

Klasifikasi Predikat Kelulusan Mahasiswa Menggunakan Algoritma C4.5

Gusmelia Testiana¹, *Catur Eri Gunawan², Vina Fitriyanti³

^{1,2,3} Universitas Islam Negeri Raden Fatah Palembang, Indonesia

Jl. Pangeran Ratu No.2085, 5 Ulu, Kecamatan Seberang Ulu I, Kota Palembang, Sumatera Selatan

Email: ¹gusmeliatestiana_uin@radenfatah.ac.id, ²caturerig@radenfatah.ac.id, ³1930803057@radenfatah.ac.id

ABSTRACT

Determination of student graduation predicate based on GPA and additional requirements of timely study period for honors predicate. Currently, there is no in-depth classification to identify graduation predicate, so understanding of the patterns that influence the results is still limited. Accumulation of student graduation data can be used to find new information. This study aims to produce a decision tree classification model using the C4.5 algorithm and evaluate its accuracy in classifying student graduation predicates at UIN Raden Fatah Palembang. The data division technique used is k-fold cross validation to divide the data into training and testing data. The k value used is k = 3 in the first data test, this is based on previous tests with several k values, where k = 3 produces higher accuracy than the others. The rules formed are 242 and the attribute that influences student graduation predicate is GPA. The accuracy of the application of the C4.5 algorithm in classifying student graduation predicates is 83.31% which is included in the Good Classification category.

Keywords: graduation predicate; data mining; classification; C4.5 algorithm

ABSTRAK

Penentuan predikat kelulusan mahasiswa berdasarkan IPK dan syarat tambahan masa studi tepat waktu untuk predikat pujian. Saat ini, belum ada klasifikasi mendalam untuk mengidentifikasi predikat kelulusan, sehingga pemahaman tentang pola yang mempengaruhi hasil masih terbatas. Akumulasi data kelulusan mahasiswa dapat digunakan untuk menemukan informasi baru. Penelitian ini bertujuan untuk menghasilkan model klasifikasi pohon keputusan menggunakan algoritma C4.5 dan mengevaluasi akurasi dalam klasifikasi predikat kelulusan mahasiswa UIN Raden Fatah Palembang. Adapun teknik pembagian data yang digunakan yaitu k-fold cross validation untuk membagi data menjadi data *training* dan *testing*. Nilai k yang digunakan adalah k=3 pada pengujian data pertama hal ini didasari oleh pengujian yang dilakukan sebelumnya dengan beberapa nilai k, dimana k=3 menghasilkan akurasi yang lebih tinggi dari yang lain. Adapun *rule* yang terbentuk sebanyak 242 dan atribut yang mempengaruhi predikat kelulusan mahasiswa adalah IPK. Akurasi dari penerapan algoritma C4.5 dalam mengklasifikasikan predikat kelulusan mahasiswa sebesar 83,31% yang termasuk dalam kategori *Good Classification*.

Kata kunci: predikat kelulusan; data mining; klasifikasi; algoritma C4.5

1. PENDAHULUAN

Pendidikan tinggi merupakan jenjang pendidikan yang ditempuh setelah menyelesaikan sekolah menengah yang diadakan oleh perguruan tinggi. Perguruan tinggi berkewajiban untuk mengontrol mahasiswanya untuk menghasilkan lulusan yang berkualitas. Indeks prestasi dan masa studi mahasiswa menghasilkan predikat kelulusan dapat digunakan untuk menilai kualitas lulusan (Suhada et al., 2021). Terdapat beberapa variabel yang dapat digunakan sebagai ukuran keberhasilan mahasiswa seperti masa studi dan IPK lulusan (Nasution et al., 2022).

Masa studi dan predikat kelulusan mahasiswa merupakan salah satu bentuk dari keberhasilan studi yang didapatkan. Penentuan predikat kelulusan didasari oleh IPK yang diperoleh serta syarat lain yang ditentukan oleh universitas masing-masing. Perguruan tinggi dapat mengupayakan agar mahasiswa dapat menyelesaikan studi tepat waktu dengan IPK yang baik (Saputra et al., 2023).

Universitas Islam Negeri Raden Fatah Palembang merupakan universitas dibawah naungan Kementerian Agama. Setiap tahunnya Universitas Islam

Negeri Raden Fatah Palembang meluluskan mahasiswanya. Sehingga kumpulan data kelulusan mahasiswa bertambah menghasilkan akumulasi data yang banyak. Selain itu, belum dilakukan klasifikasi predikat kelulusan mahasiswa. Selama ini, penentuan predikat kelulusan hanya didasari berdasarkan perolehan IPK mahasiswa serta untuk predikat pujian juga dilihat dari masa studi. Jadi, belum diketahui kelompok perolehan predikat kelulusan mahasiswa.

Dalam data kelulusan mahasiswa tentunya terdapat banyak atribut data profil mahasiswa seperti NIM, nama, jenis kelamin, provinsi, jenis sekolah, prodi, fakultas, IPK, angkatan, tanggal masuk, tanggal lulus, angkatan wisuda, dan tahun akademik. Atribut-atribut itu tentunya memberikan gambaran profil mahasiswa secara menyeluruh. Banyaknya atribut pada data kelulusan mahasiswa ini dapat digunakan untuk penggalian informasi untuk mengklasifikasikan predikat kelulusan mahasiswa. Namun, dalam proses penggalian informasi klasifikasi predikat kelulusan mahasiswa, tidak semua atribut tersebut digunakan. Hanya atribut yang relevan dan

signifikan dipilih berdasarkan kebutuhan penelitian.

