

# IMPLEMENTASI ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN TEPAT WAKTU (Studi Kasus: Program Studi Teknik Informatika)

Sarah Novia Hermawanti<sup>1</sup>, Asriyanik,<sup>2</sup> Asril Adi Sunarto<sup>3</sup>

<sup>123</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi

sarahnovia050@gmail.com

asriyanik@gmail.com

## ABSTRAK

*The Informatics Engineering Study Program is one of the Study Programs at the University of Muhammadiyah Sukabumi which every year the incoming student quota not all students can graduate on time in accordance with the study period taken so as to result in the accumulation of a number of old students who do not graduate according to the graduation period. PHP is a collection of program codes that are added to HTML which will later be processed in the processor and the results will be displayed in the browser as HTML pages. C4.5 algorithm is the result of the development of the ID3 (Iterative Dichotomiser) algorithm which was developed by Quinlan. This algorithm is used to build a decision tree that is easy to understand, flexible, and interesting because it can be visualized in the form of images. Based on this research was conducted to create a system using classification techniques that can process large amounts of data to find patterns that occur in student data. Data processing is used to predict classes that are not yet known, namely in this design prediction of student graduation. The classification technique used is decision tree with the application of C4.5 algorithm. The input used is in the form of attributes from student data including IP (Performance Index) per semester from semester 1 to 7, and BTQ grades in semester 6. The student data is training sample data used in the preparation of decision trees. This test uses training data of students who have graduated from 2015 to 2018. This knowledge can be utilized by the UMMI Informatics Engineering Study Program as a preventive measure to avoid decreasing student graduation each year.*

**Keywords :** C4.5 Algorithm, Decision Tree, PHP, Timely Graduation Prediction.

Program Studi Teknik Informatika merupakan salah satu Program Studi di Universitas Muhammadiyah Sukabumi yang setiap tahun kuota mahasiswa yang masuk tidak semua mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang ditempuh sehingga mengakibatkan penumpukan jumlah mahasiswa lama yang tidak lulus sesuai dengan masa kelulusannya. PHP adalah kumpulan dari kode-kode program yang di tambahkan ke HTML yang nantinya akan diproses didalam *processor* dan hasilnya akan ditampilkan pada *browser* sebagai halaman HTML biasa. Algoritma C4.5 adalah hasil dari pengembangan algoritma ID3 (*Iterative Dichotomiser*) yang dikembangkan oleh Quinlan. Algoritma ini digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar. Berdasarkan hal tersebut penelitian dilakukan untuk membuat sistem menggunakan teknik klasifikasi yang dapat mengolah data dalam jumlah besar untuk menemukan pola yang terjadi pada data mahasiswa. Pengolahan data tersebut digunakan untuk memprediksi kelas yang belum diketahui yaitu dalam perancangan ini prediksi kelulusan mahasiswa. Teknik klasifikasi yang digunakan yaitu *decision tree* dengan penerapan algoritma C4.5. Inputan yang digunakan yaitu berupa atribut dari data mahasiswa meliputi IP (Indeks Prestasi) per semester dari semester 1 sampai 7, dan nilai BTQ di semester 6. Data mahasiswa tersebut merupakan data contoh training yang digunakan dalam penyusunan *decision tree*. Pengujian ini menggunakan data training mahasiswa yang sudah lulus pada tahun 2015 sampai 2018. Pengetahuan ini dapat dimanfaatkan oleh pihak Prodi Teknik Informatika UMMI sebagai langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

**Kata Kunci** – Algoritma C4.5, Decision Tree, PHP, Prediksi Kelulusan Tepat Waktu

## PENDAHULUAN

Lulus tepat waktu merupakan salah satu indikator keberhasilan mahasiswa dalam memperoleh gelar sarjana. Mahasiswa dikatakan lulus tepat waktu apalagi mampu menyelesaikan studinya di perguruan tinggi selama kurang dari atau sama dengan empat tahun, sedangkan mahasiswa dikatakan tidak lulus tepat waktu apabila menyelesaikan studinya lebih dari empat tahun. Dalam praktiknya hingga saat ini, mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana dalam kurun waktu empat tahun di perguruan tingginya. Mahasiswa yang telah menyelesaikan program studi sarjana selanjutnya mendaftarkan sebagai calon wisudawan untuk nantinya mengikuti proses wisuda.

