

## PENERAPAN MACHINE LEARNING MENGGUNAKAN ALGORITMA DECISION TREE UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA

Ely Nurhalizah Nst<sup>1</sup>, Sumijan<sup>2</sup>, Gunadi Widi Nurcahyo<sup>3</sup>

<sup>1,2,3</sup> Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang  
Lubuk Begalung, Padang, Sumatera Barat

e-mail: \*<sup>1</sup>[elynurhalizahnst@gmail.com](mailto:elynurhalizahnst@gmail.com), <sup>2</sup>[sumijan@upiypk.ac.id](mailto:sumijan@upiypk.ac.id), <sup>3</sup>[gunadiwidi@yahoo.co.id](mailto:gunadiwidi@yahoo.co.id)

### Abstract

Students are an integral part of higher education institutions, where graduation rates serve as a key indicator of academic quality and institutional effectiveness. To maintain accreditation and academic standards, universities must optimize student graduation rates. Evaluating the factors influencing graduation is crucial in identifying patterns and key determinants that contribute to academic success. This study aims to predict student graduation using Machine Learning, specifically the C5.0 Decision Tree algorithm. The findings indicate a high reliability in predicting student graduation, with an accuracy of 91.35%. The model's ability to identify on-time graduates is reflected in a recall of 93.85% for the On-Time category and 87.18% for the Delayed category. The prediction accuracy is further demonstrated by a precision of 92.42% for the On-Time category and 89.47% for the Delayed category. The F1-Score, which represents the balance between recall and precision, reaches 93.12% for the On-Time category and 88.32% for the Delayed category. These evaluation metrics indicate that the C5.0 algorithm effectively classifies students based on their likelihood of graduating with high accuracy. The predictions generated can serve as a reference for universities to identify at-risk students early, allowing the implementation of appropriate academic strategies to improve graduation rates, accreditation, and institutional quality.

**Keyword:** C5.0, Decision Tree, Graduation Prediction, Machine Learning, Students

### PENDAHULUAN

Kemajuan teknologi telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam bidang pendidikan (Mundok *et al.*, 2024). Implementasi teknologi yang berkembang pesat mampu meningkatkan keberhasilan sistem pendidikan yang memiliki peran vital dalam kemajuan suatu negara (Zebua *et al.*, 2024). Salah satu indikator keberhasilan sistem pendidikan adalah tingkat kelulusan mahasiswa yang tinggi dan tepat waktu. Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi (UUPT) (Menteri Pendidikan, kebudayaan, 2023), perguruan tinggi bertanggung jawab dalam memastikan pencapaian akademik mahasiswa sesuai dengan standar nasional pendidikan tinggi (Menteri Pendidikan, Kebudayaan, 2023). Tingkat kelulusan yang rendah dapat berdampak pada kualitas institusi, penilaian akreditasi (Umar *et al.*, 2024), serta masa depan mahasiswa, yang dapat mengalami kendala finansial, keterlambatan karier, hingga risiko *drop out* (Anggriani *et al.*, 2024).

Ketepatan waktu kelulusan mahasiswa program sarjana umumnya ditargetkan dalam rentang 4 hingga 7 tahun (Menteri Pendidikan, kebudayaan, 2023)(Suyanto *et al.*, 2024). Namun, berbagai faktor dapat memengaruhi pencapaian ini, seperti kemampuan akademik, motivasi, serta kurikulum dan beban studi yang diberikan. Upaya prediktif diperlukan untuk memahami tren tingkat kelulusan dan merancang intervensi akademik yang efektif.

