

MODEL PREDIKSI KELULUSAN TEPAT WAKTU MENGUNAKAN TEKNIK DATA MINING

Tasya Salsabillah¹, Lili Tanti²

1,2) Sistem Informasi Teknik dan Ilmu Komputer Universitas Potensi Utama

Article Info	ABSTRACT
Article history: Received: 04 Juli 20205 Revised: 16 Juli 20205 Accepted: 22 Juli 20205	Abstrak Pada penelitian ini bertujuan untuk membangun model prediksi kelulusan tepat waktu dengan teknik data mining menggunakan algoritma C4.5. Model ini diharapkan mampu memberikan kontribusi nyata dalam meningkatkan efisiensi dan efektivitas proses pembelajaran di perguruan tinggi serta membantu mahasiswa dalam mencapai target kelulusan tepat waktu. Dengan memanfaatkan algoritma C4.5, prediksi kelulusan tepat waktu dapat dilakukan berdasarkan sejumlah atribut atau variabel seperti Jenis Kelamin, Indeks Prestasi Kumulatif (IPK), Masa Studi, SKS, Status Pekerjaan (Bekerja, Tidak Bekerja). Hasil prediksi ini dapat membantu pihak perguruan tinggi dalam mengidentifikasi mahasiswa yang memiliki risiko tidak lulus tepat waktu Sehingga hal ini membutuhkan solusi yang lebih efisien untuk mengatasi masalah yang terjadi. Ringkasan hasil yang di uji pada penelitian ini yaitu Akurasi: 86.4%, Precision: 92.31%, Recall: 87.46% dan F1-Score: 89.86%. Kata Kunci: Prediksi, Algoritma C4.5. Abstract <i>This study aims to build a prediction model for on-time graduation using data mining techniques using the C4.5 algorithm. This model is expected to provide a real contribution in increasing the efficiency and effectiveness of the learning process in higher education and assist students in achieving the target of on-time graduation. By utilizing the C4.5 algorithm, predictions for on-time graduation can be made based on a number of attributes or variables such as Gender, Grade Point Average (GPA), Study Period, Credits, Employment Status (Employed, Not Employed). The results of this prediction can help higher education institutions in identifying students who are at risk of not graduating on time. So this requires a more efficient solution to overcome the problems that occur. The summary of the results tested in this study are Accuracy: 86.4%, Precision: 92.31%, Recall: 87.46% and F1-Score: 89.86%.</i> Keywords: Prediction, C4.5 Algorithm. Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 (CC-BY-NC-SA). 
Corresponding Author: E-mail : aaktasya26@gmail.com	

1. PENDAHULUAN

Dengan memanfaatkan algoritma C4.5, prediksi kelulusan tepat waktu dapat dilakukan berdasarkan sejumlah atribut atau variabel seperti Jenis Kelamin, Indeks Prestasi Kumulatif (IPK), Masa Studi, SKS, Status Pekerjaan (Bekerja, Tidak Bekerja).

Hasil prediksi ini dapat membantu pihak perguruan tinggi dalam mengidentifikasi mahasiswa yang memiliki risiko tidak lulus tepat waktu Sehingga hal ini membutuhkan solusi yang lebih efisien untuk mengatasi masalah yang terjadi.

Penelitian terdahulu dari Suriani (2023) mengenai Penerapan *Data Mining* untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma *Decision Tree* C4.5. Penelitian ini menggunakan variabel dependen berupa IPK dan Jumlah Mata Kuliah, variabel independen tinggi/rendah. Hasil akurasi pada penelitian ini sebesar 99.64%[1].

Penelitian terdahulu dari Hasibuan & Mahdiana (2023) mengenai Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta, Penelitian ini menggunakan variabel dependen berupa Angkatan, Jurusan, dan Nilai, variabel independen tepat waktu/telat. Hasil akurasi pada penelitian ini sebesar 75.52% [2].