Universitas Muhammadiyah Sukabumi adalah salah satu perguruan tinggi swasta di Sukabumi yang didirikan berdasarkan kajian dan studi kelayakan yang dilakukan pengurus Pimpinan Daerah Muhammadiyah (PDM) Sukabumi, pada tanggal 1 Desember 2000 berdasarkan surat ijin operasional Menteri Pendidikan Nasional dengan Nomor: 81/D/0/2003 pada tanggal 13 Juni 2003. Hingga tahun 2019, Universitas Muhammadiyah Sukabumi sudah memiliki 10 Program Studi, salah satu diantaranya adalah Program Studi (Prodi) Teknik Informatika. Prodi Teknik Informatika berdiri dari tahun 2003 dan sudah menghasilkan sebanyak 710 lulusan hingga tahun 2018. Banyaknya lulusan tersebut menjadikan sebuah data yang dapat diteliti untuk

menghasilkan rekomendasi suatu keputusan tertentu. Sehingga pada penelitian yang diteliti, penulis akan membuat sebuah aplikasi rekomendasi kelulusan tepat waktu bagi Program Studi. Dalam rentang waktu dari tahun 2014 hingga tahun 2017 sesuai dengan tahun yang difokuskan untuk diteliti, Prodi Teknik Informatika UMMI telah menghasilkan gudang data akademik dengan jumlah lulusan sebanyak 232 orang dan jumlah mahasiswa baru setiap tahunnya rata-rata 100 orang.

Secara statistik jumlah pendaftar mahasiswa Universitas Muhammadiyah Sukabumi dari tahun 2011 sampai dengan tahun 2014 di Program Studi Teknik Informatika adalah 558 pendaftar yang secara terperinci pada tahun 2011 sebanyak 101 pendaftar, tahun 2012 sebanyak 147 pendaftar, 2013 sebanyak 100 pendaftar, 2014 sebanyak 116 pendaftar.

Apabila kita lihat, jumlah pendaftar dari tahun ketahun semakin meningkat, untuk itu maka diperlukan pengklasifikasian data agar data tersebut mudah diolah yaitu menggunakan data mining sebagai pengklasifikasian data.

Salah satu fungsi dari data mining ialah Fungsi Klasifikasi (*classification*) dengan algoritma C4.5 dalam melakukan klasifikasi akan digunakan data latihan atau data pengalaman, dimana data di histori tersebut terdapat data latihan (*training data*) karena sesuai dengan pengertiannya yaitu akan menghasilkan pengetahuan. Teknik klasifikasi terdiri dari beberapa metode, dan

*decision tree* adalah bagian dari metode klasifikasi. Kemudian metode *decision tree* memiliki algoritma, algoritma C4.5 adalah salah satu dari algoritma yang memiliki decision tree. Untuk memprediksi kelulusan mahasiswa bisa tepat waktu atau tidak, maka perlu dipersiapkan sebuah data training, dimana data tersebut diambil dari data sampel kelulusan mahasiswa Teknik Informatika sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu. Dan pada penelitian ini mengambil sampel 100 mahasiswa lulusan dari tahun 2015-2018 dengan parameter IP Semester 1-7, dan Nilai BTQ di Semester 6.

Dari jumlah mahasiswa baru tersebut ternyata dalam kurung tahun 2015-2018 tidak semua mahasiswa Teknik Informatika menyelesaikan kuliah dengan tepat waktu. Hal ini tentunya mempengaruhi penilaian akreditasi sehingga perlu adanya ini untuk

mengetahui parameter apa saja yang mempengaruhi seorang mahasiswa dapat menyelesaikan studi tepat waktu.

Hal inilah yang mendasari peneliti untuk merancang sistem yang dapat melakukan prediksi kelulusan tepat waktu dengan menggunakan salah satu model data mining yaitu algoritma C4.5. Algoritma C4.5 memiliki akurasi yang baik dalam memprediksi kelulusan mahasiswa serta memiliki kelebihan yaitu dapat menampilkan sebuah pohon keputusan dalam bentuk *Rule* (aturan) yang berguna sebagai ukuran apakah model tersebut dapat memberikan informasi yang dibutuhkan oleh pihak Prodi Teknik Informatika UMMI dalam pengambilan sebuah keputusan. Sehingga penulis akan membuat judul penelitian “Prediksi Kelulusan Tepat Waktu dengan Algoritma C4.5”.