Universitas Islam Negeri (UIN) Sjech M. Djamil Djambek Bukittinggi merupakan perguruan tinggi Islam negeri yang berlokasi di Bukittinggi, Sumatera Barat. Perguruan tinggi ini mengalami perubahan status dari Institut Agama Islam Negeri (IAIN) menjadi universitas pada tahun 2022 berdasarkan Peraturan Presiden Republik Indonesia No. 85 Tahun 2022 UIN Sjech M. Djamil Djambek Bukittinggi telah meraih akreditasi unggul, sehingga peningkatan

kualitas akademik, termasuk strategi meningkatkan tingkat kelulusan mahasiswa, menjadi prioritas utama. Fakultas Tarbiyah dan Ilmu Keguruan (FTIK) memiliki enam program studi yang bertujuan meraih akreditasi unggul, yang menjadikan penelitian ini relevan dalam mendukung pencapaian tersebut.

Salah satu pendekatan yang dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa adalah penerapan *Machine Learning* (ML) yang mampu membuat komputer untuk belajar dan membuat prediksi berdasarkan data (Qisthiano *et al.*, 2023). ML termasuk salah satu cabang dari kecerdasan buatan yang mampu memanfaatkan data historis, sehingga memungkinkan sistem untuk mengembangkan model prediktif yang memperkirakan hasil atau kejadian di masa depan berdasarkan data baru (Madhyastha *et al.*, 2024). Salah satu teknik yang populer dan efektif dalam analisis data dan prediksi adalah *Decision Tree*, dapat menjadi solusi efektif dalam memprediksi tingkat kelulusan mahasiswa (Fan *et al.*, 2024).

Beberapa Penelitian relevan terdahulu yang membahas tentang prediksi tingkat kelulusan mahasiswa menggunakan berbagai metode diantaranya sebagai berikut. Penelitian yang dilakukan oleh Falda Junisman Zebua, dkk (2024) di Universitas Prima Indonesia membandingkan Algoritma C5.0 dan *Regresi Linier* untuk prediksi kelulusan mahasiswa tepat waktu (Zebua *et al.*, 2024). Dengan menggunakan 715. Penelitian ini menemukan bahwa Algoritma C5.0 lebih unggul dalam memprediksi kelulusan dengan skor sebesar 96.85% pada fase pelatihan dan 93.72% pada fase pengujian, dibandingkan *Regresi Linier* yang hanya memperoleh skor R2 sebesar 33.31% dan 40.30%. Algoritma C5.0 terbukti lebih efektif, dengan 91% mahasiswa diprediksi lulus tepat waktu (Zebua *et al.*, 2024).

Penelitian menggunakan *Decision Tree* juga digunakan peneliti internasional yaitu: Douglas Vieira do Nascimento, dkk (2023) *Live Birth Forecasting in Brazilian Health Regions with Tree-based Machine Learning Models* yang dipublikasikan IEEE 36th *International Symposium on Computer-Based Medical Systems (CBMS)*. Di mana *Decision Tree* dalam penelitian ini berfungsi untuk mengidentifikasi pola dari variabel-variabel yang memengaruhi kelahiran hidup di *mikro-region* Brazi. Prediksi ini mendukung pemerintah dalam perencanaan kesehatan ibu dan bayi, serta membantu alokasi sumber daya yang lebih tepat, sejalan dengan pencapaian Tujuan Pembangunan Berkelanjutan (SDG) PBB terkait pengurangan angka kematian ibu (Do Nascimento *et al.*, 2023).