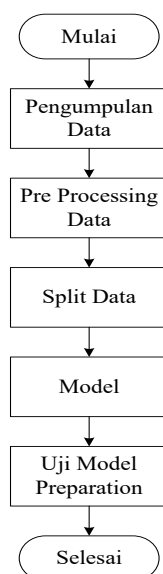
Penelitian terdahulu dari Yuliansyah et al (2021) mengenai *Predicting Students Graduate on Time Using C4.5 Algorithm*, Penelitian ini menggunakan variabel dependen berupa *Age*, *Departement*, dan GPA, variabel independen tepat waktu/tidak tepat waktu. Hasil akurasi pada penelitian ini sebesar 87.44% [3].

Penelitian terdahulu dari Dewi dan Oktaviawati (2022) mengenai Penerapan Algoritma C4.5 untuk Pemahaman Siswa SMK Pada Pelajaran Kompetensi Keahlian. Penelitian ini menggunakan Algoritma C4.5 untuk minat pemahaman siswa pada mata pelajaran kompetensi keahlian di SMK Guna Dharma Nusantara menghasilkan indikator Media pembelajaran yang memiliki *gain* paling tinggi sehingga menjadikan indikator penting dalam minat pemahaman siswa tersebut. Aplikasi yang digunakan untuk pengujian proses klasifikasi adalah Rapid Miner dengan nilai akurasi yang dihasilkan 97.22% [4].

Penelitian terdahulu dari Girsang et al (2022) mengenai Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah. Hasil Penelitian merupakan terciptanya sebuah aplikasi yang dapat memprediksi tingkat penerimaan program bantuan dengan tepat dan akurat yang nantinya diharapkan dapat membantu Kantor Kepala Desa dalam memecahkan permasalahan mengenai penerimaan program bantuan pemerintah daerah [5].

2. METODE PENELITIAN

Metodologi Penelitian merupakan bagian penting dalam proses penelitian yang menjelaskan langkah-langkah, metode, dan teknik yang digunakan untuk mengumpulkan, menganalisis, dan menginterpretasikan data. Metodologi penelitian memberikan kerangka kerja yang sistematis dan terstruktur untuk memastikan bahwa penelitian dilakukan secara ilmiah, objektif, dan dapat dipertanggungjawabkan. Metode penelitian ini menggunakan tahapan sebagai berikut:



Gambar 1. Tahapan Metodologi Penelitian

A. Pengumpulan Data

Pada tahapan ini peneliti melakukan pengumpulan data dengan melakukan observasi ke Universitas Potensi Utama agar untuk mendapatkan data mengenai Kelulusan Tepat Waktu Mahasiswa. Kegiatan pada tahap ini adalah mengumpulkan data awal, menjelaskan data, menjelajahi data, dan memverifikasi kualitas data.

Identitas Penelitian:

- 1) Peneliti: Tasya
- 2) Institusi: Universitas Potensi Utama
- 3) Tahun Penelitian: 2025
- 4) Metode: Data Mining dengan Algoritma C4.5
- 5) Objek Penelitian: Mahasiswa Universitas Potensi Utama
- 6) Jenis Data: Data Historis Kelulusan Mahasiswa Tahun 2024

Fitur-Fitur yang Digunakan dalam Model Prediksi:

- 1) Semester : Jumlah semester yang diselesaikan
- 2) Cuti : Status cuti mahasiswa/i
- 3) Nilai TA : Nilai tugas akhir (skripsi)
- 4) IPS : Index Prestasi Semester
- 5) SKS : Jumlah SKS yang sudah diselesaikan
- 6) IPK : Index Prestasi Keseluruhan

B. Pre Processing Data

Pre Processing data adalah langkah penting dalam pengolahan data yang melibatkan persiapan dan transformasi data mentah menjadi format yang lebih sesuai untuk dimodelkan.

Statistik Data:

- 1) Jumlah Data Mahasiswa: 1000
- 2) Rata-rata IPS: 3.25
- 3) Rata-rata Kehadiran: 85%
- 4) Mahasiswa yang Bekerja Sambil Kuliah: 30%
- 5) Jenis Sekolah Asal:
 - a. SMA: 60%
 - b. SMK: 30%
 - c. Madrasah Aliyah: 10%
- 6) Jurusan SMA/SMK Sesuai dengan Program Studi: 75%

C. Split Data

Split Data (Pemisahan Data) adalah proses membagi dataset menjadi beberapa subset yang berbeda untuk tujuan pelatihan, validasi, dan pengujian model dalam pengolahan data. Tujuannya adalah untuk memastikan bahwa model dapat dievaluasi secara objektif dan tidak overfitting (terlalu menyesuaikan diri dengan data pelatihan).

