

Penerapan Data Mining Algoritma C4.5 Dalam Memprediksi Predikat Kelulusan Mahasiswa Di Politeknik Kampar

Aldi Saputra¹, Triyani Arita Fitri², Karpen³, Susanti⁴

¹Stmik Amik Riau, 2110031802037@sar.ac.id, Pekanbaru, Indonesia

²Stmik Amik Riau, triyani@sar.ac.id, Pekanbaru, Indonesia

³Stmik Amik Riau, karpen@sar.ac.id, Pekanbaru, Indonesia

⁴Stmik Amik Riau, susanti@sar.ac.id, Pekanbaru, Indonesia

Informasi Makalah

Submit : May 30, 2023
Revisi : Juni 5, 2023
Diterima : Juni 16, 2023

Kata Kunci :

Data Mining
Prediksi
Algoritma C4.5
Predikat Kelulusan

Abstrak

Perguruan tinggi memiliki kewajiban dalam mengontrol mahasiswanya untuk menghasilkan lulusan yang berkualitas. Lulusan yang berkualitas dapat dilihat dari indeks prestasi dan lamanya studi yang menghasilkan predikat kelulusan. Salah satu teknik pembuatan prediksi yang dapat digunakan adalah teknik data mining. Prediksi merupakan teknik kombinasi dari beberapa teknik data mining lainnya. Penelitian ini bertujuan untuk memprediksi predikat kelulusan mahasiswa Politeknik Kampar dengan menggunakan algoritma C4.5, menganalisa tingkat akurasi dari algoritma C4.5 dan menentukan variabel yang berpengaruh terhadap prediksi predikat kelulusan mahasiswa. Data yang digunakan adalah data privat yang diperoleh dari Politeknik Kampar yang berisi data nilai akademik lulusan tahun 2015 – 2019 sebanyak 306 record dan data histori lulusan angkatan 2020 sebanyak 143 record . Hasil penelitian membuktikan bahwa Algoritma C4.5 berhasil diterapkan untuk memprediksi predikat kelulusan dengan nilai akurasi sebesar 75%, rata-rata presisi sebesar 74% dan rata-rata sensitivitas sebesar 67% dan variabel yang berpengaruh terhadap predikat kelulusan yaitu ips4, nkk, ipk dan pengalaman organisasi. Hasil penelitian ini dapat digunakan oleh Politeknik Kampar untuk mengevaluasi pola belajar mahasiswa agar nilai akademik mahasiswa dapat ditingkatkan.

Abstract

Universities have an obligation to control their students to produce quality graduates. Qualified graduates can be seen from the GPA and the length of study that results in the title of graduation. One technique for making predictions that can be used is data mining techniques. Prediction is a combination technique of several other data mining techniques. This study aims to predict the graduation predicate of Kampar Polytechnic students using the C4.5 algorithm, analyze the level of accuracy of the C4.5 algorithm and determine the variables that affect the prediction of the student graduation predicate. The data used is private data obtained from the Kampar Polytechnic which contains 306 records of academic grades for graduates from 2015 – 2019 and 143 records for historical graduates from class 2020. The results of the study prove that the C4.5 algorithm was successfully applied to predict the graduation predicate with an accuracy value of 75%, an average precision of 74% and an average sensitivity of 67% and the variables that affect the graduation predicate are ips4, nkk, ipk and organizational experience. The results of this study can be used by the Kampar Polytechnic to evaluate student learning patterns so that student academic scores can be improved.