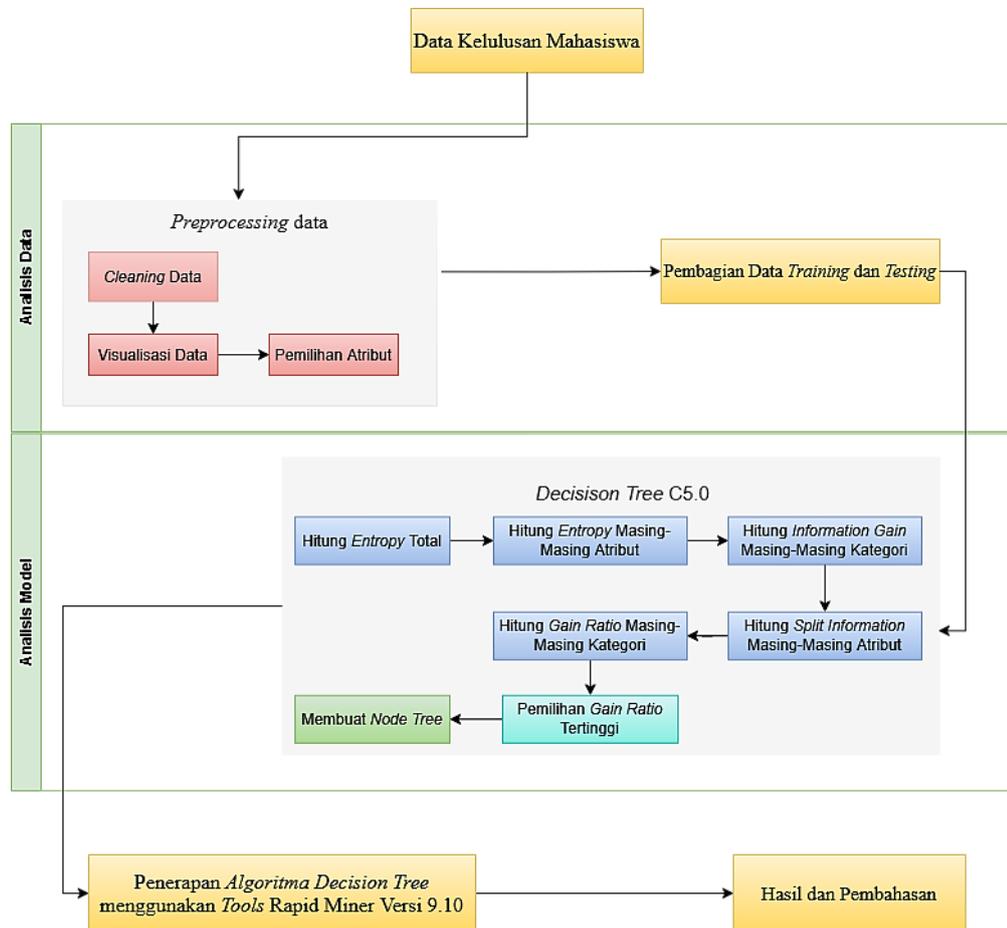
Metode *Decision Tree* ini juga dapat dikolaborasi dengan metode lainnya, seperti halnya penelitian internasional dari Guoan Zhon (2024) tentang Model prakiraan cuaca ini menggabungkan *Decision Tree* ID3 dan Neural Network (NN). Metode ID3 digunakan untuk memilih fitur terbaik dan membangun pohon keputusan dalam memprediksi tanggal kebakaran, sedangkan NN dilatih pada data suhu historis untuk memprediksi suhu maksimum dan minimum. Kombinasi ini mendukung mitigasi bencana dan peringatan dini guna mengurangi dampak cuaca ekstrem (Zhong *et al.*, 2024).

Keunggulan *Decision Tree* terletak pada interpretabilitasnya, kemampuannya untuk menangani data dengan berbagai variabel kompleks, dan menghasilkan struktur yang mudah dipahami (Firman Akbar & Rahmaddeni, 2024) (Suyanto *et al.*, 2024). Dengan penerapan *Machine Learning* menggunakan algoritma *Decision Tree*, perguruan tinggi dapat mengidentifikasi mahasiswa yang berisiko tinggi tidak lulus tepat waktu dan mengambil langkah preventif yang diperlukan (Mukrimaa *et al.*, 2024).

