

## PERBANDINGAN METODE ALGORITMA C4.5 DAN NAÏVE BAYES UNTUK MEMPREDIKSI KELULUSAN MAHASISWA

Aulia Rahmayanti, Lili Rusdiana, Suratno Suratno

STMIK Palangkaraya

auliarahmayanti125@gmail.com, fasliiana7@gmail.com, ksuratno512@gmail.com

### Abstract

*This study was conducted to compare the accuracy of two algorithm methods, namely the C4.5 algorithm and Naïve Bayes algorithm on a number of datasets. The sources of datasets used are student data of the Informatics Engineering Study Program (STMIK Palangkaraya) where each dataset has a different amount of data (instances) and number of attributes. Based on the results of the comparison study of the C4.5 and Naïve Bayes Algorithm for Predicting the On-time Graduation of STMIK Palangkaraya Students, the results obtained from the accuracy of the two algorithms show that the accuracy of the C4.5 Algorithm is 90% better than the Naïve Bayes algorithm, which is only 85%. The C4.5 algorithm also gives values on recall and precision of 92% and 94% better than the Naïve Bayes algorithm with values of 86% and 93%, respectively.*

**Keyword** : C4.5 algorithm, Naïve Bayes, Graduation Prediction, STMIK Palangkaraya

### Abstrak

Penelitian ini dilakukan untuk membandingkan akurasi dua metode algoritma, yaitu algoritma C4.5 dan *Naïve Bayes* pada sejumlah *datasets*. Adapun sumber *datasets* yang digunakan yaitu data mahasiswa Program Studi Teknik Informatika (STMIK Palangkaraya) yang mana setiap *datasets* memiliki jumlah data (*instances*) dan jumlah atribut (*attributes*) yang berbeda. Berdasarkan hasil penelitian Perbandingan Algoritma C4.5 Dan *Naïve Bayes* untuk memprediksi kelulusan tepat waktu Mahasiswa STMIK Palangkaraya, diperoleh hasil akurasi dari kedua algoritma tersebut menunjukkan tingkat akurasi Algoritma C4.5 sebesar 90% lebih baik dari pada algoritma *Naïve Bayes* yakni yang hanya sebesar 85%. Algoritma C4.5 juga memberikan nilai pada *recall* dan *precision* sebesar 92% dan 94% lebih baik dibandingkan algoritma *Naïve Bayes* dengan nilai 86% dan 93%.

**Kata Kunci:** Algoritma C4.5, *Naïve Bayes*, Prediksi Kelulusan, STMIK Palangkaraya

## 1. PENDAHULUAN

Prediksi merupakan suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin akan terjadi berdasarkan informasi sebelumnya dan sekarang yang dimiliki, agar kesalahan prediksi (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diminimalisir. Pada Perguruan Tinggi, seperti STMIK Palangkaraya, prediksi juga perlu dilakukan dalam pembelajaran formal supaya dapat memperkecil kemungkinan yang terjadi, yakni kemungkinan Mahasiswa mendapatkan nilai yang kurang baik maupun menghindari ketidakkelulusan Mahasiswa program studi Teknik Informatika dalam suatu mata kuliah. Dengan menggunakan data mining, setiap kumpulan atau gudang data dapat memberikan pengetahuan penting. Dalam data mining terdapat banyak teknik dalam pengerjaannya, seperti algoritma Naïve Bayes, algoritma C4.5, jaringan saraf tiruan, dan masih banyak lainnya.

Informasi tingkat kelulusan dari Mahasiswa STMIK Palangkaraya sangat penting untuk meningkatkan pelayanan yang dapat membuat Mahasiswa lulus tepat waktu. Penggunaan data mining dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan khususnya faktor dalam data induk mahasiswa.

Jika masa studi Mahasiswa dapat diketahui lebih dini, maka pihak program studi Teknik Informatika dapat melakukan tindakan-tindakan yang diperlukan supaya Mahasiswa dapat lulus tepat waktu.

Berdasarkan masalah tersebut, dilakukan penelitian untuk membandingkan akurasi dua metode algoritma yang merupakan bagian dari

data mining, yaitu algoritma C4.5 dan Naïve Bayes pada sejumlah *datasets*. Perbandingan metode Algoritma C4.5 dan Naïve Bayes dilakukan dalam melakukan prediksi kelulusan tepat waktu Mahasiswa. Penelitian ini juga bertujuan untuk mengetahui algoritma mana yang memiliki akurasi lebih tinggi di antara Algoritma C4.5 dan Naïve Bayes dalam melakukan prediksi kelulusan tepat waktu Mahasiswa.