1. Pendahuluan

Perguruan tinggi memiliki kewajiban dalam mengontrol mahasiswanya untuk menghasilkan lulusan yang berkualitas. Politeknik Kampar merupakan salah satu perguruan tinggi di bidang sawit yang ada di Indonesia (Fenty kurnia oktorina & Andri nofiar, 2021). Dalam hal ini, politeknik kampar memiliki tujuan dan target yang mana setiap lulusan dari perguruan tersebut menghasilkan lulusan terbaik dalam bidang industri yang dibuktikan dari nilai akademik yang di peroleh dari data tamatan lulusan yang sebelumnya (Aryadi, 2020). Lama studi mahasiswa dapat dipengaruhi oleh beberapa faktor dapat dilihat dari Indeks Prestasi Kumulatif (IPK) yang mencerminkan seluruh nilai yang diperoleh mahasiswa setiap semester yang sedang berjalan (Oon Wira Yuda et al., 2022). Lulusan yang berkualitas dapat dilihat dari indeks prestasi dan lamanya studi yang menghasilkan predikat kelulusan (Elanda & Suhada, 2021). Dari data tersebut menghasilkan sebuah data historis yang dapat dimanfaatkan untuk mengetahui tingkat pencapaian berupa sebuah predikat kelulusan yang di peroleh oleh mahasiswa dengan cara menggunakan algoritma dari data mining. Data mining merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenalan pola, teknik statistik, dan matematika (Ramadani & Alexander, 2020). Agar diketahui faktor lain yang memungkinkan mahasiswa tersebut mendapatkan predikat tersebut dengan melakukan sebuah analisis untuk memprediksi predikat kelulusan mahasiswa untuk memperoleh sebuah informasi atau pengetahuan baru. Pengetahuan baru yang dihasilkan pada penelitian ini yaitu berupa pengetahuan baru terkait faktor lain yang berpengaruh terhadap predikat kelulusan dan akurasi dari hasil prediksi yang diperoleh dalam implementasinya dengan menggunakan algoritma C4.5. Algoritma C4.5 merupakan

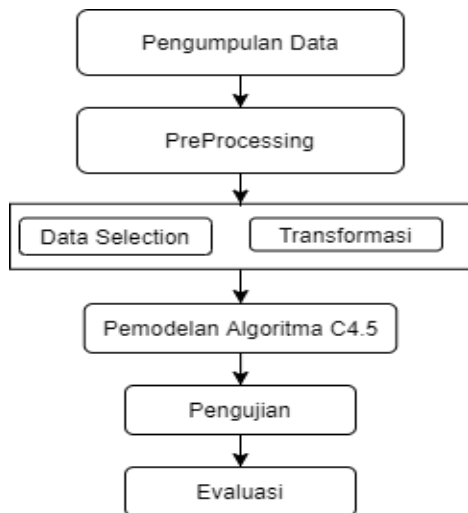
algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma lain (Jumaroh et al., 2020).

Ada beberapa peneliti sebelumnya yang telah melakukan penelitian dalam memprediksi predikat kelulusan mahasiswa di antara nya yaitu (Rohman & Mujiyono, 2021) menggunakan algoritma C4.5 yang menghasilkan atribut yang berpengaruh terhadap predikat kelulusan dan menghasilkan akurasi prediksi kelulusan sebesar 71,67%. Penelitian oleh Penelitian (Wibowo & Rohman, 2022) menggunakan algoritma C4.5 dan *Naïve bayaes* yang menghasilkan atribut yang berpengaruh terhadap predikat kelulusan dan menghasilkan akurasi prediksi kelulusan pada C4.5 sebesar 86% dan *Naïve Bayes* sebesar 60%.

Dalam proses penelitian ini dibutuhkan tahap *lifecycle* di antaranya tahap pengumpulan *dataset*, tahap *preprocessing*, pemodelan algoritma C4.5, pengujian dan evaluasi yang dilakukan menggunakan bahasa pemrograman *python* di *google collab*.

2. Metode Penelitian

Penelitian ini terdapat sebuah proses yang mencakup segala aspek kegiatan dalam mendapatkan hasil yang ingin di dapatkan. Berikut ini merupakan gambar alur dalam penelitian ini.



Gambar 1. Metodologi penelitian

2.1 Pengumpulan Data

Tahap pengumpulan data yaitu langkah awal dalam melakukan penelitian ini. Data yang dikumpulkan berupa *dataset* historis mahasiswa lulusan 2015-2019 mahasiswa politeknik Kampar yang bersifat *privat*. Data dalam bentuk *spreadsheet* di *excel* yang memiliki *variabel* sebanyak 19 *variabel*. Berikut ini merupakan *variable* data yang di peroleh :

Tabel 1. *Variabel* awal

No	Nama Variabel
1	Nama mahasiswa
2	Angkatan
3	Program studi
4	Status Hidup Ayah
5	Status Hidup Ibu
6	Pekerjaan Orang Tua
7	Asal Provinsi
8	Pengalaman Organisasi
9	Agama
10	IPK
11	Ips 1
12	Ips 2
13	Ips 3
14	Ips
15	NKK
16	Status Lulus Percobaan
17	Status Lulus Percobaan Lebih Dari Satu
18	Lulus Tepat Waktu
19	Predikat Kelulusan