Klasifikasi predikat kelulusan mahasiswa menjadi sangat penting karena selama ini hanya dilakukan penentuan saja berdasarkan perolehan IPK dan juga masa studi. Sehingga mahasiswa dengan perolehan predikat kelulusan belum terklasifikasi dan kumpulan data kelulusan mahasiswa dengan akumulasi data yang banyak, maka alternatif solusi yang dapat dilakukan yaitu dengan menggunakan data kelulusan mahasiswa, untuk menggali informasi dan pengetahuan baru. Penggalan informasi dari gudang data dapat digunakan dalam membangun model baru dalam klasifikasi predikat kelulusan.

Pemanfaatan data kelulusan mahasiswa secara maksimal dengan menggunakan olah *data mining* dapat memberikan informasi yang berguna bagi pihak perguruan tinggi. Salah satu teknik data mining yang dapat digunakan dalam penelitian ini adalah klasifikasi.

Teknik klasifikasi adalah metode dari data mining yang dimanfaatkan untuk mengklasifikasikan item data ke dalam label kelas yang sebelumnya telah ditentukan, membangun sebuah

model klasifikasi dari data *training* untuk memprediksi data baru (Firdaus et al., 2021). Terdapat banyak berbagai algoritma yang dapat digunakan pada klasifikasi seperti *decision Tree*, *K-Nearest Neighbor*, *Naive Bayes Classifier*, dan lainnya. Penelitian ini menggunakan algoritma C4.5 yang termasuk dalam *decision tree* untuk mengklasifikasikan predikat kelulusan mahasiswa.

Algoritma C4.5 merupakan algoritma pohon keputusan yang dikembangkan oleh Ross quinlan. Atribut dengan nilai gain tertinggi menjadi prioritas serta menjadi poros dari klasifikasi pohon keputusan. Dua prinsip kerja dari algoritma C4.5 yaitu membuat pohon keputusan serta *rule* atau aturan model (Sukma et al., 2019).

Tingkat akurasi dan pemodelan waktu pada algoritma C4.5 lebih baik dari algoritma lainnya dan mempunyai tingkat kesalahan lebih rendah (Suweleh et al., 2020). Selain itu, algoritma ini dapat mengatasi dataset kompleks serta mengidentifikasi fitur-fitur penting data (Zafitri & Jambak, 2023).

Pembagian data dan evaluasi yang digunakan dalam penelitian ini adalah menggunakan *k-fold cross validation*

untuk mengetahui akurasi dari model yang terbentuk.

Terdapat penelitian terdahulu terkait dengan penerapan algoritma C4.5, yaitu penelitian oleh Rohman & Mujiyono (2021) melakukan prediksi predikat kelulusan mahasiswa menggunakan data mining. Penelitian tersebut bertujuan untuk mendapatkan pola yang terbentuk dari pohon keputusan yang dihasilkan menggunakan algoritma C4.5 yang mendapatkan nilai akurasi sebesar 71,67%. Penelitian oleh Saputra et al., (2023) melakukan prediksi predikat kelulusan mahasiswa dengan akurasi yang diperoleh sebesar 75% dan variabel yang berpengaruh terhadap predikat kelulusan adalah ips4, nkk, ipk dan pengalaman organisasi. Algoritma C4.5 Algoritma C4.5 ini mampu menangani atribut yang memiliki tipe diskrit dan numerik, tingkat akurasi tinggi, dan menghasilkan pohon keputusan yang mudah untuk diinterpretasi (Novika et al., 2021). Selain itu, dalam mengkonstruksi pohon keputusan algoritma ini membaca seluruh sampel data training (Zami et al., 2022).

Merujuk pada uraian konteks latar belakang yang telah dipaparkan di atas,

maka penelitian ini bertujuan untuk mengklasifikasikan predikat kelulusan mahasiswa untuk mengetahui kelompok mahasiswa yang memperoleh predikat kelulusan tertentu dengan menggunakan algoritma C4.5. Tujuan selanjutnya yaitu mengetahui akurasi dari algoritma C4.5.

2. METODE

Penelitian ini menerapkan metode kuantitatif berdasarkan pendekatan deskriptif. Penelitian deskriptif dengan pendekatan kuantitatif adalah penelitian yang bertujuan untuk mendeskripsikan suatu fenomena secara objektif dengan menggunakan data kuantitatif. Metode kuantitatif dalam penelitian ini memanfaatkan data mining dengan perhitungan algoritma C4.5.

2.1. Metode Pengumpulan Data

Untuk mendukung penelitian ini, maka dilakukan tahap pengumpulan data. Pengumpulan data dilakukan dengan teknik wawancara, studi pustaka dan dokumentasi. Data yang digunakan dalam penelitian adalah data sekunder. Penelitian ini menggunakan data kelulusan mahasiswa dari Pusat Teknologi Informasi dan Pangkalan

Data (PUSTIPD) UIN Raden Fatah Palembang.