Penelitian ini menggunakan beberapa variabel yang mempengaruhi kelulusan Mahasiswa pada STMIK Palangkaraya yaitu IPK Semester 1 sampai 6 dan Jenis Kelamin. Data yang akan diproses hanya sebagian dari angkatan 2015 dan 2016 program studi Teknik Informatika.

*Tool* atau *Software* yang digunakan untuk pengujian metode yaitu *Weka*. *Datasets* yang digunakan yakni data lengkap tanpa missing value.

## 2. METODE

Penelitian ini merupakan penelitian eksperimen dengan langkah-langkah atau metode penelitian sebagai berikut:

### a. Pengumpulan Data

Adapun data yang digunakan dalam penelitian ini yakni data primer dan data sekunder.

- 1) Data Primer yakni data yang diperoleh secara langsung dari sumber, yaitu diperoleh secara langsung dari STMIK Palangkaraya yaitu dari data akademik mahasiswa angkatan 2015 dan 2016 pada Program Studi Teknik Informatika.
- 2) Data Sekunder yaitu data yang diperoleh secara tidak langsung, misalnya dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah

- yang diteliti yaitu penggunaan algoritma C4.5 dan *Naïve Bayes*.
- b. Pengolahan Awal Data  
Data yang digunakan yaitu data akademik mahasiswa, data terdiri dari NIM, tempat tanggal lahir, asal sekolah, IPK semester 1 sampai semester 6 dan tanggal kelulusan. Data diambil dari data mahasiswa tahun 2015 dan 2016. Dari data yang sudah ada dipecah lagi menjadi beberapa atribut yaitu NIM, Jenis Kelamin, IPK semester 1, IPK semester 2, IPK semester 3, IPK semester 4, IPK semester 5, IPK semester 6 dan kelulusan.
  - c. *Cleaning Data*  
Proses *cleaning* data atau pembersihan dilakukan terhadap data yang redundan atau ganda, inkonsisten, *missing value* atau *missing data* dan *outlier data*. Hal tersebut perlu dilakukan agar tidak mempengaruhi performa proses klasifikasi yang dilakukan. Pembersihan data yang dilakukan juga menghapus fitur yang tidak diperlukan dalam proses klasifikasi nantinya.
  - d. *Training Data* dan Pengujian Metode dengan *software Weka*.  
Pemrosesan *training* data diambil dari sebagian dataset mahasiswa. Pengujian dilakukan menggunakan *software Weka*, dengan cara memasukkan data menggunakan *file* berformat *excel*. Besarnya proporsi data yang dilakukan pengujian *training* data adalah 60%, sedangkan data lainnya digunakan untuk uji coba metode. Pengujian *training* data sehingga menghasilkan beberapa aturan dan membentuk sebuah pohon keputusan. Pada penelitian ini komparasi algoritma yang digunakan ada dua yaitu algoritma klasifikasi C4.5 dan *Naïve Bayes*.

- e. Evaluasi dan validasi penelitian  
Pada penelitian ini dilakukan evaluasi dengan cara menghitung jumlah prediksi lulus tepat waktu dan prediksi lulus tidak tepat waktu dengan pada algoritma C4.5 dan *Naïve Bayes*. Langkah selanjutnya dilakukan proses analisis data, dihitung tingkat akurasi menggunakan rumus *Confusion Matrix*.

### 3. KERANGKA TEORI

Pemilihan penggunaan Algoritma C4.5 dan *Naïve Bayes* pada penelitian ini didasarkan pada beberapa alasan, yaitu: Kedua algoritma memiliki kesamaan mudah diimplementasikan dan dapat memberikan hasil yang baik dalam kasus klasifikasi. Kedua algoritma tersebut juga mempunyai beberapa keunggulan yakni C4.5 merupakan algoritma klasifikasi pohon keputusan yang efisien dalam menangani atribut bertipe diskret dan numerik, sedangkan *Naive Bayes* hanya membutuhkan satu kali *scan data training* (Han, Kamber, & Pei, 2012).