Sumber : Politeknik Kampar

Setelah *variable* awal di peroleh, maka di perlukan data untuk mengetahui ketentuan

kriteria predikat yang akan dihasilkan untuk hasil penelitian yang dilakukan. Berdasarkan buku peraturan akademik di Politeknik Kampar, pada pasal 59 tentang status yudisium Diploma/Predikat Kelulusan di tentukan oleh unsur – unsur indeks Prestasi kumulatif (IPK), nilai kelakuan kumulatif (NKK) dan jumlah status percobaan. Berikut ini range status predikat kelulusan yang masih berlaku yaitu sebagai berikut yang dapat di lihat pada tabel dibawah ini.

Tabel 2. Kriteria Predikat Kelulusan

Predikat	Kriteria
Pujian	IPK \geq 3,50 NKK Program D3 \geq 22 NKK Program D4 \geq 30 Tidak pernah mendapat status kelulusan percobaan selama pendidikan. Lulus tepat waktu.
Sangat Memuaskan	IPK \geq 3,50 jika NKK tidak terpenuhi atau pernah menjadi lulusan percobaan. $3,01 \leq$ IPK < 3,50 NKK Program D3 \geq 20 NKK Program D4 \geq 28 Hanya pernah mendapat status kelulusan percobaan selama pendidikan sebanyak 1 Kali.
Memuaskan	IPK \geq 3,00 Jika NKK tidak terpenuhi atau pernah menjadi lulusan percobaan. $2,75 \leq$ IPK < 3,00 NKK Program D3 \geq 18 NKK Program D4 \geq 26 Hanya pernah mendapat status kelulusan percobaan selama pendidikan sebanyak 1 Kali.
Lulus	Kondisi lain yang tidak termasuk dalam dua kriteria sebelumnya

Sumber : Buku Peraturan Akademik Politeknik Kampar 2023

2.2 Preprocessing

Berdasarkan *dataset* yang diperoleh, terdapat data historis dari data lulusan sebelumnya. Sebelum proses data mining dilakukan, peneliti melakukan proses *preprocessing* data yaitu suatu tahapan dalam proses data mining. Sebelum *dataset* tersebut diproses agar menghasilkan *output* yang diharapkan. Tahap proses *preprocessing*

yang dilakukan di antaranya yaitu data *selection* dan *transformasi* data.

- A. Data *selection* merupakan tahap proses memilih data dari kumpulan yang dimiliki sebelum dilakukan tahap penggalian informasi dalam KDD dimulai (Taufiq, 2020). Pada tahap data *selection* dilakukan pemilihan *variabel* atau melakukan seleksi *variabel* apa saja yang akan nanti di lakukan pengujian menggunakan algoritma C4.5. Pada penelitian ini, variabel yang ada pada data awal tidak semuanya digunakan. Oleh karena itu, hanya variabel yang sesuai untuk dianalisis yang akan digunakan dari data awal yang diperoleh. Variabel yang digunakan pada penelitian ini sebanyak 15 kolom dan dapat dilihat pada tabel di bawah ini.

Tabel 3. *Variabel* yang digunakan

No	Variabel yang Di gunakan
1	Program Studi
2	Status Hidup Ayah
3	Status Hidup Ibu
4	Pekerjaan Orang Tua
5	Asal Provinsi
6	Pengalaman Organisasi
7	Agama
8	Angkatan
9	Ips 1
10	Ips 2
11	Ips 3
12	Ips 4
13	Nkk
14	Ipk
15	Predikat Kelulusan

- B. Transformasi data merupakan tahap mensinambungkan atribut/variabel yang nanti akan digunakan untuk memprediksi (Ginting et al., 2020). Tranformasi data pada penelitian ini yaitu mengubah data kategorik menjadi numerik yang bertujuan agar menghasilkan hasil dan performa yang baik. Berikut ini merupakan data yang dikukan transformasi yang dapat dilihat pada tabel dibawah ini.

Tabel 4. Transformasi nilai ips 1, ips 2, ips 3, ips 4 dan ipk

No	Nilai	Kategori
1	>3.51	A
2	>3.01	B
3	>2.75	C
4	<2.75	D

Dikarenakan Nkk yang digunakan menggunakan Nkk dari semester 4, maka dilakukan transformasi jumlah Nkk sebagai berikut.

