL kolom *ENTROPY* diatas dihitung dengan rumus, sebagai berikut :

$$Entropy(TOTAL) = \left(-\frac{109}{263} * \log_2\left(\frac{109}{263}\right)\right) + \left(-\frac{154}{263} * \log_2\left(\frac{154}{263}\right)\right)$$

$$Entropy(TOTAL) = 0,234697$$

Perhitungan *entropy* pada tiap nilai atribut dihitung dengan cara yang sama dengan *entropy* total

$$Entropy(SKS\ TEMPUH, TINGGI) = \left(-\frac{99}{168} * \log_2\left(\frac{99}{168}\right)\right) + \left(-\frac{69}{168} * \log_2\left(\frac{69}{168}\right)\right)$$

$$Entropy(SKS\ TEMPUH, TINGGI) = 0,66351511$$

Sementara itu nilai *GAIN* pada baris SKS TEMPUH dihitung dengan menggunakan rumus *Gain*, sebagai berikut :

$$Gain(TOTAL, SKS\ TEMPUH)$$

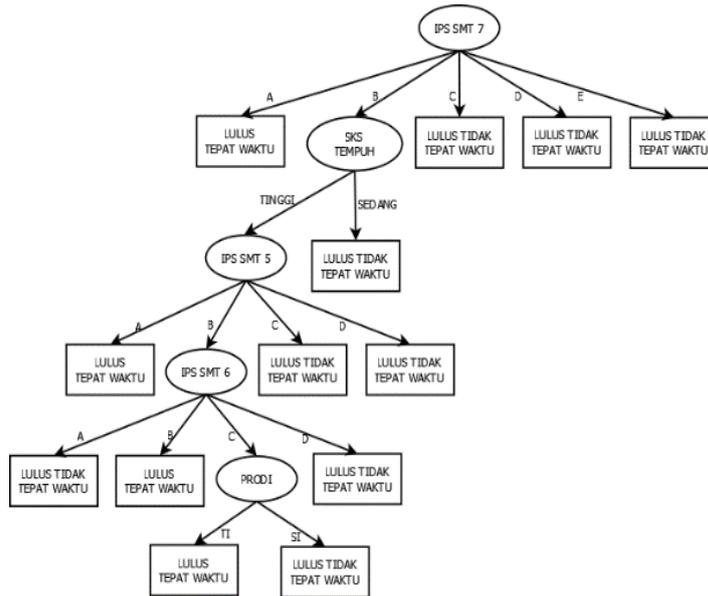
$$= Entropy(TOTAL) - \sum_{i=1}^n \frac{|sks\ tempuh|}{|total|} * Entropy(sks\ tempuh)$$

$$Gain(TOTAL, SKS\ TEMPUH)$$

$$= 0,234697 - \left(\left(\frac{168}{263} * 0,63351\right) + \left(\frac{89}{263} * -2,2924529\right) + \left(\frac{6}{263} * 0\right)\right)$$

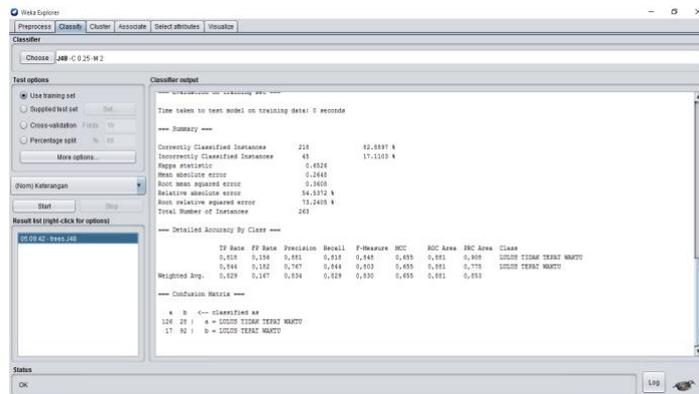
$$Gain(TOTAL, SKS\ TEMPUH) = 0,586627775$$

Dari hasil padaa tabel diatas dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah IPS SMT 7 yaitu sebesar 1,087617875 dengan demikian IPS SMT 7 menjadi *node* akar, ada 5 nilai atribut yaitu A, B, C, D, dan E. Tahapan selanjutnya adalah sama seperti sebelumnya mencari nilai *gain* tertinggi hingga menghasilkan tree akhir yaitu semua atribut menjadi tree tanpa pengulangan seperti pada gambar 3 :



Gambar 3 Hasil Pohon Keputusan

Selanjutnya akan dilakukan tahap pengujian menggunakan weka untuk melakukan pengecekan apakah tree hasil hitungan manual sudah tepat dan mengukur akurasi dari klasifikasi. Atribut yang digunakan sebagai label adalah keterangan, peneliti akan menganalisis faktor apa saja yang mempengaruhi mahasiswa dapat lulus tepat waktu. Data uji menggunakan format csv.