Perbandingan antara kedua metode tersebut pernah dilakukan untuk klasifikasi penerima beasiswa. Ditunjukkan bahwa tingkat akurasi model algoritma C4.5 lebih baik dari tingkat akurasi model algoritma *Naive Bayes* (Anam & Santoso, 2018). Perbandingan metode yang sama juga dilakukan untuk penentuan penyakit anak dengan hasil akurasi yang sama pula (Bahri, Midyanti, & Hidayati, 2018). Begitu pula untuk prediksi kegiatan penerimaan Mahasiswa baru (Yahya & Jananto, 2019) dan penentuan kelayakan penerima bantuan program keluarga harapan (Fitriani, 2020).

Terkait kelulusan Mahasiswa, penelitian yang pernah dilakukan dengan pemodelan K-Means (Rusdiana

& Sam'ani, 2017) dan algoritma *K-Means* (Rusdiana, 2018). Perbandingan metode pun pernah dilakukan terkait penelitian untuk kelulusan Mahasiswa menggunakan metode *K-Nearest Neighbor* dan *Fuzzy C-Means* (Rusdiana, 2017). Implementasi metode *Fuzzy C-Means* dalam sebuah aplikasi untuk

menentukan predikat kelulusan Mahasiswa (Rusdiana & Rosmiati, 2016). Pengukuran kinerja untuk klasifikasi pada dataset, memiliki dua kelas yaitu *class* positif dan *class* negatif, maka dapat dibuatkan seperti pada Tabel 1 (Wati, 2016).

Tabel 1  
*Class pada Confussion Matrix*

Classification		Predicted Class	
		Class = Yes	Class = No
Observed Class	Class = Yes	<i>a true positive - (TP)</i>	<i>b false negative - (FN)</i>
	Class = No	<i>c false positive - (FP)</i>	<i>d true negative - (TN)</i>

Keterangan:

*True Positive (TP)* : Proporsi positif dalam dataset yang diklasifikasikan positif

*True Negative (TP)* : Proporsi negatif dalam dataset yang diklasifikasikan negatif

*False Positive (FP)* : Proporsi negatif dalam dataset yang diklasifikasikan positif

*False Negative (FN)* : Proporsi negatif dalam dataset yang diklasifikasikan negatif

#### 4. PEMBAHASAN

*Data testing* yang diambil dari data kelulusan mahasiswa dengan perbandingan 78%:22% dari 90 data, total 70 *data training* dan 20 *data testing*.

Pengujian sistem yang dilakukan dengan data kelulusan Mahasiswa menggunakan Algoritma *C4.5* dan *Naive Bayes*. *Data testing* yang digunakan seperti pada tabel 2.

Tabel 2  
*Data Testing*

No	Nim	Gender	1	2	3	4	5	6	Kelulusan
1	C1557201140	L	C	C	SM	C	SM	C	Tepat Waktu
2	C1555201043	P	SM	SM	C	SM	C	SM	Tepat Waktu
3	C1555201011	P	SM	M	C	C	M	SM	Tepat Waktu
4	C1557401056	L	SM	C	C	C	C	SM	Tepat Waktu
5	C1555401061	L	C	SM	C	C	C	C	Tepat Waktu
6	C1557401081	P	C	C	SM	C	C	C	Tepat Waktu
7	C1555401024	P	C	SM	C	C	C	SM	Tepat Waktu
8	C1555201022	L	SM	SM	SM	C	C	SM	Tepat Waktu
9	C1555201031	L	M	SM	SM	C	C	C	Tepat Waktu
10	C1555201036	P	SM	SM	C	C	C	C	Tepat Waktu
11	C1557201031	P	M	SM	M	SM	SM	SM	Tidak Tepat Waktu
12	C1557201093	P	SM	SM	SM	C	C	SM	Tidak Tepat Waktu

No	Nim	Gender	1	2	3	4	5	6	Kelulusan
13	C1557201038	P	M	M	SM	M	SM	SM	Tidak Tepat Waktu
14	C1557201012	P	SM	SM	SM	SM	C	SM	Tidak Tepat Waktu
15	C1557201076	L	M	M	SM	SM	SM	C	Tidak Tepat Waktu
16	C1557201071	L	SM	SM	M	M	SM	C	Tidak Tepat Waktu
17	C1557201083	L	M	SM	SM	C	C	SM	Tidak Tepat Waktu
18	C1557201092	L	SM	SM	M	C	C	SM	Tidak Tepat Waktu
19	C1557201098	L	C	M	SM	SM	C	SM	Tidak Tepat Waktu
20	C1555201109	P	SM	SM	C	SM	SM	C	Tidak Tepat Waktu