## PENELITIAN YANG TERKAIT

**Tabel 1.** Tabel Penelitian yang Terkait

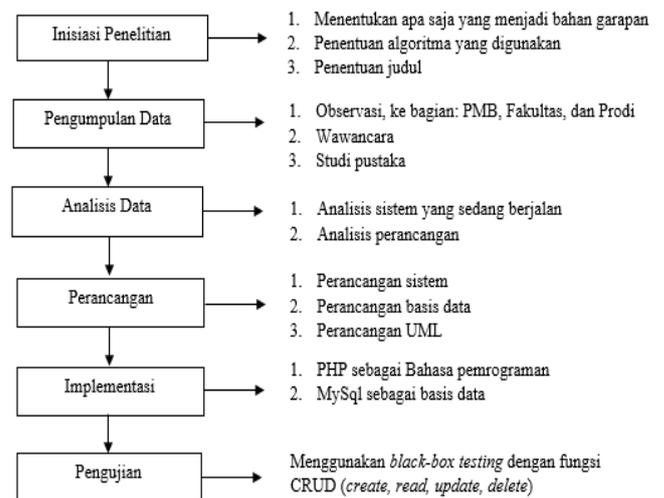
Judul Penelitian	Sumber Jurnal	Isi Ringkasan	Persamaan
Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, dan Indeks Prestasi Menggunakan Algoritma Decision Tree	Jurnal Teknologi Informasi, Volume 13 Nomor 1, Januari 2017, ISSN 1907-3380	Dari penelitian ini ditemukan bahwa prediksi masa studi siswa dipengaruhi oleh siswa yang masuk usia, jenis kelamin, IPK semester 1 hingga 4 semester IPK dan yang paling berpengaruh adalah IPK semester 4 mahasiswa pascasarjana tepat waktu dengan nilai gain 0,340 dari semua atribut.	Kesamaan jurnal dengan penelitian yang akan dilaksanakan yaitu menggunakan IPK sebagai salah satu penilaian dalam memprediksi kelulusan mahasiswa dan sama-sama menggunakan data mining algoritma Decision Tree C4.5
Implementasi Data Mining	Jurnal Media	Pada penelitian dibahas dengan memantau hasil belajar di	Kesamaan jurnal dengan penelitian yang akan

untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5	Infotama Vol. 11 No. 2, September 2015, ISSN 1858-2680	universitas berupa nilai IPK dan Jumlah SKS yang belum akurat untuk menentukan seorang mahasiswa lulus tepat waktu atau tidak.	dilaksanakan yaitu memantau hasil belajar dari nilai IPK dan jumlah SKS untuk memprediksi kelulusan mahasiswa.
---	--	--	--

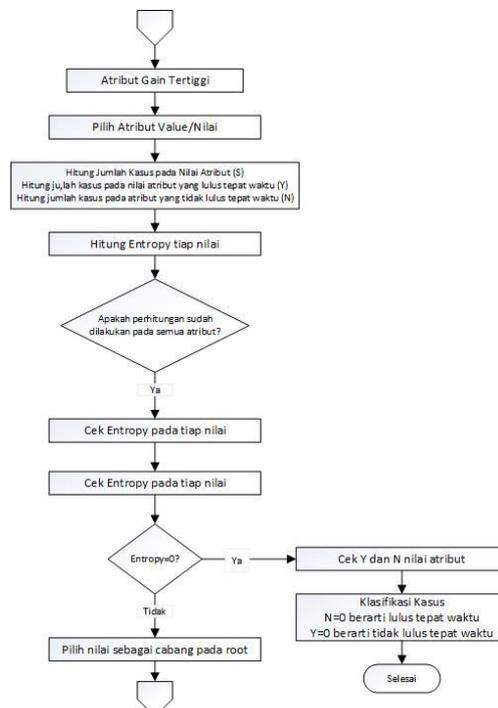
**METODE PENELITIAN**

Metode penelitian yang dilakukan oleh penulis ini yaitu menggunakan algoritma C4.5. Algoritma C4.5 ditemukan oleh John Ross Quinlan, merupakan pengembangan dari algoritma ID3. Berbeda dengan ID3 yang menggunakan *Information Gain*, algoritma C4.5 menggunakan *Gain Ratio* agar tidak bias dalam penentuan atribut pemilah terbaik (*the best split attribute*). Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.

Adapun tahapan penelitian yang dilakukan penulis yaitu menggunakan SDLC (*Software Development Life Cycle*), dengan susunan berikut:



**Gambar 1.** Tahapan Penelitian



**HASIL DAN PEMBAHASAN**

Input dari sistem yang akan dibangun oleh penulis yaitu berupa data mahasiswa lama dalam bentuk tabel yang berisi beberapa

atribut meliputi IP semester 1-7 dan nilai BTQ disemester 6. Data mahasiswa tersebut akan diklasifikasikan berdasarkan target yang akan ditentukan dan dihitung menggunakan metode *Decission Tree* yaitu algoritma C4.5 untuk mencari nilai *Entropy*, *Information Gain*, *Gain Ratio*, dan *Split Information*. Setelah proses perhitungan selesai maka akan menghasilkan *rule* atau kondisi yang digunakan dalam penentuan keputusan pada proses prediksi. *Output* dari sistem ini yaitu keterangan lulus tepat waktu atau tidaknya tiap mahasiswa yang diprediksi.