D. Model

Membuat dan menilai berbagai model berdasarkan beberapa teknik pemodelan yang berbeda. Pada tahap ini, ada empat tugas, yaitu memilih teknik pemodelan,

menghasilkan desain pengujian, membangun model, dan yang terakhir menilai model. Model akhir yang dihasilkan adalah prediksi kelulusan mahasiswa tepat waktu.

E. Uji Model

Pada tahapan ini peneliti melakukan uji model menggunakan algoritma C4.5. Ini adalah tahap terakhir dan paling penting dari proses CRISP-DM. Perencanaan untuk deployment dimulai dari fase business understanding dan harus menggabungkan tidak hanya untuk menghasilkan nilai model, tetapi juga cara mengonversi skor keputusan dan penggabungan keputusan tersebut.

3. HASIL DAN PEMBAHASAN

A. Data Set/Data Selection

Berikut tabel pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai:

Tabel 1. Data Kelulusan Mahasiswa/i

NO. Alumni	NIM	NAMA	JK	SEMESTER	CUTI	NILAI TA	IPS	SKS	IPK
1	1210000244	Nanda Harty Fathana	Perempuan	8	Tidak	B	3,07	148	3,27
2	1310000225	Randy Pramana Pandiangan	Laki-laki	8	Tidak	B	2,97	148	3,00
3	1310000004	Andy Kurnia Z	Laki-laki	8	Tidak	B	3,50	148	3,32
4	1310000017	Ilham	Laki-laki	8	Tidak	A	2,16	148	3,55
5	1310000019	Irfannur Diah	Laki-laki	8	Tidak	A	2,99	148	3,73
NO. Alumni	NIM	NAMA	JK	SEMESTER	CUTI	NILAI TA	IPS	SKS	IPK
6	1310000023	Muhammad Hagi Sandria	Laki-laki	8	Tidak	B	2,94	148	3,27
7	1310000086	Puteri Noviani Fajrina	Perempuan	8	Tidak	B	2,31	148	3,18
8	1310000110	Jenal Abidin	Laki-laki	8	Tidak	B	2,73	148	3,06
9	1310000119	Muhammad Juheryansah S	Laki-laki	8	Tidak	B	2,21	148	2,98
10	1310000130	Adji Syahendra	Laki-laki	8	Tidak	B	3,07	148	3,27
...	...	...	...	...	...	...	...	...	...
1000	2012000285	Muhammad Rozi Ichwan	Laki-Laki	8	Tidak	B	3,29	148	3,26

B. Pre-processing Data

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD, bisa dilihat pada tabel 2 berikut:

Tabel 2. Hasil Pre-processing/cleaning

NO. Alumni	SEMESTER	CUTI	NILAI TA	IPS	SKS	IPK
1	8	Tidak	B	2,49	142	3,27
2	8	Tidak	B	2,06	142	3,00
3	8	Tidak	B	2,23	142	3,32
4	8	Tidak	A	2,62	142	3,55
5	8	Tidak	A	2,39	142	3,73
6	8	Tidak	B	2,08	142	3,27
7	8	Tidak	B	3,37	142	3,18
8	8	Tidak	B	2,27	142	3,06
9	8	Tidak	B	2,14	142	2,98
10	8	Tidak	B	3,43	142	3,20

...	...	...	...	...	...	...
1000	8	Tidak	A	3,22	148	3,53

C. Split Data

Split Data adalah proses membagi dataset menjadi beberapa subset.

1) Training Set (Data Latihan)

- Bagian terbesar dari dataset yang digunakan untuk melatih model.
- 80% dari total data.

2) Validation Set (Data Validasi)

- Digunakan untuk menyetel parameter model (hyperparameter tuning).
- 20% dari total data.

3) Test Set (Data Uji)

- Digunakan untuk menguji performa akhir model setelah pelatihan selesai.
- 20% dari total data.