Data yang digunakan adalah data kelulusan mahasiswa angkatan 2017 dan 2018 yang menempuh masa studi dari tahun 2020-2023. Pemilihan angkatan dan tahun lulus ini dikarenakan terdapat variasi waktu kelulusan mahasiswa, dimana ada mahasiswa yang masa studinya kurang dari delapan semester sementara yang lain lulus di semester selanjutnya. Jumlah sampel yang digunakan sebanyak 5.125 dari jumlah keseluruhan data kelulusan mahasiswa sebanyak 5.153. Pengambilan jumlah sampel ini berdasarkan pengolahan data mining yaitu *cleaning data* pada *data preparation*. Data yang tidak memiliki nilai yang kosong maka data tersebut dihapus. Sehingga diperoleh data sampel tersebut. Adapun jumlah *field* yang digunakan berdasarkan data kelulusan yaitu 8 *field*, dimana 7 *field* sebagai atribut dan 1 *field* sebagai label.

2.2. Algoritma C4.5

Decision tree adalah algoritma pengembangan dari algoritma sebelumnya yaitu ID3 (*Iterative Dichotomiser*) oleh J. Ross Quinlan di akhir tahun 1970-an dan awal 1980-an.

Quinlan mempresentasikan algoritma C4.5 (penerus ID3) (Han et al., 2012).

Algoritma C4.5 adalah struktur pohon keputusan yang memuat simpul yang menjelaskan mengenai atribut-atribut data, cabangnya mendeskripsikan hasil uji atribut, serta setiap daunnya mendeskripsikan kelas target. Setiap simpul keputusan dikunjungi oleh algoritma C4.5, memilih pembagian secara optimal hingga tidak dapat dibagi kembali. Untuk memilih pembagian yang optimal maka algoritma ini menggunakan konsep *entropy*, *information gain* dan *gain ratio* (Han et al., 2012).

Menurut Jollyta et al., (2020) tahapan dari proses algoritma C4.5 dapat dipaparkan sebagai berikut.

1. Persiapkan data *training*.
2. Menghitung nilai *entropy*.

Nilai Entropy dihitung dengan Persamaan 1.

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

S adalah himpunan kasus, N adalah jumlah partisi S , p_i adalah proporsi himpunan kasus ke-1 terhadap himpunan kasus.

3. Menghitung nilai gain seperti pada Persamaan 2.

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

$$Entropy(S_i) \quad (2)$$

S adalah himpunan kasus, A adalah atribut, n adalah jumlah partisi atribut A, |Si| adalah jumlah kasus pada partisi ke i dan |S| adalah jumlah kasus dalam S.

4. Menghitung nilai *split info*.

Nilai *split info* didapatkan dengan Persamaan 3.

$$SplitInfo(S,A) = - \sum_{i=1}^n \frac{S_i}{S} * \log_2 \frac{S_i}{S} \quad (3)$$

S adalah ruang sampel, A adalah atribut, dan Si adalah jumlah sampel atribut ke i.

5. Menentukan gain ratio dengan Persamaan 4.

$$GainRatio(S,A) = \frac{Gain(S,A)}{Split(S,A)} \quad (4)$$

Gain (S,A) adalah *information gain*, A adalah atribut, Split (S,A) adalah *split information* pada atribut (S,A).

6. Pengulangan pada proses ke 2 sampai semua cabang memiliki kelas yang sama.

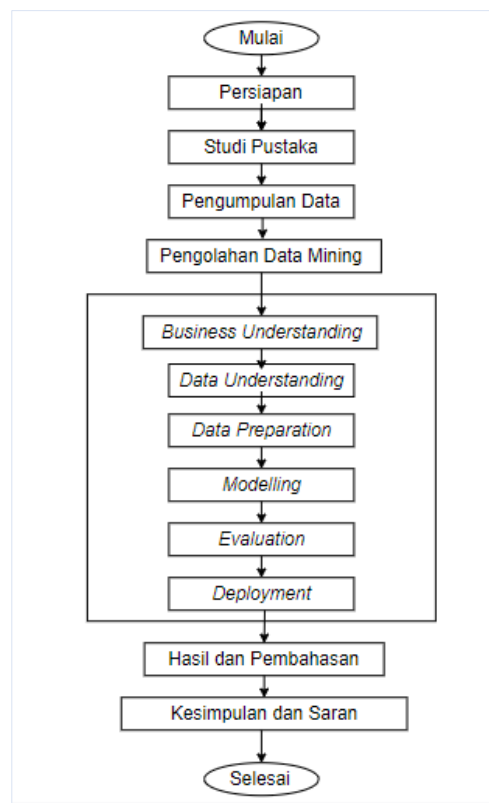
Percabangan pohon keputusan akan berhenti jika setiap kasus dari simpul n memiliki kelas yang sama, pada kasus tidak terdapat variabel independen yang akan dipartisi lagi,

serta tidak ada kasus di dalam cabang yang kosong.

Atribut dengan nilai gain ratio tertinggi akan dijadikan *root* dari pohon keputusan.

2.3. Tahapan Penelitian

Tahapan penelitian menggambarkan tentang tindakan yang dilakukan pada penelitian ini. Berikut Gambar 1 yang menjelaskan tahapan pada penelitian ini.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Pengolahan data mining adalah bagian utama dari hasil pada penelitian ini yang selanjutnya akan dibahas pada

bagian pembahasan. Pengolahan data mining dilakukan dengan menggunakan tahapan CRISP-DM yang meliputi enam tahapan yaitu *business/research understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, hingga *deployment*.