Berdasarkan penelitian sebelumnya, algoritma *Decision Tree* telah diterapkan dalam berbagai bidang, termasuk analisis data akademik, prediksi kelulusan mahasiswa, serta perencanaan strategis di berbagai sektor internasional, seperti kesehatan dan demografi. Keunggulannya dalam mengolah data historis dan mengidentifikasi pola relevan menjadikannya metode yang banyak digunakan untuk pengambilan keputusan berbasis data. Penelitian ini untuk menganalisis kelulusan mahasiswa di UIN Sjech M. Djamil Djambek Bukittinggi. Dengan metode *Decision Tree* yang terbukti efektif, penelitian ini diharapkan membantu perguruan tinggi mengidentifikasi mahasiswa berisiko terlambat lulus dan menyusun strategi akademik yang lebih tepat guna meningkatkan mutu pendidikan dan capaian akademik.

## METODE PENELITIAN

Metode Kerangka penelitian adalah konsep yang mengatur alur, metode, dan tahapan penelitian supaya sesuai dengan rencana awal dan menghasilkan solusi yang tepat untuk masalah yang diteliti. Kerangka kerja yang jelas sangat penting untuk mendukung keberhasilan penelitian ini. Adapun kerangka penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1. Tahapan penelitian ini menunjukkan alur penelitian dalam menganalisis data kelulusan mahasiswa menggunakan algoritma *Decision Tree* C5.0. Proses penelitian dibagi menjadi dua tahap utama, yaitu Analisis Data dan Analisis Model. Setelah melalui tahap *preprocessing* data, data dibagi menjadi *training* dan *testing* sebelum diterapkan pada model *Decision Tree*. Hasil akhirnya berupa penerapan algoritma menggunakan RapidMiner serta analisis hasil dan pembahasan.

### A. Data Kelulusan Mahasiswa

Data kelulusan mahasiswa merupakan sumber utama dalam penelitian ini. Data tersebut mencakup berbagai atribut yang berpotensi memengaruhi kelulusan, seperti jumlah SKS yang ditempuh, IPK, dan faktor lainnya. Data ini akan dianalisis guna memperoleh pola yang dapat digunakan dalam prediksi kelulusan mahasiswa.

### B. *Preprocessing* Data

Tahap ini melibatkan proses persiapan data agar siap dianalisis, dengan memastikan data bersih, terstruktur, dan relevan untuk membangun model yang akurat.

1. *Cleaning Data*  
Menghapus data tidak valid, menangani nilai hilang, duplikat, dan menyeragamkan format untuk menjaga kualitas data dan menghindari bias.
2. *Visualisasi Data*  
Dilakukan untuk memahami distribusi dan pola dalam data, serta mengidentifikasi hubungan antarvariabel yang relevan.
3. *Pemilihan Atribut*  
Menyaring atribut yang berpengaruh signifikan terhadap prediksi, serta mengeliminasi fitur yang tidak relevan guna meningkatkan akurasi dan efisiensi model.

#### C. Pembagian Data Training dan *Testing*

Data dibagi menjadi dua subset, yaitu *training* data dan *testing* data, dengan rasio tertentu, misalnya 80:20 atau 70:30. *Training* data digunakan untuk membangun model, sedangkan *testing* data digunakan untuk mengevaluasi performa model dalam melakukan prediksi.

#### D. *Decision Tree C5.0*

Tahapan analisis model dengan algoritma *Decision Tree C5.0* yang diterapkan untuk membangun model klasifikasi berdasarkan atribut-atribut yang telah dipilih. Model ini bekerja dengan membagi dataset ke dalam node berdasarkan nilai atribut yang paling berpengaruh terhadap hasil prediksi.

1. Hitung *Entropy* Total

*Entropy* mengukur tingkat ketidakteraturan dalam dataset. Semakin tinggi nilai entropi, semakin tidak teratur data tersebut. Entropi dihitung menggunakan rumus (Saputra & Kristiyanti, 2024):

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

Entropi ini melihat proporsi ( $p_i$ ) dari setiap kelas dalam dataset. Jika data terbagi secara merata, entropinya tinggi, sedangkan jika data lebih terpusat pada satu kategori, entropinya lebih rendah.