Tabel 5. Transformasi jumlah Nkk

No	Jumlah	Kategori
1	>15	A
2	>13	B
3	<13	C

Tahap selanjutnya yaitu mengubah dari data bertipe kategorik menjadi numerik agar bisa di olah di pyton dengan menerapkan algoritma C4.5. Transformasi yang dilakukan dapat di lihat pada tabel dibawah ini.

Tabel 6. Transformasi data

No	Nama variabel	Hasil tranformasi
1	Program Studi	3,2,1,0
2	Angkatan	0,1,2,3,4
3	Status Hidup Ayah	0,1
4	Status Hidup Ibu	0,1
5	Pekerjaan Orang Tua	0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,6,17,18,19,20
6	Asal Provinsi	0,1,2,3,4,5,6,7,8,9,10,11,12
7	Pengalaman Organisasi	0,1
8	Agama	0,1,2,3,4
9	Ips 1	0,1,2,3,4
10	Ips 2	0,1,2,3,4
11	Ips 3	0,1,2,3,4
12	Ips 4	0,1,2,3,4
13	Nkk	0,1,2
14	Ipk	0,1,2

2.3 Pemodelan Algoritma C4.5

Algoritma C4.5 merupakan turunan dari ID3 yang menangani suatu atribut yang berbentuk numerik, missing value beserta noise pada sebuah dataset, dan rule-rule yang didapatkan dari terbentuknya sebuah model yang berbentuk pohon (Made et al., 2022). Algoritma ini dapat digunakan untuk memprediksi atau mengklarifikasi suatu kejadian dengan pembentukan *Decision Tree* atau pohon keputusan (Romli & Zy, 2020). Decision Tree atau pohon keputusan adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (leaf) merepresentasikan kelompok kelas tertentu (Farida et al., 2020).

Beberapa tahapan dalam algoritma C4.5 memilih atribut sebagai akar, berdasarkan nilai gain tertinggi dari atribut-atribut yang ada. Selanjutnya yaitu menentukan akar dari pohon. Akar tersebut diambil dari atribut, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain tertinggi akan dijadikan akar pertama. Sebelum menentukan nilai gain hitung terlebih dahulu nilai entropy menggunakan rumus sebagai berikut :

$$Entropy(S) = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Keterangan :

S = kumpulan data

k = banyaknya kelas dalam S

p_j = probabilitas kelas S

Setelah itu tentukan nilai gain, menggunakan rumus sebagai berikut :

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

Keterangan :

S = kumpulan data

A = atribut

A_i = nilai atribut ke-i

$|S_i|$ = jumlah data untuk A_i

$|S|$ = jumlah data dalam S

k = jumlah nilai atribut

Ulangi langkah ke-2 hingga kasus dalam cabang memiliki kelas yang sama. Ulangi langkah ke-3 hingga kasus dalam cabang tidak bisa dihitung lagi. Setelah semua proses perhitungan dilakukan, tahap selanjutnya yaitu membentuk pohon keputusan yang menghasilkan rule yang di tampilkan pada grafik *decision tree*.

2.4 Pengujian

Jika perhitungan model algoritma c4.5 telah dilakukan dan telah menghasilkan rule atau aturan yang telah terbentuk, maka di lakukan tahap pengujian dengan diimplementasikan menggunakan bahasa pemrograman *python*. *Python* merupakan bahasa pemrograman yang bebas digunakan. tidak ada larangan dalam penyalinannya atau mendistribusikannya (Medy & Alviva, 2019). Tujuan dari tahap pengujian ini yaitu untuk menghasilkan prediksi berdasarkan dari rule yang telah bentuk dari data training dan data testing yang dilakukan pada model c4.5.

2.5 Evaluasi

Setelah menghasilkan akurasi, maka di lakukan evaluasi model terhadap hasil prediksi yang ada. Hal tersebut dilakukan untuk mengetahui seberapa besar keakuratan berupa hasil akurasi dari hasil prediksi menggunakan algoritma c4.5. Proses evaluasi dilakukan dengan melihat, menganalisis dan membandingkan hasil pengujian dengan menggunakan *confusion matrix* (Agustian & Ramadhani, 2022). Confusion Matrix merupakan sebuah metode yang kegunaannya untuk melakukan perhitungan akurasi pada konsep dalam data mining, evaluasi dengan menggunakan metode confusion matrix menghasilkan nilai akurasi, presisi dan recall (Normawati & Prayogi, 2021). Confusion matrix memiliki tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Berikut ini merupakan tabel confusion matrixnya :