Tabel 3  
Kategori IPK

IPK	Predikat	Keterangan
2,00 - 2,75	Memuaskan	M
2,76 - 3,50	Sangat Memuaskan	SM
3,51 - 4,00	Dengan Pujian ( <i>Cumlaude</i> )	C

Tabel 3 menunjukkan keterangan dari penggunaan penjelas pada data testing (tabel 2). Gambaran pengujian Algoritma C4.5 dan *Naïve Bayes* menggunakan *tools WEKA*:

a. Preprocessing Data

1) Pemrosesan data awal

Data awal yang belum diklasifikasikan diproses menjadi data yang sudah diklasifikasikan sesuai dengan kelas kategorial/kriteria yang digunakan untuk pengujian Algoritma C4.5 dan *Naïve Bayes* dengan menggunakan *excel*.

2) Menghubungkan Data dengan Aplikasi *Weka*

*Data training* digunakan dengan data dengan presentase 78% dari data keseluruhan 90 data. Untuk melakukan proses *data mining* pada *WEKA*, dilakukan beberapa hal yaitu *data training* yang tersimpan di *excel*, lalu di *edit* menggunakan *notepad* kemudian disimpan dalam ekstensi *.csv* seperti pada gambar 1.

```
Nim,jenis_kelamin,ipk_1,ipk_2,ipk_3,ipk_4,ipk_5,ipk_6,kelulusan
C155281069,L,SM,SM,SM,SM,SM,Tepat Waktu
C155281001,L,SM,SM,C,C,C,C,Tepat Waktu
C155281074,L,SM,SM,C,C,C,C,Tepat Waktu
C155281056,L,SM,SM,SM,C,SM,SM,Tepat Waktu
C155281095,L,C,C,C,C,C,C,Tepat Waktu
C155281024,L,SM,SM,SM,SM,C,SM,Tepat Waktu
C155281049,L,SM,SM,C,C,C,C,Tepat Waktu
C155281071,L,SM,SM,SM,C,C,C,Tepat Waktu
C155281041,L,SM,M,SM,C,SM,SM,Tidak Tepat Waktu
C155281079,L,SM,SM,SM,C,C,SM,Tepat Waktu
C155281050,L,M,SM,SM,SM,M,SM,Tidak Tepat Waktu
C155281040,L,SM,M,SM,SM,SM,SM,Tepat Waktu
C155281098,L,SM,SM,SM,C,C,C,Tepat Waktu
C155281019,L,SM,M,M,SM,M,SM,Tidak Tepat Waktu
C155281046,L,SM,M,SM,C,SM,SM,Tidak Tepat Waktu
C155281035,L,SM,SM,SM,M,SM,SM,Tidak Tepat Waktu
C155281011,L,SM,M,SM,SM,SM,SM,Tidak Tepat Waktu
C1557201011,P,SM,SM,SM,SM,SM,SM,Tidak Tepat Waktu
C1557201048,L,SM,SM,SM,C,SM,C,Tidak Tepat Waktu
C1557201004,L,SM,SM,M,SM,SM,SM,Tidak Tepat Waktu
C1557201003,P,SM,SM,C,C,C,C,Tepat Waktu
C1557201025,L,SM,SM,C,SM,C,C,Tidak Tepat Waktu
E1657401013,P,SM,C,C,C,M,C,Tepat Waktu
E1657401020,L,SM,SM,SM,SM,SM,SM,Tepat Waktu
E1657401009,P,M,SM,C,SM,M,SM,Tepat Waktu
E1657401015,P,M,SM,SM,SM,M,SM,Tidak Tepat Waktu
C1657201001,L,SM,SM,M,M,M,M,Tidak Tepat Waktu
C1655201052,L,C,C,C,C,C,C,Tepat Waktu
E1657401005,L,SM,SM,SM,SM,C,SM,Tepat Waktu
C1655201012,P,C,C,C,C,C,C,Tepat Waktu
C1655201006,P,SM,C,C,SM,C,C,Tepat Waktu
C1655201066,P,C,SM,C,C,C,C,Tepat Waktu
C1657201042,L,SM,C,C,C,SM,C,Tepat Waktu
C1657201069,L,C,C,C,C,SM,C,Tepat Waktu
C1657201026,P,SM,SM,C,SM,SM,C,Tepat Waktu
C1657201014,P,C,C,C,C,C,C,Tepat Waktu
```