Sistem yang akan dibangun memiliki kemampuan sebagai berikut:

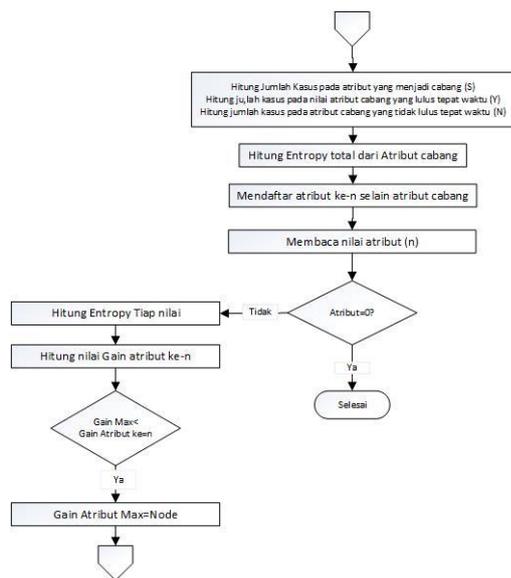
1. Melakukan perhitungan terhadap data yang telah diperoleh dengan mengubah formatnya ke dalam bentuk tabel dan sampel training
2. Melakukan perhitungan data untuk menentulan nilai *Gain* dan *Entropy*, *Gain Ratio*, dan *Split Information*.
3. Melakukan pembaruan dengan mentraining kembali sampel data yang baru
4. Memberikan pengetahuan baru dalam menentukan target yang akan diprediksi, yaitu tepat atau tidaknya waktu kelulusan

mahasiswa dengan masa studi yang ditempuh.

**Gambar 2.** *Flowchart* penentuan akar

Berikut sistem yang berjalan pada saat penentuan akar (*root*):

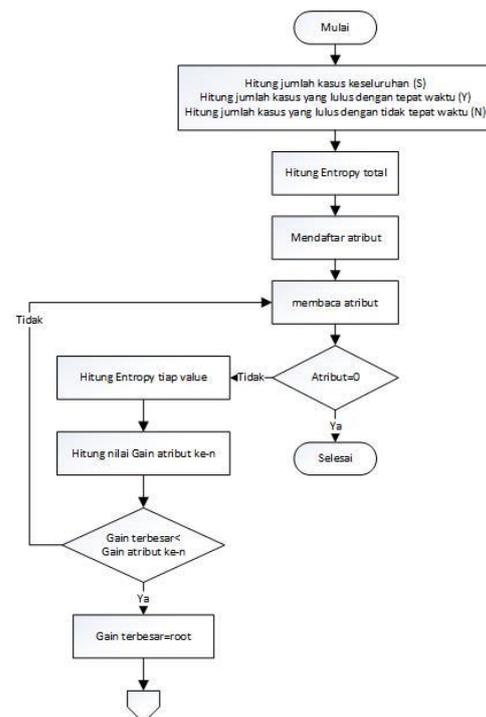
1. Sistem menghitung total nilai informasi dari data trainingnya
2. Sistem mendaftarkan atribut A
3. Masing-masing atribut akan dihitung nilai *Entropy* dan *Gain*nya
4. Sistem membandingkan nilai *Gain* terbesar dari masing-masing atribut
5. Setelah nilai *Gain* terbesar didapat, maka sistem akan memilih atribut dengan *Gain* terbesar sebagai atribut untuk dijadikan akar
6. Sistem melakukan proses ini hingga semua atribut dalam daftar selesai dihitung seluruhnya



**Gambar 3.** Flowchart Penentuan Cabang

Berikut sistem yang akan berjalan pada saat penentuan cabang pada *Tree* yang terbentuk:

1. Sistem akan menentukan atribut dengan *Gain* tertinggi
2. Nilai pada atribut tertinggi akan diklasifikasikan berdasarkan target yang ingin dicapai yaitu lulus tepat waktu (Y) atau tidak lulus tepat waktu (N)
3. Tiap nilai atribut akan dihitung *entropy* masing-masing hingga semua atribut selesai dihitung
4. Nilai *entropy* yang nol (0) akan dikoreksi untuk penentuan klasifikasi kasus.
5. Apabila nilai *entropy* lebih dari nol (0) maka nilai tersebut dijadikan cabang pada node berikutnya.



**Gambar 4.** Flowchart penentuan node

Berikut sistem yang akan berjalan saat penentuan node pada pembentukan *tree*:

1. Sistem menghitung jumlah kasus pada atribut dengan *Gain* tertinggi
2. Menghitung nilai *Entropy* total dari atribut dengan *Gain* tertinggi
3. Mendaftar atribut a selain atribut dari *gain* tertinggi
4. Menghitung *entropy* dari tiap nilai atribut ke-n hingga habis terhitung semua
5. Menghitung nilai *Gain* dari tiap atribut
6. Menentukan *Gain* maksimal untuk menentukan node berikutnya

Setelah node berikutnya terpilih maka proses perhitungan akan berulang lagi mulai dari penentuan cabang dan penentuan node hingga semua atribut habis dieksekusi dan mencapai *end of Tree*.