2. Hitung *Entropy* Masing-Masing Atribut

Entropi dihitung untuk setiap atribut dalam dataset guna menentukan seberapa besar kontribusinya dalam membagi data ke dalam kategori tertentu. Atribut dengan entropi yang lebih rendah memiliki peran yang lebih signifikan dalam pemisahan data.

3. Hitung *Information Gain* Masing-Masing Kategori

*Information Gain* mengukur seberapa besar suatu atribut dapat mengurangi ketidakpastian dalam dataset. *Information Gain* dihitung menggunakan rumus (Saputra & Kristiyanti, 2024):

$$IG(S, A) = E(S) - \sum_{v \in A} \frac{|S_v|}{|S|} E(S_v) \quad (2)$$

Semakin tinggi nilai *Information Gain*, semakin efektif atribut tersebut dalam membagi data berdasarkan kategori tertentu. *Information Gain* digunakan untuk memilih atribut terbaik dalam membangun pohon keputusan.

4. Hitung *Split Information* Masing-Masing Atribut

*Split Information* digunakan untuk menyesuaikan *Information Gain* terhadap jumlah kategori dalam atribut. *Split Information* dihitung menggunakan rumus (Saputra & Kristiyanti, 2024):

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

Perhitungan ini berguna untuk menghindari kecenderungan pemilihan atribut dengan jumlah kategori yang banyak tetapi kurang bermakna.

5. Hitung *Gain Ratio* Masing-Masing Kategori

*Gain Ratio* merupakan perbaikan dari *Information Gain* dengan mempertimbangkan jumlah kategori dalam atribut. *Gain Ratio* dihitung menggunakan rumus (Saputra & Kristiyanti, 2024):

$$GR(A) = \frac{IG(S, A)}{SI(A)} \quad (4)$$

Atribut dengan nilai *Gain Ratio* tertinggi akan dipilih sebagai simpul utama (*root node*) dalam *Decision Tree*. *Gain Ratio* mengatasi bias yang mungkin muncul dalam *Information Gain* terhadap atribut dengan banyak kategori.

6. Pemilihan *Gain Ratio* Tertinggi

Setelah perhitungan dilakukan untuk setiap atribut, atribut dengan nilai *Gain Ratio* tertinggi akan dipilih sebagai titik awal dalam pembentukan pohon keputusan. Pemilihan atribut yang optimal sangat menentukan efektivitas model dalam melakukan klasifikasi.

7. Membuat *Node Tree*

Setelah simpul utama (*root node*) ditentukan, algoritma akan membentuk cabang-cabang pohon berdasarkan aturan yang telah dihitung. Proses ini berlanjut hingga seluruh data terklasifikasi dengan baik.

E. Penerapan *Algoritma Decision Tree menggunakan RapidMiner 9.10*

Setelah Model diterapkan menggunakan RapidMiner 9.10 yang mendukung proses pemodelan, pelatihan, dan evaluasi dengan alat analisis yang lengkap. Penggunaannya meningkatkan efisiensi penelitian dan akurasi pengolahan data.

F. Hasil dan Pembahasan

Tahap akhir penelitian meliputi evaluasi akurasi model *Decision Tree* C5.0 dan analisis faktor-faktor yang paling memengaruhi prediksi kelulusan. Temuan ini menjadi dasar rekomendasi kebijakan untuk meningkatkan tingkat kelulusan mahasiswa.

## HASIL DAN PEMBAHASAN

Penelitian Penelitian ini menggunakan algoritma *Decision Tree* untuk menganalisis dan memprediksi tingkat kelulusan mahasiswa. Data kelulusan dikumpulkan dan diproses melalui tahap *preprocessing* guna memastikan kualitas sebelum analisis. Algoritma kemudian digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi kelulusan serta mengevaluasi kinerja model.