Tabel 80% dari total data:

Tabel 3. Data 80%

NO. Alumni	SEMESTER	CUTI	NILAI TA	IPS	SKS	IPK
1	8	Tidak	B	2,49	142	3,27
2	8	Tidak	B	2,06	142	3,00
3	8	Tidak	B	2,23	142	3,32
4	8	Tidak	A	2,62	142	3,55
5	8	Tidak	A	2,39	142	3,73
6	8	Tidak	B	2,08	142	3,27
7	8	Tidak	B	3,37	142	3,18
8	8	Tidak	B	2,27	142	3,06
9	8	Tidak	B	2,14	142	2,98
10	8	Tidak	B	3,43	142	3,20
...	...	...	...	...	...	...
800	9	Tidak	A	2,25	148	3,53

Tabel 20% dari total data:

Tabel 4. Data 20%

NO. Alumni	SEMESTER	CUTI	NILAI TA	IPS	SKS	IPK
801	8	Tidak	B	2,5	148	3,14
802	8	Tidak	B	3,13	148	3,2
803	8	Ya	B	2,75	148	3,1
804	8	Tidak	A	2,78	148	3,69
805	8	Tidak	A	2,03	148	3,74
806	8	Tidak	B	2,16	148	3,39
807	8	Tidak	B	3,39	148	3,13
808	8	Tidak	B	2,44	148	3,24
809	8	Tidak	A	2,69	148	3,4

810	8	Ya	A	3,05	148	3,47
...	...	...	...	...	...	...
1000	8	Tidak	A	3,22	148	3,53

D. Model

Model yang digunakan adalah model yang terkait dengan penerapan algoritma C4.5.

Statistik Data:

1) Jumlah Data Mahasiswa: 1000

2) Rata-rata IPS: 3.25

3) Jenis Sekolah Asal:

- SMA: 60%
- SMK: 30%
- Madrasah Aliyah: 10%

Berikut ini tabel perhitungan 80% data (800 Mahasiswa/i):

Tabel 5. Entrophy Total 80%

TTW	TW	Entrophy
346	454	
$346/800=0,4325$	$454/800=0,5675$	
$\text{Log}(0,4325)=-1,209$	$\text{Log}(0,5675)=-0,8173077$	
$(0-1,209)*1,209=-0,522991$	$(0-0,817307)*-0,817307=0,4638221$	0,986813

Berikut ini tabel total pasangan:

Tabel 6. Total Pasangan

	SM		CUTI		NILAI TA			IPS		SKS		IPK	
	8	>8	Ya	Tdk	A	B	C	≥ 2.5	< 2.5	≥ 142	< 143	≥ 3.2	< 3.2
TW	454	0	454	0	207	247	0	306	148	454	0	121	332
TTW	303	43	323	23	155	191	0	249	97	334	12	105	241
Total	757	43	777	23	362	438	0	555	245	788	12	227	573

Berikut ini tabel pembagian dengan total:

Tabel 7. Pembagian dengan Total

	SM/Total		CUTI/Total		NILAI TA/Total			IPS/Total		SKS/Total		IPK/Total	
TW	0,6	0	0,6	0	0,6	0,6	0	0,6	0,6	0,6	0	0,5	0,6
TTW	0,4	1	0,4	1	0,4	0,4	0	0,4	0,4	0,4	1	0,5	0,4

Berikut ini tabel Log (Hasil Pembagian):

Tabel 8. Log (Hasil Pembagian)

	SM/Total		CUTI/Total		NILAI TA/Total			IPS/Total		SKS/Total		IPK/Total	
TW	-0,7	0,0	-0,8	0,0	-0,8	-0,8	0,0	-0,9	-0,7	-0,8	0,0	-0,9	-0,8
TTW	-1,3	0,0	-1,2	0,0	-1,2	-1,2	0,0	-1,2	-1,3	-1,2	0,0	-1,1	-1,3

$$\text{Entropy} = (E.\text{Total} - ((H.\text{Bagi1}/800) * \text{Hasil Log1}) + (H.\text{Bagi2}/800) * \text{Hasil Log2}))$$

Berikut ini tabel hasil Entrophy:

Tabel 9. Entrophy

SM		CUTI		NILAI TA		
0,988027096	0	0,988037	0	0,98804454	0,988048	0

Berikut ini tabel hasil Entrophy:

Tabel 10. Entrophy

IPS		SKS		IPK	
0,988054	0,988024	0,988047	0	0,98804454	0,988048

$$\text{Information Gain} = (E.\text{Total} - ((\text{Total1}/800) * \text{Entropy1}) + (\text{Total2}/800) * \text{Entropy2}))$$

$$\text{Split Info} = 0 - ((\text{Total1}/800) * \text{LOG}(\text{Total1}/800) - ((\text{Total2}/800) * \text{LOG}(\text{Total2}/800)))$$

Berikut ini tabel IG, SI dan Akar Pohon:

Tabel 11. IG, SI dan Akar Pohon

	SM	CUTI	NILAI TA	IPS	SKS	IPK
IG	0,051892575	0,027182	-0,0012335	-0,00123	0,013592	-0,00123
SI	0,302118004	0,188084	0,99348001	0,888797	0,112361	0,86051
GR	0,171762604	0,14452	-0,0012416	-0,00139	0,120964	-0,00143
Akar	1	2	4	5	3	6

Dari hasil perhitungan diatas maka diperoleh puncak pohon yaitu:

- a) Semester
- b) Cuti
- c) SKS
- d) Nilai TA
- e) IPS
- f) IPK

Struktur Pohon keputusan:

Semester <= 8?

├─ Tidak → Tidak Tepat Waktu

└─ Ya:

Cuti = Tidak?

├─ Ya → Tidak Tepat Waktu

└─ Tidak:

SKS ≥ 144 ?

└─ Tidak → Tidak Tepat Waktu

└─ Ya:

Nilai TA = A/B/C?

└─ C → Tidak Tepat Waktu

└─ A atau B:

IPS ≥ 2.5 ?

└─ Tidak → Tidak Tepat Waktu

└─ Ya:

IPK ≥ 3.0 ?

└─ Ya → ** Tepat Waktu **

└─ Tidak → **Tidak Tepat Waktu **

Berikut ini tabel perhitungan 20% data (200 Mahasiswa/i):

Tabel 12. Entrophy Total 20%

TTW	TW	Entrophy
110	90	
$110/200=0,55$	$90/200=0,45$	
$\text{Log}(0,55)=-0,8625$	$\text{Log}(0,45)=-1,152$	
$(0-0,55)*0,8625 = -0,522991$	$(0-0,45)* -1,152=0,516125$	0,992774

Berikut ini tabel total pasangan:

Tabel 13. Total Pasangan

	SM		CUTI		NILAI TA			IPS		SKS		IPK	
	8	>8	Ya	Tdk	A	B	C	≥ 2.5	<2.5	>142	≤ 143	≥ 3.2	<3.2
TW	89	0	89	0	31	58	0	58	31	89	0	62	27
TTW	95	16	99	12	36	75	0	80	31	107	4	73	38
Total	184	16	188	12	67	133	0	138	62	196	4	135	65

Berikut ini tabel pembagian dengan total:

Tabel 14. Pembagian dengan Total

	SM/Total		CUTI/Total		NILAI TA/Total			IPS/Total		SKS/Total		IPK/Total	
TW	0,5	0	0,5	0	0,5	0,4	0	0,4	0,5	0,5	0	0,5	0,4
TTW	0,5	1	0,5	1	0,5	0,6	0	0,6	0,5	0,5	0	0,5	0,6

Berikut ini tabel Log (Hasil Pembagian):

Tabel 15. Log (Hasil Pembagian)

	SM/Total		CUTI/Total		NILAI TA/Total			IPS/Total		SKS/Total		IPK/Total	
TW	-1,04	0	-1,1	0	-1,1	-1,2	0	-1,3	-1	-1,1	0	-1,1	-1,3
TTW	-0,95	0	-0,9	0	-0,9	-0,8	0	-0,8	-1	-0,9	0	-0,9	-0,8

Entrophy=(E.Total-((H.Bagi1/800)*Hasil Log1)+(H.Bagi2/800)*Hasil Log2)))

Berikut ini tabel Entrophy:

Tabel 16. Entrophy

SM		CUTI		NILAI TA		
0,997771	0	0,997764	0	0,997754	0,99401	0

Berikut ini tabel Entrophy:

Tabel 17. Entrophy

IPS		SKS		IPK		
0,997682	0,994024	0,997744	0	0,99775	0,993999	

Berikut ini tabel IG, SI dan Akar Pohon:

Tabel 18. IG, SI dan Akar Pohon

	SM	CUTI	NILAI TA	IPS	SKS	IPK
IG	0,074825	0,054876	-0,00249	-0,00377	0,014985	-0,00376
SI	0,402179	0,327445	0,919953	0,893173	0,141441	0,909736
GR	0,18605	0,167589	-0,00271	-0,00423	0,105948	-0,00423
Akar	1	2	4	5	3	6

Dari hasil perhitungan diatas maka diperoleh puncak pohon yaitu:

- a) Semester
- b) Cuti
- c) SKS
- d) Nilai TA
- e) IPS
- f) IPK

Struktur Pohon keputusan:

Semester ≤ 8 ?

└─ Tidak → Tidak Tepat Waktu

└─ Ya:

Cuti = Tidak?

└─ Ya → Tidak Tepat Waktu

└─ Tidak:

SKS ≥ 144 ?

└─ Tidak → Tidak Tepat Waktu

└─ Ya:

Nilai TA = A/B/C?

└─ C → Tidak Tepat Waktu

└─ A atau B:

IPS $\geq$ 2.5?

└─ Tidak → Tidak Tepat Waktu

└─ Ya:

IPK $\geq$ 3.0?

└─ Ya → ** Tepat Waktu **

└─ Tidak → **Tidak Tepat Waktu **

E. Uji Model

Pada tahapan ini peneliti melakukan uji model menggunakan algoritma C4.5. Ini adalah tahap terakhir dan paling penting dari proses CRISP-DM.

1) Split Data

○ Hitung Jumlah Data Training dan Testing

- Training Data (80%) = $80\% \times 1000 = 800$ data
- Testing Data (20%) = $20\% \times 1000 = 200$ data

○ Split Confusion Matrix

Berikut tabel dari confusion matrix:

Tabel 19. Confusion Matrix

	Prediksi Tepat Waktu	Prediksi Tidak Tepat Waktu
Aktual Tepat Waktu	TP = 600	FN = 86
Aktual Tidak Tepat Waktu	FP = 50	TN = 264

Bagikan 80% ke Training dan 20% ke Testing

Training Data (80%)

- a. TP_train = $80\% \times 600 = 480$
- b. FN_train = $80\% \times 86 = 69$
- c. FP_train = $80\% \times 50 = 40$
- d. TN_train = $80\% \times 264 = 211$

Training Data (20%)

- a. TP_train = $20\% \times 600 = 120$
- b. FN_train = $20\% \times 86 = 17$
- c. FP_train = $20\% \times 50 = 10$

$$d. TN_{train} = 20\% \times 264 = 53$$

o Hasil Split Data

Berikut ini tabel trining set (800 data):

Tabel 20. Trining Set (800 data)

	Prediksi Tepat Waktu	Prediksi Tidak Tepat Waktu	Total Aktual
Aktual Tepat Waktu	TP = 480	FN = 69	549
Aktual Tidak Tepat Waktu	FP = 40	TN = 211	251

Berikut ini tabel trining set (200 data):

Tabel 21. Training Set (200 data)

	Prediksi Tepat Waktu	Prediksi Tidak Tepat Waktu	Total Aktual
Aktual Tepat Waktu	TP = 120	FN = 17	137
Aktual Tidak Tepat Waktu	FP = 10	TN = 53	63

2) Akurasi:

$$\text{Accuracy} = \frac{600 + 264}{600 + 50 + 86 + 264} = \frac{864}{1000} = 0.864 \text{ (86,4\%)}$$

3) Confusion Matrix:

Confusion matrix terdiri dari:

- True Positive (TP): Jumlah data yang benar-benar Tepat Waktu dan diprediksi Tepat Waktu.
- True Negative (TN): Jumlah data yang benar-benar Tidak Tepat Waktu dan diprediksi Tidak Tepat Waktu.
- False Positive (FP): Jumlah data yang sebenarnya Tidak Tepat Waktu tetapi diprediksi Tepat Waktu.
- False Negative (FN): Jumlah data yang sebenarnya Tepat Waktu tetapi diprediksi Tidak Tepat Waktu.