Pada langkah awal pada usulan sistem akan ditentukan beberapa atribut yang digunakan sebagai parameter dalam pengklasifikasian data sampel. Atribut

menyatakan suatu parameter yang dibuat sehingga kriteria dalam pembentukan pohon.

Berikut langkah-langkah perhitungannya :

1. Data dikelompokkan berdasarkan atribut berserta nilai didalamnya
2. Menghitung jumlah data pada tiap nilai atribut yang ada
3. Mengklasifikasi data yang sudah dihitung menjadi dua kelompok berdasarkan target tujuan yaitu lulus tepat waktu (Y) dan lulus tidak tepat waktu (N).
4. Menghitung *entropy* total dari 100 data mahasiswa
5. Menghitung *entropy* dari masing-masing nilai atribut
6. Menghitung *gain* dari tiap atribut
7. Mencari atribut dengan *gain* tertinggi untuk dijadikan *root*
8. Menentukan nilai atribut yang akan dijadikan cabang
9. Menentukan node selanjutnya dari atribut yang dipilih berdasarkan nilai *gain* tertinggi

Berikut rumus algoritma C4.5 yang digunakan pada pembentukan *decision tree* :

**a. Mencari nilai Entropy :**

Data dikelompokkan berdasarkan atribut dan nilai atributnya lalu dihitung jumlah keseluruhan, jumlah mahasiswa yang lulus tepat waktu dan tidak, lalu dihitung nilai *entropy* dan *gain* masing-masing atributnya.

Baris total kolom *entropy* pada tabel diatas dihitung dengan rumus sebagai berikut:

$$Entropy (S) = \sum_i^c - p_i * \log_2 p_i$$

Ket : S: himpunan kasus

N: jumlah partisi S

Pi: proporsi dari Si terhadap S

Entropy (Total) =

$$- \frac{75}{100} \log_2 \frac{75}{100} - \frac{25}{100} \log_2 \frac{25}{100}$$

Entropy (Total) = 0,8113

Jadi, dari data yang telah diambil memiliki nilai *entropy* sebesar 0,8113.

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

Entropy (IP1,A) =

$$- \frac{18}{18} \log_2 \frac{18}{18} - \frac{0}{18} \log_2 \frac{0}{18}$$

= galat matematika / 0

Entropy (IP1,B) =

$$- \frac{51}{68} \log_2 \frac{51}{68} - \frac{17}{68} \log_2 \frac{17}{68}$$

= 0,8113

Entropy (IP1,C) =

$$- \frac{6}{13} \log_2 \frac{6}{13} - \frac{7}{13} \log_2 \frac{7}{13}$$

= 0,9957

Entropy (IP1,D) =  $- \frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}$

= Galat Matematika => 0

Jadi, dari data di atas dapat diambil kesimpulan bahwa nilai *entropy* dari IP Semester 1 yang bernilai atribut A menghasilkan galat matematika, nilai atribut B menghasilkan 0,8113, nilai atribut C 0,9957, dan atribut D bernilai galat matematika.

**b. Rumus mencari nilai Gain**

Sementara itu nilai *gain* pada baris IP semester 1 dihitung dengan menggunakan rumus *gain*, sebagai berikut :

$$Gain(S,A) = Entropy(S) - \sum_{I=1}^C \frac{|Si|}{|S|} * Entropy(Si)$$

Dengan :

S : Himpunan Kasus

A : Atribut

N : Jumlah Partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Gain(Total, IP1) = Entropy(Total)$$

$$- \sum_{i=1}^n \left( \frac{IP1}{Total} \right) * Entropy(IP1)$$

$$Gain(Total, IP1) = 0,8113$$

$$- \left( \left( \frac{18}{100} \right) * 0 + \left( \frac{68}{100} \right) * 0,8113 + \left( \frac{13}{100} \right) * 0,9557 + - \left( \frac{1}{100} \right) * 0 \right)$$

$$= 0,1302$$

Jadi, nilai *gain* dari data di atas berjumlah 0,1302.

**c. Rumus mencari nilai Split Information**

$$Split\ Information(S,A) \equiv \sum_{I=1}^C$$

$$\frac{|Si|}{|S|} + \log_2 \frac{|Si|}{|S|}$$

$$Split\ Information =$$

$$- \left( \left( \frac{18}{100} \right) \log_2 \left( \frac{18}{100} \right) - \left( \left( \frac{68}{100} \right) \log_2 \left( \frac{68}{100} \right) - \left( \left( \frac{13}{100} \right) \log_2 \left( \frac{13}{100} \right) - \left( \left( \frac{1}{100} \right) \log_2 \left( \frac{1}{100} \right) \right) \right) \right)$$

$$= 1,302$$

Jadi, nilai *Split Information* dari data diatas berjumlah 1,302.