3.1. Research

Understanding(Pemahaman Penelitian)

Proses awal data mining diawali dengan melakukan pemahaman penelitian (*research understanding*) bertujuan untuk menetapkan tujuan yang akan dicapai melalui proses pengolahan data mining. Sebagaimana permasalahan yang telah diuraikan dalam latar belakang masalah, maka penerapan data mining dalam penelitian ini bertujuan mendapatkan hasil klasifikasi predikat kelulusan mahasiswa untuk mengetahui mahasiswa yang seperti apa yang masuk dalam masing-masing kelompok predikat kelulusan tersebut menggunakan algoritma C4.5. Perhitungan menggunakan algoritma C4.5 dilakukan dengan menentukan nilai *entropi* dari setiap atribut, *information gain*, *split info* dan *gain ratio*. Nilai *gain ratio* tertinggi dari

atribut akan dijadikan sebagai *root* dari pohon keputusan.

3.2. Data Understanding (Pemahaman Data)

Tahapan selanjutnya yaitu pemahaman data (*data understanding*). Tahap ini bertujuan untuk memahami karakteristik yang telah dikumpulkan. Adapun karakteristik data yang akan diproses pada penelitian ini adalah data kelulusan mahasiswa angkatan 2017-2018 yang menempuh masa studi dari tahun 2020-2023 Universitas Islam Negeri di Sumatera Selatan. Jumlah sampel data sebanyak 5125 data. Terdapat 13 atribut pada dataset awal yang diperoleh yaitu NIM, nama, jenis kelamin, provinsi, jenis sekolah, prodi, fakultas, IPK, angkatan, tanggal masuk, tanggal lulus, angkatan wisuda, dan tahun akademik. Berikut Tabel 1 merupakan dataset awal.

Tabel 1. Dataset Awal

No	NIM	Nama	Thn Akademik
1	169..41	A	... 20201
2	173..77	B	... 20201
3	173...97	C	... 20201
4	173...17	D	... 20201
5	173...35	E	... 20201
6	171...03	F	... 20201
7	171...17	G	... 20201
8	172...35	H	... 20201
9	172...40	I	... 20201
10	172...47	J	... 20201
...
5153	183...48	Z	... 20221

(Sumber: Pusat Teknologi Informasi dan Pangkalan Data)

Dari 13 atribut tersebut yang digunakan hanya 8 field, dimana 7 sebagai atribut dan 1 sebagai label. Berikut Tabel 2 atribut-atribut terpilih seperti jalur masuk, jenis kelamin, asal daerah, jenis sekolah, prodi, fakultas, IPK dan predikat kelulusan.

Tabel 2. Atribut Terpilih

Nama Field	Fungsi
Jalur Masuk	Atribut
Jenis Kelamin	Atribut
Asal Daerah	Atribut
Jenis Sekolah	Atribut
Prodi	Atribut
Fakultas	Atribut
IPK	Atribut
Predikat Kelulusan	Label

3.3. Data Preparation (Persiapan Data)

Persiapan data merupakan tahapan setelah melakukan pengumpulan data awal dan pemahaman data sebelumnya. Tahap ini akan mempersiapkan data sebelum data diolah dengan memanfaatkan teknik dari data mining. pada persiapan data akan dilakukan dengan proses seleksi data, *cleaning data*, *construct* dan transformasi data.

a. Seleksi Data

Dalam proses seleksi data maka dilakukan penentuan atribut yang paling

relevan untuk dianalisis yaitu dengan melakukan seleksi atribut. Atribut yang relevan dipilih berdasarkan literatur dari artikel, buku panduan akademik UIN Raden Fatah mengenai penentuan predikat kelulusan dan pertimbangan teoritis untuk memastikan pengaruhnya terhadap predikat kelulusan. Adapun atribut yang terpilih sebanyak 8 yang meliputi jalur masuk (diidentifikasi dari NIM), jenis kelamin, provinsi, jenis sekolah, prodi, fakultas, IPK dan predikat kelulusan sebagai target.

b. *Cleaning Data*

Cleaning data yaitu pembersihan data untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik. Melakukan pembersihan terhadap data yang *missing value* sebanyak 28 record dari jumlah dataset awal. Berhubung jumlah data yang mengalami *missing value* lebih sedikit jadi salah satu cara yang dapat digunakan untuk mengatasi data tersebut yaitu menghapus baris data yang mengandung *missing value* pada dataset (Rudy Hendrawan et al., 2022). Jumlah dataset setelah dilakukan pembersihan data sebanyak 5.125 data dari yang sebelumnya berjumlah 5.153.

c. *Construct Data*

Construct data adalah proses untuk mempersiapkan data konstruktif yaitu membangun atribut turunan dari atribut yang ada. Pada tahap ini ada atribut predikat kelulusan yang dijadikan *class target* diperoleh dari IPK mahasiswa serta masa studi untuk predikat pujian. Berikut Tabel 3 predikat kelulusan.

Tabel 3. Predikat Kelulusan

IPK	Predikat Kelulusan
3,51-4,00	Pujian
3,01-3,50	Sangat Memuaskan
2,76-3,00	Memuaskan
2,51-2,75	Baik
2,00-2,50	Cukup

d. *Transformasi Data*

Transformasi data adalah mengubah atribut tertentu dari tipe data numerik menjadi kategorik ataupun sebaliknya. Adapun atribut yang akan ditransformasi adalah jalur masuk, provinsi, dan IPK. Berikut Tabel 4 Transformasi Atribut.