Gambar 1  
Data dengan file Ekstensi *.csv*

Relation: data_testing_predicted											
No.	1: Nim	2: jenis_kelamin	3: ipk_1	4: ipk_2	5: ipk_3	6: ipk_4	7: ipk_5	8: ipk_6	9: prediction margin	10: predicted Kelulusan	11: Kelulusan
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	L		SM	SM	SM	SM	SM	SM	-0.272727	Tidak Tepat Waktu	Tepat Waktu
2	P		C	C	C	C	C	C	0.904762	Tepat Waktu	Tepat Waktu
3	P		C	M	C	SM	M	C	0.904762	Tepat Waktu	Tepat Waktu
4	L		C	SM	C	SM	C	C	0.904762	Tepat Waktu	Tepat Waktu
5	L		SM	C	C	SM	C	SM	0.904762	Tepat Waktu	Tepat Waktu
6	P		SM	SM	SM	SM	C	SM	0.777778	Tepat Waktu	Tepat Waktu
7	P		SM	C	C	SM	C	C	0.904762	Tepat Waktu	Tepat Waktu
8	L		C	C	SM	SM	C	C	0.777778	Tepat Waktu	Tepat Waktu
9	L		M	C	SM	SM	C	SM	0.777778	Tepat Waktu	Tepat Waktu
10	P		C	C	C	SM	C	SM	0.904762	Tepat Waktu	Tepat Waktu
11	P		M	C	M	C	SM	C	1.0	Tidak Tepat Waktu	Tidak Tep...
12	P		C	C	SM	SM	C	C	-0.777778	Tepat Waktu	Tidak Tep...
13	P		M	M	SM	M	SM	C	0.272727	Tidak Tepat Waktu	Tidak Tep...
14	P		C	C	SM	C	C	C	-0.777778	Tepat Waktu	Tidak Tep...
15	L		M	M	SM	C	SM	SM	0.272727	Tidak Tepat Waktu	Tidak Tep...
16	L		C	C	M	M	SM	SM	1.0	Tidak Tepat Waktu	Tidak Tep...
17	L		M	C	SM	SM	C	C	-0.777778	Tepat Waktu	Tidak Tep...
18	L		C	C	M	SM	C	C	1.0	Tidak Tepat Waktu	Tidak Tep...
19	L		SM	M	SM	C	C	C	-0.777778	Tepat Waktu	Tidak Tep...
20	P		C	C	C	C	SM	SM	-0.904762	Tepat Waktu	Tidak Tep...

Gambar 2  
Pengujian Data Testing pada Algoritma C4.5

b. Pengujian Algoritma C4.5  
Pengujian *Data Testing* pada Algoritma C4.5 seperti pada gambar 2.  
Dari proses pada *WEKA* data testing yang diproses

menghasilkan 6 *data error*, dengan presentase 22% *data testing* memiliki data sebanyak 20 data. Hal ini diketahui dari hasil *confusion matrix* seperti tabel 4.

Tabel 4 :  
*Confusion Matrix Data Testing Algoritma C4.5*

N = 20	Prediksi Tepat Waktu (1)	Prediksi Tidak Tepat Waktu (0)	
<b>Aktual Tepat Waktu (1)</b>	TP = 9	FN = 1	FN+TP = 10
<b>Aktual Tidak Tepat Waktu (0)</b>	FP = 5	TN = 5	FP+TN = 10
	TP+FP = 14	FN+TN = 6	Total = 20