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Gambar 2. Gambar tabel confusion matrix

Keterangan :

TP (True Positive) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1

TN (True Negative) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (False Positive) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (False Negative) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus confusion matrix untuk menghitung accuracy, precision, dan recall adalah sebagai berikut :

$$accuracy = \frac{TP+TN}{Total}$$

$$precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang dikumpulkan berupa dataset historis mahasiswa lulusan 2015-2019 mahasiswa politeknik dalam bentuk spreadsheet di excel yang memiliki variabel sebanyak 19 variabel dengan teknik observasi dan wawancara secara langsung pada objek atau tempat studi kasus penelitian yang dilakukan yaitu di politeknik kampar. Data tersebut dapat dilihat pada gambar berikut :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Nama Mahasiswa	Program Studi	Status Hidup Ayah	Status Hidup Ibu	Pekerjaan Orang Tua	Asal Provinsi	Pengalaman Organisasi	Agama	Ips 1	Ips 2	Ips 3	Ips 4	Wkk	Ipk	Status Lulus Perbaikan (Lebih dari Satu)	Lulus Tepat Waktu	Lulus Peringkat Kelulusan	Predikat Kelulusan	
2	Fajri	TPS	2015 Hidup	Hidup	Montir	RAJU	ada	Islam	3,45	3,39	3,44	3,3	14	3,4	ya	tidak	ya	Sangat Memuaskan
3	Ridha	TPS	2015 Hidup	Hidup	buruh	RAJU	ada	Kristen	3,4	2,99	3,59	3,12	14	3,20	tidak	tidak	ya	Sangat Memuaskan
4	Ridha	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Kristen	3,65	3,93	3,98	3,58	15	3,75	ya	tidak	ya	Pujian
5	Lulus	TPS	2015 Hidup	Hidup	gejati	RAJU	tidak ada	Islam	3,5	3,46	3,45	3,42	14	3,45	ya	tidak	ya	Sangat Memuaskan
6	Lulus	TPS	2015 Hidup	Hidup	Montir	RAJU	tidak ada	Islam	3,23	3,56	3,68	3,44	14	3,40	tidak	tidak	ya	Sangat Memuaskan
7	Dewa	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,07	3,02	3,51	3	14	3,15	ya	tidak	ya	Sangat Memuaskan
8	Melita	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,23	2,92	3,32	2,95	14	3,11	tidak	tidak	ya	Sangat Memuaskan
9	Melita	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,05	3,12	3,33	3,07	14	3,14	ya	tidak	ya	Sangat Memuaskan
10	Siswanto	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,28	3,22	3,28	3,05	14	3,21	ya	tidak	ya	Sangat Memuaskan
11	Nani	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,07	3,12	3,38	3,52	14	3,27	tidak	tidak	ya	Sangat Memuaskan
12	Alham	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,05	2,71	2,51	2,9	12	2,89	ya	tidak	ya	Memuaskan
13	Yusuf	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,07	3,07	3,32	3	14	3,13	ya	tidak	ya	Sangat Memuaskan
14	Yusuf	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,35	3,39	3,89	3,65	15	3,57	ya	tidak	ya	Pujian
15	Yusuf	TPS	2015 Hidup	Hidup	gejati	SUMATRA	ada	Islam	3,42	3,49	3,42	3,54	14	3,47	tidak	tidak	ya	Sangat Memuaskan
16	Adnan	TPS	2015 Hidup	Hidup	gejati	RAJU	ada	Islam	3,1	3,19	3,51	3,37	14	3,29	ya	tidak	ya	Sangat Memuaskan
17	Febry	TPS	2015 Hidup	Hidup	BURUK	RAJU	ada	Islam	3,1	2,85	3,32	3	14	3,07	ya	tidak	ya	Sangat Memuaskan
18	Adnan	TPS	2015 Hidup	Hidup	Montir	RAJU	ada	Islam	3,15	2,85	3,06	2,9	12	2,97	ya	tidak	ya	Memuaskan
19	Siti	Ha	TPS	2015 Hidup	Hidup	WIRAUAS/SUMATRA	ada	Islam	3,35	3,42	3,88	3,65	15	3,63	ya	tidak	ya	Pujian
20	Mahar	TPS	2015 Hidup	Hidup	Karyawan	RAJU	tidak ada	Islam	2,75	2,76	2,72	2,82	12	2,76	tidak	tidak	ya	Memuaskan
21	Adel	TPS	2015 Hidup	Hidup	Karyawan	RAJU	tidak ada	Islam	2,72	2,84	2,91	3,12	12	2,85	ya	tidak	ya	Memuaskan
22	Dimas	TPS	2015 Hidup	Hidup	WIRAUAS/RAJU	tidak ada	Islam	2,65	2,54	2,81	3,12	12	2,78	ya	tidak	ya	Memuaskan	
23	Reddy	TPS	2015 Hidup	Hidup	buruh	RAJU	tidak ada	Islam	2,77	2,61	2,79	2,86	12	2,76	tidak	tidak	ya	Memuaskan