Tabel 4. Transformasi Atribut

No	Atribut	Sub Atribut	Konversi
1	Jalur Masuk	1	SNMPTN/SPAN-PTKIN
		2	SBMPTN/UM-PTKIN
		3	Mandiri
		9	Internasional
2	Asal Daerah	Sumatera Selatan	Sumatera Selatan

Aceh,	Luar
Bangka	Sumatera
Belitung,	Selatan
Banten,	
Bengkulu,	
DKI	
Jakarta,	
Jambi,	
Jawa	
Barat,	
Jawa	
Tengah,	
Jawa	
Timur,	
Kepulauan	
Riau,	
Lampung,	
Riau,	
Sumatera	
Barat, dan	
Sumatera	
Utara	
Malaysia	Luar
	Negeri

3	IPK	3,51-4,00	0
		3,01-3,50	1
		2,76-3,00	2
		2,51-2,75	3
		2,00-2,50	4

Setelah dilakukan proses transformasi selesai, maka dataset siap untuk digunakan pada tahap berikutnya. Berikut Tabel 5 menunjukkan dataset kelulusan mahasiswa setelah dilakukan tahap preparation data mulai dari seleksi, *cleaning* data, *construct* dan transformasi data sesuai ketentuan di atas.

Tabel 5. Dataset Hasil *Preparation Data*

No	Jalur Masuk	Jenis Kelamin	Predikat Kelulusan
1	Internasional	L	Sangat Memuaskan
2	Mandiri	P	Pujian

3	Mandiri	P	.	Pujian
			.	
4	Mandiri	L	.	Pujian
			.	
5	Mandiri	P	.	Pujian
			.	
...
			.	
512	Mandiri	P	.	Sangat
5	i		.	Memuaskan
			.	

3.4. Modeling (Pemodelan)

Pada tahapan pemodelan ini adalah tahap menerapkan teknik dan algoritma data mining yang dipilih. Pada penelitian ini akan diterapkan klasifikasi dengan metode algoritma C4.5. Tools yang digunakan untuk perhitungan manual adalah *microsoft excel* sedangkan pada tahap pengujian menggunakan bantuan *software rapidminer*. Data yang digunakan adalah data kelulusan mahasiswa angkatan 2017 dan 2018 yang menempuh masa studi dari tahun 2020-2023. Data kelulusan ini akan dijadikan sampel untuk mengetahui klasifikasi predikat kelulusan dengan proses perhitungan algoritma C4.5.

Pada tahap ini dataset akan dibagi menjadi 2 yaitu data *training* dan data *testing*. Pembagian data berdasarkan nilai k yang digunakan pada saat evaluasi yaitu *k-fold Cross Validation*. Jumlah keseluruhan dataset sebanyak 5.125 data, lalu data akan diacak menggunakan fungsi *randomization* pada *Microsoft Excel* untuk memastikan bahwa setiap data memiliki peluang yang sama untuk dimasukkan ke dalam subset mana pun. Setelah data diacak, data tersebut akan dibagi menjadi tiga bagian yang sama besar, yaitu F1, F2, dan F3, sesuai dengan nilai k yang digunakan dalam *k-fold cross-validation*, yaitu $k=3$. Pengujian pertama, dengan data *training* F1 sebanyak 3.417 dan data *testing* F2 dan F3 sebanyak 1.708 data. Setelah pembagian data, selanjutnya yaitu perhitungan berdasarkan tahap algoritma C4.5 sebagai berikut:

a. Mempersiapkan *Data Training*

Dataset kelulusan mahasiswa diacak menggunakan fungsi *random* pada *excel*, lalu data tersebut dibagi menjadi 3 bagian sama besar yaitu F1, F2 dan F3. Data *training* yang digunakan berjumlah 3.417 data, dimana data dari F2 dan F3. Data ini mencakup urutan nomor 1709-5125,

yang merupakan nomor urut setelah data diacak, bukan berdasarkan nomor urut awal sebelum pengacakan.

b. Perhitungan Nilai Entropy

Perhitungan entropy berdasarkan rumus pada Persamaan 1. Berikut Tabel 6 nilai entropy total yang diperoleh.

Tabel 6. Perhitungan Entropy

Jml. Kasus	Pujian	Sangat Memuaskan	Memuaskan	ENTROPY
3417	1761	1654	2	1,005836353

c. Menghitung *Information Gain*

Mencari nilai *information gain* dengan rumus pada persamaan (2). Berikut Tabel 7 nilai *information Gain* pada setiap atribut.

Tabel 7. Perhitungan *Information Gain*

Atribut	<i>Information Gain</i>
Jalur Masuk	0,525802193
Jenis kelamin	0,013192245
Asal Daerah	0,036216145
Jenis Sekolah	0,382035952
Prodi	0,991999917
Fakultas	0,66972916
IPK	1,005836353

d. Menghitung *Split Info*

Menghitung *split info* menggunakan rumus pada persamaan (3). Berikut Tabel 8 Perhitungan *Split Info*.