Gambar 4 Hasil Classify WEKA

Dari gambar 4 dapat dilihat dalam proses *classify* data yang memiliki 5 atribut yaitu sks tempuh, prodi, ips smt5, ips smt6, ips smt7 dan 1 atribut label keterangan. Mempunyai jumlah *record* atau full data sebanyak 263 data. Berdasarkan hasil pengujian menunjukkan kinerja algoritma mengklasifikasikan data sebanyak 218 data dengan presentase sebesar (82,8897%) serta data tidak berhasil di klasifikasikan dengan benar (*error*) sebanyak 45 data dengan presentase sebesar (17,1103%).

Selanjutnya hasil uji pohon keputusan yang dilakukan oleh algoritma C.45 (J48 di WEKA) menunjukkan bahwa ips smt 7 menjadi *root node* sementara sks tempuh menjadi *child node*. Gambar pohon keputusan dapat dilihat seperti pada gambar 5.



### Gambar 5 Pohon Keputusan pada WEKA

Pada hasil *decision tree* yang diolah oleh WEKA terdapat beberapa data janggal seperti jika IPS SMT 7 B dan SKS TEMPUH RENDAH maka akan LULUS TEPAT WAKTU hal ini dikarenakan pada uji coba ini tidak ada data yang diuji pada IPS SMT 7 B dan SKS TEMPUH RENDAH sehingga WEKA secara otomatis mengklasifikasikan LULUS TEPAT WAKTU, lalu pada *tree* jika IPS SMT 7 B, SKS TEMPUH TINGGI dan nilai IPS SMT 5 E maka akan LULUS TEPAT WAKTU pada uji coba ini juga tidak terdapat data dalam *rule* tersebut sehingga WEKA mengklasifikasikannya sebagai LULUS TEPAT WAKTU hal ini juga berlaku pada *rule* jika IPS SMT 7 B, SKS TEMPUH TINGGI, IPS SMT 5 B dan IPS SMT 6 E maka akan LULUS TEPAT WAKTU tidaklah *valid* dikarenakan tidak ada data yang digunakan dalam pengujian. Selanjutnya pada *rule* jika IPS SMT 7 B, SKS TEMPUH TINGGI, IPS SMT 5 B, dan IPS SMT 6 A maka LULUS TIDAK TEPAT WAKTU belum tentu *valid* dikarenakan pada pengujian ini hanya menggunakan 2 data.

Didapatlah hasil prediksi dari *decision tree* sebagai bahan acuan pihak fakultas untuk mahasiswa lulus tepat waktu sebagai berikut :

#### Prediksi Lulus Tepat Waktu :

1. Jika IPS SMT 7 A (3,51-4,00) maka LULUS TEPAT WAKTU
2. Jika IPS SMT 7 B (3,00-3,50) , SKS TEMPUH TINGGI (101-150) dan IPS SMT 5 A (3,51-4,00) maka LULUS TEPAT WAKTU
3. Jika IPS SMT 7 B (3,00-3,50) , SKS TEMPUH TINGGI (101-150), IPS SMT 5 B (3,00-3,50) dan IPS SMT 6 B (3,00-3,50) maka LULUS TEPAT WAKTU
4. Jika IPS SMT 7 B (3,00-3,50), SKS TEMPUH TINGGI (101-150), IPS SMT 5 B(3,00-3,50), IPS SMT 6 C (2,51-2,99) dan PRODI TI maka LULUS TEPAT WAKTU