Berikut ini tabel prediksi waktu:

Tabel 22. Prediksi Waktu

	Prediksi Tepat Waktu	Prediksi Tidak Tepat Waktu
Aktual Tepat Waktu	TP = 600	FN = 86
Aktual Tidak Tepat Waktu	FP = 50	TN = 264

4) Precision, Recall, dan F1-Score:

Precision :

$$\text{Precision} = \frac{600}{600 + 50} = \frac{600}{650} = 0.9231 \text{ (92,31\%)}$$

Recall :

$$\text{Recall} = \frac{600}{600 + 86} = \frac{600}{686} = 0.8746 \text{ (87,46\%)}$$

F1-Score :

$$\text{F1 - Score} = 2 \times \frac{0.9231 \times 0.8746}{0.9231 + 0.8746} = 0.8986 \text{ (89.86\%)}$$

Ringkasan Hasil:

a. Akurasi: 86.4%

b. Precision: 92.31%

c. Recall: 87.46%

F1-Score: 89.86%

4. SIMPULAN

Pada penelitian ini Universitas Potensi Utama dapat memprediksi kelulusan tepat waktu mahasiswa berdasarkan data-data yang lalu. Dalam menerapkan Algoritma C4.5 dapat memprediksi kelulusan tepat waktu mahasiswa. Pada hasil uji dari prediksi kelulusan tepat waktu mahasiswa dengan teknik data mining menghasilkan Akurasi: 86.4%, Precision: 92.31%, Recall: 87.46% dan F1-Score: 89.86%.

REFERENCES

- [1] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5", *Journal of Computer and Information Systems Ampera.*, Vol. 3, No. 2, May, 2023e-ISSN: 2775-2496.
- [2] T. H. Hasibuan, & D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta", *istem Komputer dan Teknik Informatika*, Volume 6, Nomor 1, Januari2023, Halaman 61-74, E-ISSN:2721-4788.
- [3] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, & M. Wibowo, *Predicting Students Graduate on Time Using C4.5 Algorithm*, *Journal of Information Systems Engineering and Business Intelligence*, Vol.7, No.1, April 2021, ISSN 2443-2555.
- [4] S. Dewi, & Oktaviawati. Penerapan Algoritma C4.5 untuk Pehamanan Siswa SMK Pada Pelajaran Kompetensi Keahlian. *INTERNAL (Information System Journal)*. Volume 5, No.2, 116-125, Desember2022, DOI:10.32627.
- [5] R. Girsang, E. Fahmi G., & Masyuni H. Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah. *Jurnal Sistem Informasi TGD*. Volume 1, Nomor 4, Hal 449-459, Juli 2022.
- [6] Ananda, R., Sumarno, S., & Gunawan, I. (2022). Prediksi Promosi Jabatan Karyawan JNE Pematangsiantar Dengan Algoritma C4.5. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(3), 358.