Tabel 5 menunjukkan nilai untuk mencari *accuracy*, *precision*, dan *recall*.  
Dengan perhitungan sebagai berikut :

$$Accuracy = (TP+TN)/Total$$

$$= (9+5)/20$$

$$= 0,7$$

$$= 70%$$

$$Precision = TP/(TP+FP)$$

$$= 9/(9+5)$$

$$= 0,64$$

$$= 64%$$

$$Recall = TP/(TP+FN)$$

$$= 9/(9+1)$$

$$= 0,9$$

$$= 90%$$

Tabel 5 : Hasil Pengujian Data Testing

Jumlah Data	Akurasi	Presisi	Recall
20	70%	64%	90%

22% dari keseluruhan data yaitu sebanyak 20 data, dengan jumlah lulus tepat waktu sebanyak 10 data dan 10 data dengan lulus tidak tepat waktu.

c. Pengujian Data Testing Naïve Bayes Data Testing digunakan untuk menguji atau mengevaluasi aturan yang dihasilkan dari pengujian data training menggunakan Naïve Bayes. Data testing memiliki presentase

Relation: data_testing_predicted											
No.	1: Nim	2: jenis_kelamin	3: ipk_1	4: ipk_2	5: ipk_3	6: ipk_4	7: ipk_5	8: ipk_6	9: prediction margin	10: predicted Kelulusan	11: Kelulusan
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1		L	SM	SM	SM	SM	SM	SM	0.990959	Tepat Waktu	Tepat Waktu
2		P	C	C	C	C	C	C	0.726284	Tepat Waktu	Tepat Waktu
3		P	C	M	C	SM	M	C	-0.395568	Tidak Tepat Waktu	Tepat Waktu
4		L	C	SM	C	SM	C	C	0.994297	Tepat Waktu	Tepat Waktu
5		L	SM	C	C	SM	C	SM	0.998861	Tepat Waktu	Tepat Waktu
6		P	SM	SM	SM	SM	C	SM	0.999245	Tepat Waktu	Tepat Waktu
7		P	SM	C	C	SM	C	C	0.991864	Tepat Waktu	Tepat Waktu
8		L	C	C	SM	SM	C	C	0.342664	Tepat Waktu	Tepat Waktu
9		L	M	C	SM	SM	C	SM	0.581974	Tepat Waktu	Tepat Waktu
10		P	C	C	C	SM	C	SM	0.997456	Tepat Waktu	Tepat Waktu
11		P	M	C	M	C	SM	C	0.99505	Tidak Tepat Waktu	Tidak Tep...
12		P	C	C	SM	SM	C	C	-0.591793	Tepat Waktu	Tidak Tep...
13		P	M	M	SM	M	SM	C	0.998693	Tidak Tepat Waktu	Tidak Tep...
14		P	C	C	SM	C	C	C	0.400056	Tidak Tepat Waktu	Tidak Tep...
15		L	M	M	SM	C	SM	SM	0.970319	Tidak Tepat Waktu	Tidak Tep...
16		L	C	C	M	M	SM	SM	0.921499	Tidak Tepat Waktu	Tidak Tep...
17		L	M	C	SM	SM	C	C	0.566662	Tidak Tepat Waktu	Tidak Tep...
18		L	C	C	M	SM	C	C	0.290384	Tidak Tepat Waktu	Tidak Tep...
19		L	SM	M	SM	C	C	C	0.64154	Tidak Tepat Waktu	Tidak Tep...
20		P	C	C	C	C	SM	SM	-0.86399	Tepat Waktu	Tidak Tep...

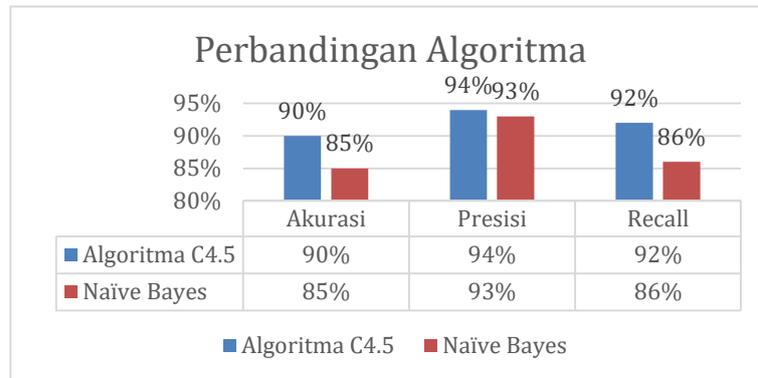
Gambar 3 Pengujian Data Testing

Dari proses pada WEKA data testing yang diproses menghasilkan 3 data error, dengan presentase 22% data testing

memiliki data sebanyak 20 data. Hal ini diketahui dari hasil *confusion matrix*

Tabel 6 Confusion Matrix Data Testing Naïve Bayes

N = 20	Prediksi Tepat Waktu (1)	Prediksi Tidak Tepat Waktu (0)	
Aktual Tepat Waktu (1)	TP = 9	FN = 1	FN+TP = 10
Aktual Tidak Tepat Waktu (0)	FP = 2	TN = 8	FP+TN = 10
	TP+FP = 11	FN+TN = 9	Total = 20