Gambar 3. Data awal

3.2. Preprocessing

Tahap pertama dalam melakukan preprocessing adalah menyeleksi data variabel yang digunakan menjadi 14 variabel dan 1 label yang di proses di google collab menggunakan python yang dapat dilihat pada gambar dibawah ini :

```
atribut = ["Program Studi", "Status Hidup Ayah", "Status Hidup Ibu", "Pekerjaan Orang Tua", "Asal Provinsi", "Pengalaman Organisasi", "Agama", "Angkatan", "Ips 1", "Ips 2", "Ips 3", "Ips 4", "Wkk", "Ipk"]
```

Gambar 4. Seleksi data

Selanjutnya yaitu tahap transformasi data yang telah di seleksi di lakukan perubahan dari data kategorik mejadi data numerik. Proses tersebut dapat dilihat pada gambar di bawah ini :

```
from sklearn.preprocessing import LabelEncoder
enc = LabelEncoder()
dataset['Program Studi'] = enc.fit_transform(dataset['Program Studi'].values)
dataset['Status Hidup Ayah'] = enc.fit_transform(dataset['Status Hidup Ayah'].values)
dataset['Status Hidup Ibu'] = enc.fit_transform(dataset['Status Hidup Ibu'].values)
dataset['Pekerjaan Orang Tua'] = enc.fit_transform(dataset['Pekerjaan Orang Tua'].values)
dataset['Asal Provinsi'] = enc.fit_transform(dataset['Asal Provinsi'].values)
dataset['Pengalaman Organisasi'] = enc.fit_transform(dataset['Pengalaman Organisasi'].values)
dataset['Agama'] = enc.fit_transform(dataset['Agama'].values)
dataset['Angkatan'] = enc.fit_transform(dataset['Angkatan'].values)
dataset['Ips 1'] = enc.fit_transform(dataset['Ips 1'].values)
dataset['Ips 2'] = enc.fit_transform(dataset['Ips 2'].values)
dataset['Ips 3'] = enc.fit_transform(dataset['Ips 3'].values)
dataset['Ips 4'] = enc.fit_transform(dataset['Ips 4'].values)
dataset['Wkk'] = enc.fit_transform(dataset['Wkk'].values)
dataset['Ipk'] = enc.fit_transform(dataset['Ipk'].values)

dataset
```

Gambar 5. Transformasi Data

Hasil dari data yang dilakukan pada tahap processing yang dapat dilihat pada gambar berikut ini :

Program Studi	Angkatan	Status Hidup Ayah	Status Hidup Ibu	Pekerjaan Orang Tua	Asal Provinsi	Pengalaman Organisasi	Agama	Ips 1	Ips 2	Ips 3	Ips 4	Mak	Tpk	Predikat Kelulusan
0	3	0	0	0	7	8	0	0	1	1	1	1	1	Sangat Memuaskan
1	3	0	0	0	0	8	0	1	1	2	0	1	1	Sangat Memuaskan
2	3	0	0	0	10	8	0	1	0	0	0	0	0	Pujian
3	3	0	0	0	10	8	1	0	1	1	1	1	1	Sangat Memuaskan
4	3	0	0	0	7	8	1	0	1	0	0	1	1	Sangat Memuaskan
...
301	0	3	0	0	10	8	1	0	1	1	1	1	1	Sangat Memuaskan
302	0	3	0	0	18	10	1	0	0	0	0	0	0	Pujian
303	0	3	0	0	10	10	1	0	0	0	0	1	0	Pujian