**d. Rumus mencari nilai Gain Ratio**

$$GainRatio \equiv \frac{Gain(S,A)}{SplitInformation(S,A)}$$

$$GainRatio(Total, IP1) \equiv \frac{0,1302}{1,302}$$

$$\equiv 0,1$$

Jadi, nilai *gain ratio* dari data di atas bernilai 0,1.

Sekarang, lakukan diskritisasi data bernilai numerik pada Tabel 4.4 ke dalam dua nilai (*binary split*). Pertama, nilai-nilai numerik didalam suatu atribut diambil yang unik (duplikasi dihilangkan) lalu diurutkan dari yang kecil ke besar (*ascending*).

Selanjutya kita harus memilih nilai batas *v* yang menghasilkan partisi terbaik berdasarkan ukuran ketidak murnian *gain ratio*. Untuk atribut IP1, IP2, IP3, IP4, IP5, IP6, IP7, dan Nilai BTQ, kita mendapatkan nilai batas terbaik secara berturut-turut.

**Tabel 2.** Tabel Hasil *Entropy*, *Gain*, *Split Information*, dan *Gain Ratio*

Atribut	Nilai Atribut	S	Y	N	Entropy	Information Gain	Split Info	Gain Ratio
<b>Total</b>	<b>Total</b>	80	60	20	0.8113			
IP1								
	A	17	17	0	0	0.1304	1.302	0.1002
	B	53	38	15	0.8595	0.1304	1.302	0.1002
	C	9	4	5	0.9911	0.1304	1.302	0.1002
	D	1	1	0	0	0.1304	1.302	0.1002
IP2								
	A	11	0	11	0	0.993	1.3566	0.0732
	B	47	10	37	0.46767	0.993	1.3566	0.0732
	C	22	10	12	0.994	0.993	1.3566	0.0732
IP3								
	A	11	1	10	0.4395	0.2365	1.4876	0.159
	B	42	3	39	0.3712	0.2365	1.4876	0.159
	C	26	15	11	0.9829	0.2365	1.4876	0.159
	D	1	1	0	0	0.2365	1.4876	0.159
IP4								
	A	11	0	11	0	0.195	1.444	0.135
	B	49	8	41	0.6421	0.195	1.444	0.135
	C	18	10	8	0.9911	0.195	1.444	0.135
	D	2	2	0	0	0.195	1.444	0.135
IP5								
	A	6	0	6	0	0.2982	1.6167	0.1844
	B	40	3	37	0.3843	0.2982	1.6167	0.1844
	C	27	10	17	0.951	0.2982	1.6167	0.1844
	D	7	7	0	0	0.2982	1.6167	0.1844
IP6								
	A	4	0	4	0	0.2012	1.3183	0.1526
	B	53	7	46	0.5631	0.2012	1.3183	0.1526
	C	19	9	10	0.998	0.2012	1.3183	0.1526
	D	4	4	0	0	0.2012	1.3183	0.1526
IP7								
	A	7	0	7	0	0.2791	1.4997	0.1861
	B	46	4	42	0.4262	0.2791	1.4997	0.1861
	C	19	9	10	0.9986	0.2791	1.4997	0.1861
	D	4	4	0	0	0.2791	1.4997	0.1861
Nilai BTQ								
	A	19	3	16	0.6292	0.0358	1.6951	0.0211
	B	36	9	27	0.8113	0.0358	1.6951	0.0211
	C	21	6	15	0.8631	0.0358	1.6951	0.0211
	D	3	2	1	0.9183	0.0358	1.6951	0.0211

Dari hasil diatas menunjukkan bahwa atribut IP7 dengan *gain ratio* paling besar di antara ke 8 atribut yang ada, menjadi atribut pemilah terbaik (*the best split attribute*) untuk diletakkan sebagai *root*.

Pemanfaatan algoritma dalam prediksi kelulusan tepat waktu ini harus dilakukan terlebih dahulu menghitung dari data training untuk dijadikan pola prediksi kedepannya, dalam penelitian ini menggunakan data training 100 sampel data mahasiswa lama 2015-2018, seperti yang tertera pada gambar dibawah ini :