Grafik 1

Perbandingan Algoritma dengan menggunakan Data Training

Pada Tabel 7 didapatkan untuk mencari *accuracy*, *precision*, dan *recall*. Dengan perhitungan sebagai berikut:

$$Accuracy = (TP+TN)/Total$$

$$= (9+8)/20$$

$$= 0,85$$

$$= 85\%$$

$$Precision = TP/(TP+FP)$$

$$= 9/(9+2)$$

$$= 0,81$$

$$= 81\%$$

$$Recall = TP/(TP+FN)$$

$$= 9/(9+1)$$

$$= 0,9$$

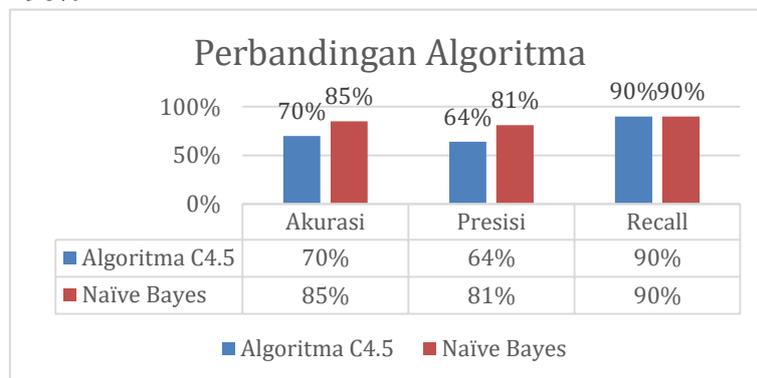
$$= 90\%$$

Tabel 7: Hasil Pengujian Data Testing

Jumlah Data	Akurasi	Presisi	Recall
20	85%	81%	90%

### Perbandingan Algoritma C4.5 dan Naive Bayes

Berikut grafik perbandingan kinerja kedua algoritma berdasarkan akurasi, presisi, dan *recall* ditampilkan pada Grafik 1 dan Grafik 2.



Grafik 2

Perbandingan Algoritma dengan menggunakan Data Testing

Hasil pengujian dan proses yang dilakukan dalam memprediksi

kelulusan Mahasiswa berdasarkan parameter Jenis Kelamin dan nilai IPK

Semester 1-6, Prodi Teknik Informatika STMIK Palangkaraya, dihasilkan suatu pola, informasi, dan pengetahuan baru sesuai dengan tujuan peneliti yaitu pola perhitungan data yang berisi *Data Training* dan *Data Testing* serta mencari probabilitas dari setiap atribut berdasarkan Data Training dan Data Testing untuk menghasilkan suatu informasi baru, pada data mahasiswa Prodi Teknik Informatika STMIK Palangkaraya akan di ketahuhi lebih dini dalam memprediksi kelulusan mahasiswa berdasarkan parameter yang sudah di tentukan, kemudian *data testing* yang akan di cari dengan memanfaatkan *data training* sebagai data historis atau pembelajaran untuk mengetahui dan memprediksi kelulusan mahasiswa kedepannya dan *data testing* sebagai data uji untuk merealisasikan suatu prediksi, sehingga setelah di uji nilai akhir dari penelitian ini selain untuk menguji data ke dalam Algoritma C4.5 dan *Naïve Bayes* yaitu nilai akhir berupa akurasi, presisi, dan *recall* yang mempresentasikan prediksi kelulusan Mahasiswa. Kemudian untuk menguji tingkat keakurasiannya maka digunakan *tool Weka* sebagai alat bantu dalam proses pengujian data. Pengujian data dilakukan dengan menggunakan *data training* sebesar 78% dan *data testing* sebesar 22%. Sebelumnya sudah dilakukan pengujian dengan 60% *data training* dan 40% *data testing* yang menghasilkan akurasi sebesar 88%, *precision* 89%, dan *recall* 88% untuk Algoritma C4.5, sedangkan untuk *Naïve Bayes* dihasilkan akurasi sebesar 83%, *precision* 86%, dan