Gambar 6. Hasil data di processing

3.3. Pemodelan algoritma c4.5

Untuk memulai tahap pemodelan menggunakan algoritma c4.5, maka dilakukan split data untuk membagi data training dan testing dalam pemodelannya. Pada pemodelan algoritma c4.5 ini dilakukan dengan data split dengan perbandingan

sebesar 70:30 dari data yang yang dimiliki. Penerapan algoritma tersebut dapat dilihat pada gambar dibawah ini :

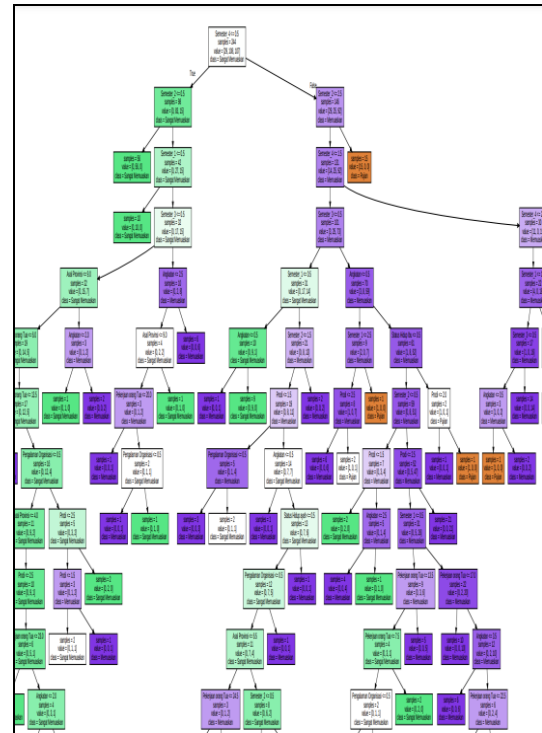
```
dataTraining, dataTesting, targetTraining, targetTesting = train_test_split(data, target, test_size=0.3)
```

Gambar 7. Split data training dan data testing

```
modelKlasifikasi = tree.DecisionTreeClassifier(criterion='entropy')
modelKlasifikasi = modelKlasifikasi.fit(dataTraining, targetTraining)
```

Gambar 8. Pemodelan decision tree

Hasil dari algoritma tersebut di tampilkan dalam bentuk grafik decision tree yang dapat dilihat pada gambar dibawah ini :



Gambar 9. Hasil pemodelan algoritma c4.5 membentuk decision tree

3.4. Pengujian

Selanjutnya pada penerapan algoritma menggunakan python, dilakukan pengujian dalam memprediksi dengan menginput kriteria data yang dimiliki dan menghasilkan label prediksi yang dapat dilihat pada gambar dibawah ini :

```
targetPrediksi = modelKlasifikasi.predict([[3, 0, 0, 10, 11, 0, 0, 0, 1, 2, 1, 1]])
```

Gambar 10. Pengujian hasil prediksi kelulusan predikat

Berikut hasil prediksi yang diperoleh :

```
print(targetPrediksi)
['Sangat Memuaskan']
```

Gambar 11. Hasil prediksi

3.5. Evaluasi

Selanjutnya dilakukan evaluasi terhadap hasil model dalam memprediksi untuk melihat ke akuratan hasil prediksi dengan

melihat hasil akurasi dan confusion matrix yang dapat dilihat pada gambar dibawah ini :

```
akurasi = accuracy_score(targetTesting, targetPrediksi)
```

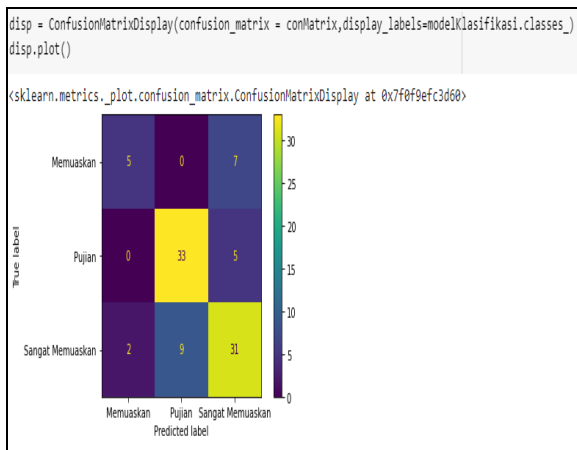
Gambar 12. Menghitung akurasi dari hasil prediksi