Tabel 8. Perhitungan *Split Info*

Atribut	<i>Split Info</i>
Jalur Masuk	1,514637217
Jenis kelamin	0,822109744
Asal Daerah	0,229333942

Jenis Sekolah	2,205895228
Prodi	0,966461984
Fakultas	2,751623105
IPK	0,803355234

e. Menghitung *Gain Ratio*

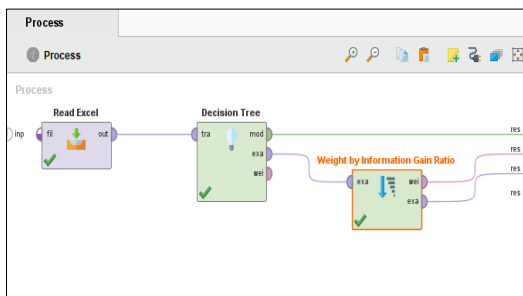
Setelah diperoleh *split info* selanjutnya yaitu mencari nilai *gain ratio* untuk menentukan akar dari pohon keputusan seperti pada Tabel 9.

Tabel 9. Perhitungan nilai *Gain Ratio*

Atribut	Nilai <i>Gain Ratio</i>
IPK	1,252044313
Prodi	1,026424147
Jalur Masuk	0,347147282
Fakultas	0,24339422
Jenis Sekolah	0,173188621
Asal Daerah	0,157918817
Jenis Kelamin	0,016046817

Berdasarkan nilai *gain ratio* yang telah diperoleh pada Tabel 9, maka *root* dari pohon keputusan adalah atribut IPK, karena IPK memiliki nilai *gain ratio* tertinggi.

Setelah dilakukan pengujian menggunakan microsoft excel selanjutnya melakukan pengujian menggunakan rapidminer. Berikut Gambar 2 hasil klasifikasi dari setiap kelompok predikat kelulusan mahasiswa.

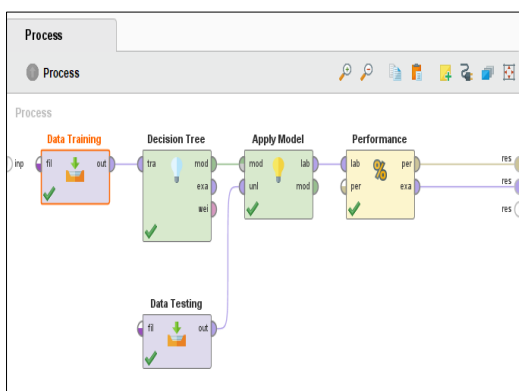


Gambar 2. Pemodelan Pada Rapidminer

Pada gambar 2 merupakan pemodelan menggunakan tools *Rapidminer* untuk menghasilkan pohon keputusan.

Berdasarkan pohon keputusan, maka diperoleh *rule* yang terbentuk sebanyak 242 *rule* dengan atribut yang memiliki gain tertinggi yaitu IPK. Sehingga atribut ini menjadi *root* dari pohon keputusan.

Selanjutnya yaitu melakukan pengujian terhadap data testing dari hasil pembagian data menggunakan *3-fold cross validation*. Adapun jumlah data testing yaitu 1.708 data. Berikut Gambar 3 pengujian *data testing*.



Gambar 3. Pengujian *Data Testing* Rapidminer

Dari pemodelan yang telah dilakukan di atas, maka diperoleh sebuah keputusan yaitu klasifikasi predikat kelulusan mahasiswa yang dapat dilihat pada Gambar 4 berikut.

predikat	predictionP_	confidence_	confidence_	confidence_	Jalur Masuk	Jenis Kelamin	Asal Daerah	Jenis Sekolah	Prodi	Fakultas
Sangat Mem.	Sangat Mem.	1	0	0	Internasional	L	Luar Negeri	SNK	SI Hukum K.	Fakultas Sa
Sajan	Pujian	0.333	0.667	0	Wanderi	P	Sumatera Sel.	SNK	SI Hukum K.	Fakultas Sa
Sajan	Pujian	0.333	0.667	0	Wanderi	P	Sumatera Sel.	SNK	SI Hukum K.	Fakultas Sa
Sajan	Sangat Mem.	0.529	0.471	0	Wanderi	L	Sumatera Sel.	SNK	SI Hukum K.	Fakultas Sa
Sajan	Pujian	0.333	0.667	0	Wanderi	P	Luar Sumatera.	SNK	SI Hukum K.	Fakultas Sa
Sajan	Pujian	0.184	0.816	0	SNMPNUSP.	L	Sumatera Sel.	SNK	SI Hukum PL.	Fakultas Sa
Sajan	Pujian	0.184	0.816	0	SNMPNUSP.	P	Luar Sumatera.	SNK	SI Hukum PL.	Fakultas Sa
Sajan	Sangat Mem.	0.400	0.400	0	SNMPNUSM.	L	Sumatera Sel.	SNK	SI Hukum PL.	Fakultas Sa
Sajan	Sangat Mem.	0.400	0.400	0	SNMPNUSM.	L	Sumatera Sel.	SNK	SI Hukum PL.	Fakultas Sa

Gambar 4. Hasil Klasifikasi Setiap Kelompok Predikat Kelulusan

Gambar 4 merupakan hasil dari klasifikasi setiap kelompok predikat kelulusan yang diuji menggunakan rapidminer. Pada hasil ini menunjukkan prediksi dari kelompok predikat kelulusan mahasiswa berdasarkan data *testing*, diperoleh sejumlah informasi yaitu 1258 mahasiswa diprediksi mendapatkan predikat pujian, sebanyak 449 mahasiswa diprediksi mendapatkan predikat sangat memuaskan dan 1 mahasiswa diprediksi mendapatkan predikat memuaskan.