*recall* 83%. Untuk tingkat akurasi berdasarkan proses klasifikasi menggunakan Algoritma C4.5 dan *Naïve Bayes*, dengan melalui beberapa tahapan dan pengujian yang semua tahapan dipastikan tidak ada bagian-bagian penting yang terlewatkan, sehingga dihasilkan tingkat akurasi sebesar 90%, *precision* 94%, dan *recall* 92% untuk Algoritma C4.5. Kemudian untuk *Naïve Bayes* dihasilkan tingkat akurasi sebesar 85%, *precision* 93%, dan *recall* 86%, Berdasarkan hasil proses dan pengujian data tingkat akurasi dengan menggunakan *Tool Weka* untuk percobaan tersebut dapat di bandingkan bahwa nilai akurasi algoritma terbaik adalah Algoritma C4.5 yaitu dengan tingkat akurasi sebesar 90%, *precision* 94%, dan *recall* 92%, dari hasil tersebut bisa juga disebabkan oleh beberapa faktor – faktor yaitu: faktor dalam pemilihan parameter, faktor dalam banyaknya parameter, faktor banyaknya data yang diujikan berkaitan data training dan data testingnya, dan bisa juga kurang kompleksitas data yang mengakibatkan model dapat memprediksi cukup akurat. Faktor-faktor diatas tersebut sangatlah mempengaruhi dalam pemrosesan data pada sistem.

## 5. KESIMPULAN

Berdasarkan uraian hasil penelitian, maka dapat dikemukakan kesimpulan sebagai berikut:

- a. Data diuji menggunakan *tools Weka* kemudian model yang diuji akan menghasilkan nilai *accuracy*,

- precision*, dan *recall* dari setiap algoritma.
- b. Berdasarkan hasil penelitian Perbandingan Algoritma C4.5 Dan *Naïve Bayes* untuk memprediksi kelulusan tepat waktu mahasiswa STMIK Palangkaraya, dengan mengukur kinerja kedua algoritma tersebut menggunakan *Confussion Matrix*, diperoleh hasil akurasi dari kedua algoritma tersebut menunjukkan tingkat akurasi algoritma C4.5 sebesar 90% lebih baik dari pada algoritma *Naïve Bayes* sebesar 85%.
  - c. Algoritma C4.5 juga memberikan nilai pada *recall* dan *precision* sebesar 92% dan 94% lebih baik dibandingkan algoritma *Naïve Bayes* dengan nilai 86% dan 93%.

## References

- Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Energy*, 13-19.
- Bahri, S., Midyanti, D. M., & Hidayati, R. (2018). Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)* (hal. 24-31). Yogyakarta: Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Jurnal Sistemasi*, 103-115.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques. Third Edition*. Morgan Kaufmann Publishers.
- Rusdiana, L. (2017). Perbandingan Metode K-Nearest Neighbor dan Fuzzy C-Means dalam Menentukan Predikat Kelulusan Mahasiswa. *PROSIDING SNSebatik* (hal. 21-26). Samarinda: STMIK Widya Cipta Dharma Samarinda.
- Rusdiana, L., & Rosmiati. (2016). Aplikasi Berbasis Fuzzy C-Means Dalam Penentuan Predikat Kelulusan Mahasiswa. *Jurnal Ilmu Komputer*, 1-9.
- Rusdiana, Lili. (2018). K-Means Algorithm to Group Students' Academic Status at STMIK Palangka Raya. *CSRID (Computer Science Research and Its Development Journal)*, 124-134.
- Rusdiana, Lili; Sam'ani. (2017). Pemodelan K-Means Pada Penentuan Predikat Kelulusan Mahasiswa STMIK Palangka Raya. *Jurnal Saintekom*, 1-15.
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *Jurnal Evolusi*, 4(1), 25-31.
- Yahya, N., & Jananto, A. (2019). Komparasi Kinerja Algoritma C.45 Dan Naïve Bayes Untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru. *Seminar Nasional Multi Disiplin Ilmu* (hal. 221-228). Semarang: Proceeding SENDI\_U.

This page intentionally left blank.