Berikut hasil akurasi prediksi :

	precision	recall	f1-score	support
Pujian	0.71	0.42	0.53	12
Sangat Memuaskan	0.79	0.87	0.82	38
Memuaskan	0.72	0.74	0.73	42
accuracy			0.75	92
macro avg	0.74	0.67	0.69	92
weighted avg	0.75	0.75	0.74	92

Gambar 13. Hasil akurasi prediksi

Berikut hasil akurasi menggunakan confusion matrix :



Gambar 14. Hasil evaluasi

4. Simpulan

4.1. Kesimpulan

Berdasarkan dari hasil penelitian yang di lakukan, maka dapat di simpulkan sebagai berikut :

- Analisa prediksi predikat kelulusan mahasiswa menggunakan algorithma C4.5 telah berhasil dilakukan di Politeknik Kampar terhadap data uji

lulusan tahun 2020 sebanyak 143 record.

- Hasil pengujian menggunakan Algorithma C4.5 dalam memprediksi predikat kelulusan menghasilkan nilai akurasi sebesar 75%, rata-rata presisi sebesar 74% dan rata-rata sensitivitas sebesar 67%.
- Variabel yang berpengaruh terhadap predikat kelulusan adaah IPS4, NKK, IPK dan pengalaman organisasi.

4.2. Saran

Dari hasil penelitian yang dilakukan maka di perlukan saran pada penelitian ini di antaranya :

- Untuk menghasilkan akurasi yang stabil terhadap model, maka di perlukan data yang memiliki varibel dan data yang bagus.
- Dengan di peroleh nya hasil dari penelitian ini, politeknik kampar dapat memanfaatkan hasil prediksi tersebut guna untuk evaluasi pola belajar agar nilai akademik bisa di tingkatkan.
- Diharapkan penelitian ini bisa dikembangkan lagi ke tahap sebuah sistem atau aplikasi agar lebih mudah dalam pemanfaatanya.

5. Referensi

- Agustian, S., & Ramadhani, S. (2022). *Jurnal Computer Science and Information Technology (CoSciTech) menggunakan algoritma lexrakn*. 3(3), 371–381
- Aryadi. (2020). *Kampus Bina Widya Km. 12,5 Simpang Baru Panam Pekanbaru.Telp (0761) 63277*. 1(421), 1–15.
- Elanda, & Suhada. (2021). *Klasifikasi Predikat Tingkat Kelulusan Mahasiswa*

- Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang). *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, 1(2), 14–27. <https://doi.org/10.35969/dirgamaya.v1i2.182>
- Farida, Martaleli, & Nurfadila. (2020). Implementasi Algoritma C4.5 Untuk Menentukan Uang Kuliah Tunggal (Ukt) Mahasiswa (Studi Kasus : Mahasiswa Umrah Tahun 2018). 88–97.
- Fenty kurnia oktorina, & Andri nofiar, A. nofiar. (2021). 9 th Applied Business and Engineering Conference Implementasi Dashboard Business Intelligence Untuk 9 Th Applied Business And Engineering Conference. 10–19.
- Ginting, V. S., Kusriani, K., & Luthfi, E. T. (2020). Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python. *Jurnal Teknologi Informasi*, 4(1), 1–6. <https://doi.org/10.36294/jurti.v4i1.1101>
- Jumaroh, Sorikhi, & Tezhar. (2020). *Prediksi Penerima Beasiswa Menggunakan Algoritma C4 . 5. I(1)*, 36–42.
- Made, D., Mertha, A., & Paramitha, A. A. I. I. (2022). Penerapan Data Mining untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4 . 5 : Studi Kasus STMIK Primakara. 6, 84–97.
- Medy, & Alviva. (2019). Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen. *Progres*, 49–55. <https://jurnal.stmikprofesional.ac.id/index.php/Progress/article/view/146/22>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. (2022). Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest. *SATIN - Sains Dan Teknologi Informasi*, 8(2),122–131.
- Ramadani, & Alexander. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 262–276. <https://doi.org/10.35957/jatisi.v7i2.195>
- Rohman, A., & Mujiyono, S. (2021). Permodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Decision Tree C4 . 5. II(2), 1–5.
- Romli, I., & Zy, A. T. (2020). Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 4(2), 694–702.
- Taufiq. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44. <https://doi.org/10.35585/inspir.v10i1.2535>
- Wibowo, A., & Rohman, A. (2022). *Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ*. 8(200)