- <https://doi.org/10.30865/json.v3i3.3763>
- [7] Andikos, A. F. (2019). Perancangan Aplikasi Multimedia Interaktif Sebagai Media Pembelajaran Pengenalan Hewan Pada Tk Islam Bakti 113 Koto Salak. (*Indonesia Jurnal Sakinah*) *Jurnal Pendidikan Dan Sosial Islam*, 1(1), 34–49. <http://jurnal.konselingindonesia.com/>
 - [8] Anwar, F. F., Jaya, A. I., & Abu, M. (2022). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dengan Penerapan Algoritma C4.5. *Jurnal Ilmiah Matematika Dan Terapan*, 19(1), 19–28. <https://doi.org/10.22487/2540766x.2022.v19.i1.15880>
 - [9] Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
 - [10] Gaol, L. Y. L., Safii, M., & Suhendro, D. (2021). *Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi*. 2(2), 97–106.
 - [11] Hasibuan, T. H., & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *Skanika*, 6(1), 61–74. <https://doi.org/10.36080/skanika.v6i1.2976>
 - [12] Kamal, F. (2020). Model Pembelajaran Sorogan Dan Bandongan. *Paramurobi: Jurnal Pendidikan Agama Islam*, 3(2), 15–26. <https://doi.org/10.32699/paramurobi.v3i2.1572>
 - [13] Kania Nisa Fauziah, Sudianto, S., & Septa Diana Nabella. (2022). Pengaruh Kelengkapan Data, Ketelitian, Kecepatan Dan Ketepatan Waktu Terhadap Kepuasan Konsumen Pada Pt Federal International Finance (Fif) Cabang Batam. *Postgraduate Management Journal*, 2(1), 40–51. <https://doi.org/10.36352/pmj.v2i1.418>
 - [14] Kholik, N., & Adhiwibowo, W. (2022). Sistem Informasi Perpustakaan Berbasis Web Kabupaten Demak. *Information Science and Library*, 3(1), 41. <https://doi.org/10.26623/jisl.v3i1.5106>
 - [15] Mustofa, I., & Kholil, A. N. (2022). Strategi Peningkatan Kualitas Peserta Didik dalam Penjaminan Mutu Eksternal. *Cermin Jurnal*, 2(1), 54–63. <https://ejournal.staidapondokkrempyang.ac.id/index.php/cjmp/article/view/98%0Ahttps://ejournal.staidapondokkrempyang.ac.id/index.php/cjmp/article/download/98/90>
 - [16] Permatasari, R., Arifin, M., & Padilah, R. (2021). Studi Deskriptif Dampak Psikologis Mahasiswa Program Studi Bimbingan dan Konseling Universitas PGRI Banyuwangi Dalam Penyusunan Skripsi Di Masa Pandemi COVID-19. *Jurnal Bina Ilmu Cendekia*, 2(1), 128–141. <http://jurnal.icjambi.id/index.php>
 - [17] Izzatul, H., I., Mulyadien, M., K., & Enri, U. (2022) Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidika*, 8 (10), 116-125, *Jully 2022*, DOI: <https://doi.org/10.5281/zenodo.6791722>.
 - [18] C. Rizky A. N. & Titin K. Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko Online Parfume Chantik, *Jurnal Algoritme*. Vol. 3, No. 1, Hal. 10-21, Oktober 2022, E-ISSN: 275-8796.
 - [19] F. M. A. Sofyan, A. Voutama, & Y. Umaidah. Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*. Vol. 7 No. 2, April 2023.
 - [20] M. D. Wahyudi, Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan Buku, *JurnalTeknorama (Informatika dan Teknologi ElRahma)*, Vol. 1, No. 1, pp. 1-6, Februari2023.
 - [21] A. B. I. Putra, F. Marisa, & F. Nurdiansyah, Penerapan Algoritma C4.5 Terhadap Penentuan Stok, *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 9, No. 3, Juni 2025.
 - [22] M. A. Melfia, K. A. Ramadhini, M. Ajmal M., A. Laura D., A. Aria B., & B. Oloan, L., Penerapan Data Mining Untuk Klasifikasi Data Penjualan Sembako Terlaris Dengan Algoritma C4.5, *JATI (Jurnal Mahasiswa Teknik Informatika)*. Vol. 9No. 3, Juni 2025.
 - [23] Renaldi, G., & Tanti, L. (2024, December). Penerapan Data Mining Dalam Sistem Penilaian Kinerja Karyawan Menggunakan Metode Algoritma C4. 5 Pada Pt Raharja Sinergi Komunikasi. In *Prosiding Seminar Nasional Multi Disiplin Ilmu (Senadimu)* (Vol. 1, No. 1, Pp. 447-459).
 - [24] Pala, T. N., & Tanti, L. (2024, December). Data Mining Dalam Memprediksi Tingkat Penjualan Bbm Menggunakan Metode Random Forest. In *Prosiding Seminar Nasional Multi Disiplin Ilmu (Senadimu)* (Vol. 1, No. 1, Pp. 692-704)
 - [25] Thanri, Y. Y., Iriani, J., Tanti, L., & Mukti, S. (2025). Penerapan Algoritma C4. 5 Dalam Klasifikasi Kualitas Buah Pisang Berdasarkan Karakteristik Fisik Dan Organoleptik: Application Of The C4. 5 Algorithm In Classifying Banana Fruit Quality Based On Physical And Organoleptic Characteristics. *Journal Of Computer Science And Technology Research (J-Cstr)*, 1(2), 64-72.