3.5. *Evaluation* (Evaluasi)

Evaluasi memiliki tujuan mengetahui performa dari model pada tahap sebelumnya. Proses evaluasi yang akan digunakan pada model ini yaitu proses *k-fold cross validation* dengan

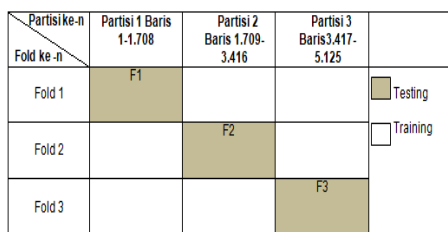
melihat tabel *confusion matrix* untuk mengetahui akurasinya.

Tahap penentuan nilai k yang digunakan ini sebelumnya telah melakukan pengujian terhadap beberapa nilai k yaitu 2, 3, 4, 5, 10 dan 15. Nilai k dengan akurasi tertinggi yang akan digunakan pada penelitian ini. Berikut Tabel 10 hasil akurasi nilai k yang diperoleh.

Tabel 10. Akurasi setiap Nilai k

K	Accuracy	weighted_mean_recall	weighted_mean_precision
k=15	79.88%	57.37%	60.29%
k=10	79.49%	60.55%	63.12%
k=5	79.61%	70.64%	73.20%
k=4	79.84%	75.79%	78.43%
k=3	79.94%	83.99%	87.24%
k=2	78.07%	66.61%	68.29%

Berdasarkan Tabel 10 di atas, maka nilai k dengan akurasi tertinggi yaitu k=3. Oleh karena itu, nilai k yang digunakan dalam penelitian ini adalah 3. Selanjutnya yaitu pembagian dataset menjadi 3 partisi sama besar.



Gambar 5. Pembagian Data Uji

Berdasarkan Gambar 5, maka dataset sebanyak 5.125 data dibagi menjadi 3 partisi yang samabesar yaitu 1.708 recordset yang nantinya akan

dijadikan sebagai data *testing* dan sisanya menjadi data *training*.

Setelah pembagian data, maka dilakukan klasifikasi menggunakan algoritma C4.5 dengan setiap nilai k dimasukkan. Sehingga diperoleh hasil berikut.

Tabel 11. Hasil Perhitungan k=3

Pengujian	Data Train	Data Test	Akurasi	Precision	Recall
1	F2,F3	F1	83.31%	81.66%	87.74%
2	F1,F3	F2	69.32%	82.70%	81.20%
3	F1,F2	F3	77.41%	85.07%	83.19%
Rata-rata (%)			76,68%	83,14%	84,04%

Berdasarkan Tabel 11. di atas, maka dapat dilihat bahwasanya akurasi yang paling tinggi terdapat pada pengujian yang ke satu dengan nilai sebesar 83,31% dengan model data *training* yang digunakan adalah F2 dan F3 serta data *testing* menggunakan F1. Maka model pada pengujian pertama merupakan model yang tepat digunakan karena memiliki akurasi akurasi tertinggi. Akurasi algoritma C4.5 dalam klasifikasi predikat kelulusan mahasiswa sebesar 83,31% termasuk ke dalam kategori *Good Classification*.

3.6. Deployment (Penyebaran)

Tahap ini merupakan tahap terakhir dari data mining. Adapun

pengetahuan dan pengujian yang diperoleh yaitu model klasifikasi predikat kelulusan mahasiswa. Model ini dapat digunakan untuk mengklasifikasikan predikat kelulusan mahasiswa serta dapat dijadikan rujukan untuk mengevaluasi dan memperbaiki pelaksanaan proses pembelajaran untuk menghasilkan luaran mahasiswa yang berkualitas.

Hasil evaluasi akan disajikan dalam bentuk tabel *confusion matrix* untuk mengetahui akurasi model yang dibangun. Akurasi dapat dikategorikan menjadi 5, yaitu sebagaimana pada Tabel 12 berikut.

Nilai Akurasi	Kategori
0,90-1,00	<i>Excellent Classification</i>
0,80-0,90	<i>Good Classification</i>
0,70-0,80	<i>Fair Classification</i>
0,60-0,70	<i>Poor Classification</i>
0,50-0,60	<i>Failure Classification</i>

Dalam penyebaran ini, dapat dipaparkan hasil akurasi model dengan C4.5 yaitu 83,31% yang termasuk dalam kategori “*Good Classification*”. Dari data uji yang berjumlah 1708 diperoleh 1199 mahasiswa yang diprediksi mendapat predikat pujian, 223 mahasiswa diprediksi mendapat predikat sangat memuaskan dan 1 mahasiswa diprediksi mendapat

predikat memuaskan. Ada 59 mahasiswa yang diprediksi mendapatkan predikat pujian tetapi masuk dalam kategori true sangat memuaskan dan 226 mahasiswa yang diprediksi mendapat predikat sangat memuaskan tetapi masuk kategori true pujian.

4. KESIMPULAN

Berdasarkan hasil dari serangkaian tahapan penelitian yang telah dilakukan, maka dapat diambil kesimpulan yaitu proses data mining dengan klasifikasi menggunakan algoritma C4.5 menghasilkan sebuah informasi model pohon keputusan dan menghasilkan *rule* untuk mengetahui predikat kelulusan mahasiswa UIN Raden Fatah Palembang. Perhitungan yang dilakukan menggunakan microsoft excel dan rapidminer keduanya menunjukkan bahwa tidak ada perbedaan signifikan. Hasil ini mengindikasikan bahwa metode perhitungan yang digunakan, baik secara manual maupun otomatis, menghasilkan akurasi yang sama dalam proses klasifikasi.