No	Nama	IP Semester 1	IP Semester 2	IP Semester 3	IP Semester 4	IP Semester 5	IP Semester 6	IP Semester 7	Mata Kuliah	Tahun Masuk	Tanggal Ujian	Tahun Lulus	Status	Uraian/Keperluan
1	1305121001 MUHAMMAD FAHRIZAL	B	B	B	B	B	B	B	C	2012	29.08.2016	2016	Ya	Data Training
2	1305121002 DEWI KURNIA APRILIAN	B	B	B	B	C	B	B	C	D	2012	19.08.2017	Ya	Data Training
3	1305121003 YOGA FERDINA	B	B	B	B	B	B	B	B	B	2012	18.08.2017	Ya	Data Training
4	1305121004 ADISETIA ADYANUS PRIMA	A	B	C	B	B	B	B	B	B	2012	23.08.2016	Ya	Data Training
5	1305121005 SONNY SUPRANCI	B	B	B	C	C	B	B	B	B	2012	29.08.2016	Ya	Data Training
6	1305121007 IRIE HARSHADI	B	C	C	B	C	B	C	B	B	2012	28.08.2016	Ya	Data Training
7	1305121008 KEADY ANU SANTOSA PUTRI DIVYANTI	B	B	B	B	C	B	B	C	2012	28.08.2016	Ya	Data Training	
8	1305121009 M. HOSBY	B	B	B	B	B	B	B	C	2012	22.08.2016	Ya	Data Training	
9	1305121011 BAKU LINDIANA	B	A	C	C	B	B	B	B	2012	28.08.2016	Ya	Data Training	
10	1305121032 MOCHAMAD SEP SETIABAN	B	B	B	A	B	B	B	C	2012	18.08.2016	Ya	Data Training	
11	1305121033 RHEZAM FIDYUS	B	B	C	C	C	B	C	C	2012	22.08.2016	Ya	Data Training	
12	1305121035 YOSHY ALPHANI MOCHAMAD	A	B	B	B	B	B	B	B	2012	18.08.2016	Ya	Data Training	
13	1305121036 DEDE WAHYU DAMALAN	B	B	C	B	C	B	B	C	2012	22.08.2016	Ya	Data Training	
14	1305121037 AGI HARINDO PUTRA	B	B	C	B	C	B	B	C	2012	18.08.2016	Ya	Data Training	
15	1305121038 END KUSUMA PUTRA	B	B	C	B	B	B	B	B	2012	13.08.2017	Ya	Data Training	
16	1305121039 ELKA SILVASTIN	A	B	B	B	B	B	A	C	2012	28.08.2016	Ya	Data Training	
17	1305121030 ANINDA PRYA PAKA	A	B	B	B	B	B	B	C	2012	18.08.2016	Ya	Data Training	
18	1305121021 DIMAS GIBRAN CAPOPIA	B	B	C	C	B	B	C	B	2012	28.08.2016	Ya	Data Training	

Gambar 5. Data Training Mahasiswa Lulusan 2015-2018

Dari data training yang telah di *inputkan* maka sistem akan menemukan pola parameter mana yang menjadi atribut pemilah terbaik, dalam kasus ini yaitu pada atribut Indeks Prestasi di semester 7.

Untuk proses mining C4.5 bisa dilihat pada gambar dibawah ini

No	Atribut	Nilai	Gain Ratio
1	IP_Sem_7	Total	0.8193
2	IP_Sem_7	IP_Sem_7 A	0.1304
3	IP_Sem_7	IP_Sem_7 B	0.1302
4	IP_Sem_7	IP_Sem_7 C	0.1302
5	IP_Sem_7	IP_Sem_7 D	0.1302
6	IP_Sem_7	IP_Sem_7 E	0.1302
7	IP_Sem_7	IP_Sem_7 F	0.1302
8	IP_Sem_7	IP_Sem_7 G	0.1302
9	IP_Sem_7	IP_Sem_7 H	0.1302
10	IP_Sem_7	IP_Sem_7 I	0.1302

Gambar 6. Proses Mining C4.5

Dan untuk kinerja C4.5 dapat dilihat pada gambar 7. dibawah ini

Attribute	Value	Performance
IP_Sem_7	10	0.8193
IP_Sem_7	4	0.1302
IP_Sem_7	3	0.1302
IP_Sem_7	2	0.1302
IP_Sem_7	1	0.1302
IP_Sem_7	5	0.1302
IP_Sem_7	6	0.1302
IP_Sem_7	7	0.1302
IP_Sem_7	8	0.1302
IP_Sem_7	9	0.1302
IP_Sem_7	10	0.1302

Gambar 7. Kinerja C4.5

Hasil dari pola yang telah didapat maka dapat digunakan untuk memprediksi kelulusan mahasiswa yang akan datang seperti yang tertera pada gambar berikut :

Attribute	Value	Decision
IP_Sem_7	10	Ya
IP_Sem_7	4	Tidak
IP_Sem_7	3	Tidak
IP_Sem_7	2	Tidak
IP_Sem_7	1	Tidak
IP_Sem_7	5	Tidak
IP_Sem_7	6	Tidak
IP_Sem_7	7	Tidak
IP_Sem_7	8	Tidak
IP_Sem_7	9	Tidak
IP_Sem_7	10	Tidak