#### Prediksi Lulus Tidak Tepat Waktu :

1. Jika IPS SMT 7 C (2,51-2,99) maka LULUS TIDAK TEPAT WAKTU
2. Jika IPS SMT 7 D (2,00-2,50) maka LULUS TIDAK TEPAT WAKTU
3. Jika IPS SMT 7 C (2,51-2,99) maka LULUS TIDAK TEPAT WAKTU
4. Jika IPS SMT 7 B (3,00-3,50) dan SKS TEMPUH RENDAH (0-50) maka LULUS TIDAK TEPAT WAKTU
5. Jika IPS SMT 7 B (3,00-3,50), SKS TEMPUH TINGGI (101-150), dan IPS SMT 5 C (2,51-2,99), maka LULUS TIDAK TEPAT WAKTU
6. Jika IPS SMT 7 B (3,00-3,50), SKS TEMPUH TINGGI (101-150), dan IPS SMT 5 D (2,00-2,50) maka LULUS TIDAK TEPAT WAKTU
7. Jika IPS SMT 7 B (3,00-3,50), SKS TEMPUH TINGGI (101-150), IPS SMT 5 B (3,00-3,50) dan IPS SMT 6 D (2,00-2,50) maka LULUS TIDAK TEPAT WAKTU
8. Jika IPS SMT 7 B (3,00-3,50), SKS TEMPUH TINGGI (101-150), IPS SMT 5 B (3,00-3,50), IPS SMT 6 C (2,51-2,99) dan PRODI SI maka LULUS TIDAK TEPAT WAKTU

## 4. Kesimpulan

Setelah melakukan analisis dan penerapan metode *decision tree* pada data mahasiswa FTI UMBY Angkatan 2012-2015 dapat diperoleh kesimpulan bahwa :

1. Dari hasil uji coba menggunakan 263 data sampel didapatkan akurasi menunjukkan kinerja algoritma mengklasifikasikan data sebanyak 215 data dengan presentase sebesar (82,8897%) serta data tidak berhasil di klasifikasikan dengan benar (*error*) sebanyak 48 data dengan presentase sebesar (17,1103%).
2. Pada hasil penerapan Algoritma C4.5 pada 263 data sampel Mahasiswa FTI UMBY Angkatan 2012-2015 dengan atribut sks tempuh kumulatif semester 7, prodi, ips semester 5-ips semester 7 menghasilkan suatu *rule* dimana mahasiswa yang lulus tepat waktu ialah mahasiswa pada saat semester 7 memiliki IPS dengan indeks A (3,51-4,00) dan indeks B (3,00-3,50) serta telah menempuh SKS 101-150 SKS pada semester 7.

## Daftar Pustaka

## 5. Bibliography

- [1] Menteri Pendidikan Dan Kebudayaan, "Peraturan," Februari 2020. [Online]. Available: [Http://Lldikti3.Ristekdikti.Go.Id/V6/Wp-Content/Uploads/2020/02/Salinan-PERMENDIKBUD-3-TAHUN-2020-FIX-GAB.Pdf](http://Lldikti3.Ristekdikti.Go.Id/V6/Wp-Content/Uploads/2020/02/Salinan-PERMENDIKBUD-3-TAHUN-2020-FIX-GAB.Pdf). [Diakses 26 July 2020].
- [2] I. P. Astuti, "Fountain Of Informatics Journal," *Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5*, 2017.
- [3] J. H. Jaman, *Prediksi Kelulusan Mahasiswa Dengan Metode Algoritma C4.5*, 2018.
- [4] N. D. Prayoga, *Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : Smik Royal Kisaran)*, 2018.
- [5] R. D. Pambudi, A. A. Suprianto Dan N. Y. Setiawan, "Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer," *Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya*, 2019.
- [6] E. P. Rohmawan, *Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network*, 2018.
- [7] D. Larose, *Discovering Knowledge In Data : An Introduction To Data Mining*, John Willey & Sons. Inc., 2005.
- [8] M. J. Berry Dan G. Linof, *Data Mining Techniques For Marketing, Sales, Customer, Relationship Management*. Second Edition., Wiley Publishing, Inc., 2004.
- [9] M. R. Romadhon Dan F. Kurniawan, "A Comparison Of Naive Bayes Methods, Logistic Regression And Knn For Predicting Healing Of Covid-19 Patients In Indonesia," Dalam *3rd East Indonesia Conference On Computer And Information Technology (Eiconcit)*, Bali, 2021.
- [10] A. Basuki Dan I. Syarif, *Modul Ajar Decision Tree*, Surabaya: Pens-Its, 2003.