Adapun *rule* yang terbentuk sebanyak 242 *rule*, dimana atribut yang paling tinggi nilai gain rasionya yaitu

IPK mahasiswa. Atribut yang berpengaruh dapat digunakan untuk mengevaluasi dan memperbaiki pelaksanaan proses pembelajaran untuk membantu mahasiswa mendapatkan predikat kelulusan yang baik.

Akurasi yang diperoleh berdasarkan proses evaluasi menggunakan *k-fold cross validation* dengan nilai $k=3$ yaitu pada pengujian pertama yang disajikan dalam tabel *confusion matrix* yaitu 0,8331 atau 83,31%. Hasil akurasi tersebut menunjukkan bahwa model yang dibangun masuk dalam kategori “*Good Classification*”.

Penelitian ini memiliki kelemahan seperti jumlah atribut yang digunakan terbatas yaitu 8, sehingga hal ini dapat membatasi cakupan analisis. Lalu, penelitian ini hanya menerapkan satu algoritma tanpa mengeksplorasi atau membandingkan beberapa algoritma sehingga dapat memberikan hasil yang lebih beragam dalam membangun model klasifikasi.

Untuk mengatasi kelemahan yang telah diidentifikasi dalam penelitian ini, maka beberapa saran untuk penelitian selanjutnya yaitu menambahkan atribut-atribut lain yang lebih banyak dan bervariasi. Agar akurasi dan klasifikasi

predikat kelulusan mahasiswa mendapatkan kategori “*Excellent Classification*”. Lalu, agar lebih berkembang maka penelitian selanjutnya dapat melakukan perbandingan algoritma klasifikasi untuk predikat kelulusan mahasiswa. sehingga memberikan hasil beragam menemukan model terbaik dari berbagai algoritma.

DAFTAR PUSTAKA

- Firdaus, A. F., Saedudin, R., & Andeswari, R. (2021). Implementation of Naive Bayes Classification Method in Predicting Student Graduation. *E-Proceeding of Engineering*, 8(5), 9274–9279.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Data mining concepts and techniques. In *Elsevier* (Third Edit). <https://doi.org/10.1109/ICMIRA.2013.45>
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. DEEPUBLISH.
- Nasution, N. B., Hartanto, D., Silitonga, D. J., Lasimin, & Mardhiyana, D. (2022). Prediksi Lama Studi dan Predikat Kelulusan Mahasiswa Menggunakan Algoritma Supervised Learning. *G-Tech : Jurnal Teknologi Terapan*, 7(2), 386–395. <https://doi.org/10.33379/gtech.v7i2.2077>
- Novika, T., Okprana, H., Windarto, A. P., & Siahaan, H. (2021).

- Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika. *MEDIA INFORMATIKA BUDIDARMA*, 5(1), 9–17. <https://doi.org/10.30865/mib.v5i1.2498>
- Rohman, A., & Mujiyono, S. (2021). Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4.5. *Jurnal Prodi Teknik Informatika UNW "Multimatrix*, III(2), 1–5.
- Rudy Hendrawan, I. N., Budhi Saputra, I. M. A., Cahya Dewi, G. A. P., Adi Pranata, I. G. S., & Wedasari, N. L. N. (2022). Klasifikasi Lama Studi dan Predikat Kelulusan Mahasiswa menggunakan Metode Naïve Bayes. *Jurnal Eksplor Informatika*, 11(1), 50–56. <https://doi.org/10.30864/eksplor.a.v11i1.606>
- Saputra, A., Fitri, T. A., Karpen, & Susanti. (2023). Penerapan Data Mining Algoritma C4 . 5 Dalam Memprediksi Predikat Kelulusan Mahasiswa Di Politeknik Kampar. *Sains Dan Teknologi Informasi*.
- Suhada, K., Elanda, A., & Aziz, A. (2021). Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang). *Manajemen Dan Sistem Informasi*, 1(2), 14–27. <https://doi.org/10.35969/dirgamma.v1i2.182>
- Sukma, A. R., Halfis, R., & Hermawan, A. (2019). Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5. *Jurnal Teknik Komputer AMIK BSI*, 5(1), 21–28. <https://doi.org/10.31294/jtk.v4i2>
- Suweleh, A. S., Susilowati, D., & Hairani. (2020). Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5. *Jurnal BITE*, 2(1), 12–21. <https://doi.org/10.30812/bite.v2i1.798>
- Zafitri, Z., & Jambak, M. I. (2023). Karakteristik Pembatalan Reservasi Kamar Hotel Pada Online Travel Agent Menggunakan Algoritma C4.5. *Indonesian Journal of Computer Science*, 12(4), 2010–2023. <https://doi.org/10.33022/ijcs.v12i4.3268>
- Zami, A. Z., Nurdiawan, O., & Dwilestari, G. (2022). Klasifikasi Kondisi Gizi Bayi Bawah Lima Tahun Pada Posyandu Melati Dengan Menggunakan Algoritma Decision Tree. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(3), 305–310. <https://doi.org/10.30865/json.v3i3.3892>