Gambar 8. Penentu Keputusan

## KESIMPULAN

### Kesimpulan

Kesimpulan yang diperoleh dari hasil pengerjaan skripsi tentang prediksi kelulusan tepat waktu dengan algoritma C4.5 yaitu:

1. Dari hasil ujiacoba menggunakan 100 data sampel dari mahasiswa lulusan Teknik Informatika tahun 2015-2018, pola yang dibentuk mempunyai akurasi kecocokan sebesar 68.42 %, nilai presisi sebesar 42.86 %, dan penarikan sebesar 60%.
2. Semakin banyak data sampel yang digunakan maka semakin kuat pula kebenaran prediksinya, sehingga algoritma yang dibuat dapat digunakan untuk memprediksi kelas baru yang belum diketahui dalam hal ini yaitu memprediksi kelulusan mahasiswa yang lulus tepat waktu atau tidak tepat waktu.
3. Sistem memprediksi kelulusan mahasiswa ini telah berhasil dibuat menggunakan metode *Decision Tree* yaitu dengan algoritma C4.5.

### Saran

Dalam pembuatan prediksi kelulusan tepat waktu dengan algoritma C4.5 ini masih banyak hal yang yang dapat dikembangkan seperti:

1. Sistem ini belum bisa menampilkan pencarian otomatis data nim dan nama mahasiswa yang akan diprediksi
2. Dapat dilakukan perbandingan dengan algoritma lain yang bisa mendukung pengujian data yang ada sehingga bisa didapat tingkat akurasi yang lebih baik lagi kedepannya.

## UCAPAN TERIMA KASIH

Ucapan terimakasih penulis sampaikan kepada pihak yang membantu dan memberikan dukungan kepada Prodi Teknik Informatika Universitas Muhammadiyah Sukabumi yang telah mendukung penelitian ini.

## DAFTAR PUSTAKA

### *Journal Article*

- [1] Ahmad, H.N., Suhartono, V., & Dewi, I.N. (2017). Penentuan Tingkat Kelulusan Tepat Waktu Mahasiswa STMIK Subang Menggunakan Algoritma C4.5. *Jurnal Teknologi Informasi*, Volume 13 Nomor 1, ISSN: 1907-3380.
- [2] Astuti, I.P. (2017). Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5. *Fountain of Informatics Journal*, Volume 2 No. 2, ISSN: 2541-4313.
- [3] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 di Universitas Bengkulu. *Jurnal Media Informasi*, Vol. 11 No. 2, ISSN: 1858-2680.
- [4] Mustafa, M.S., & Simpen, W. (2014). Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining di Data Akademik Mahasiswa STMIK Dipanegara Makassar. *Citec Journal*, Vol. 1, No. 4, ISSN: 2354-5771.

- [5] Pambudi, H.R., Setiawan, B.D., & Indriati. (2018). Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, NO. 7, hlm. 2637-2643, e-ISSN: 2548-964X.
- [6] Putri, R.P.S., Waspada, I. (2018). Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. *Jurnal Ilmu Komputer dan Informatika*, Vol. 4 No. 1, ISSN: 2621-038X.
- [7] Raharjo, B. (2011). *Membuat Database Menggunakan MySql*. Bandung: Informatika.
- [8] Rismayanti. (2018). Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika di Fakultas Teknik dan Komputer Universitas Harapan Medan. *Jurnal Sistem Informasi*, Volume 02, Nomor 1, ISSN: 2579-5341.
- [9] Romadhona, A., Suprapedi, Hilmawan, H. (2017). Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, dan Indeks Prestasi Menggunakan Algoritma Decision Tree. *Jurnal Teknologi Informasi*, Volume 13 Nomor 1, ISSN 1907-3380.
- Monograph, edited book, book*
- [10] Suyanto. 2018. *Machine Learning Tingkat Dasar dan Lanjut* Bandung: Informatika.
- [11] Sukamto, R. A., & Shalahuddin, M. (2014). *Rekayasa Perangkat Lunak*. Bandung: Informatika.
- [12] Sidik, B. (2012). *Pemrograman Web dengan PHP*. Bandung: Informatika.
- [13] Kustiyaningsih, Y., & Anamisa, D. R. (2011). *Pemrograman Basis Data Berbasis Web Menggunakan PHP & MySQL*. Yogyakarta: Graha Ilmu.
- [14] Buana, I. K. (2014). *Jago Pemrograman PHP Untuk Pemula dan Orang Awam*. Jakarta Timur: Dunia Komputer.
- Theses, Dissertation*
- [15] Susi Mashlahah. 2013. Prediksi Kelulusan Mahasiswa Menggunakan metode Decision Tree dengan Penerapan Algoritma C4.5. Malang (ID): Universitas Islam Negeri Maulana Malik Ibrahim Malang.