A. Data Kelulusan Mahasiswa

Data penelitian diambil dari mahasiswa Angkatan ke-VII yang mengikuti wisuda pada 12 September 2024. Sebanyak 520 data mahasiswa Fakultas Tarbiyah dan Ilmu Keguruan (FTIK) Universitas Islam Sjech M. Djamil Djambek Bukittinggi sebanyak 520 data mahasiswa yang terdiri dari 6 program studi, yaitu Pendidikan Teknik Informatika dan Komputer (PTIK), Pendidikan Matematika (PMTK), Pendidikan Bahasa Inggris (PBI), Pendidikan Bahasa Arab (PBA), Pendidikan Agama Islam (PAI), dan Pendidikan Bimbingan Konseling (PBK). Atribut dalam data meliputi nama, NIM, jenis kelamin, program studi, akreditasi, total SKS, IPK, yudisium, durasi pengerjaan skripsi, dan lama masa studi.

B. *Preprocessing* Data

*Preprocessing* data adalah langkah awal yang sangat penting dalam analisis data, *preprocessing* melibatkan tiga tahap utama: pembersihan data (*cleaning data*), visualisasi data, dan pemilihan atribut.

### 1. Cleaning Data

Pembersihan data dilakukan manual menggunakan Excel untuk menghapus duplikasi dan nilai kosong dari 520 data kelulusan mahasiswa yang terdiri dari 12 kolom. Hasilnya, tidak ditemukan duplikat atau nilai kosong, sehingga jumlah data tetap. Tiga atribut NIM, Nama, Jenis Kelamin, dan Lama Studi dihapus karena dianggap tidak relevan atau sudah terwakili oleh atribut lain. Selanjutnya, dilakukan standarisasi format data. Atribut Total SKS, IPK, dan Lama Skripsi/Semester diubah tipenya dari Integer/Float menjadi String. Hasil transformasi ditampilkan pada Tabel 1.

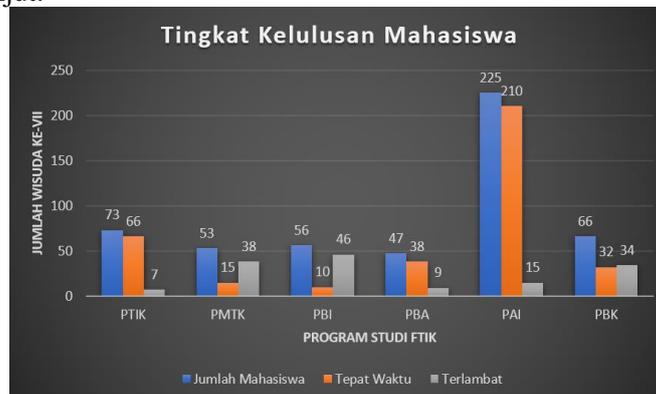
**Tabel 1.** Tabel Hasil *Cleaning* dan Transformasi Data

No	ID	Prodi	Akreditasi	Total SKS	IPK	Yudisium	Lama Masa Skripsi	Label
1	Mhs 1	PTIK	Baik Sekali	Banyak	Tinggi	Dengan Pujian	Cepat	Tepat Waktu
2	Mhs 2	PTIK	Baik Sekali	Banyak	Tinggi	Dengan Pujian	Cepat	Tepat Waktu
3	Mhs 3	PTIK	Baik Sekali	Banyak	Sedang	Sangat Memuaskan	Cepat	Tepat Waktu
4	Mhs 4	PTIK	Baik Sekali	Banyak	Tinggi	Dengan Pujian	Cepat	Tepat Waktu
5	Mhs 5	PTIK	Baik Sekali	Banyak	Tinggi	Sangat Memuaskan	Cepat	Tepat Waktu
6	Mhs 6	PTIK	Baik Sekali	Banyak	Rendah	Sangat Memuaskan	Lama	Terlambat
7	Mhs 7	PTIK	Baik Sekali	Banyak	Sedang	Sangat Memuaskan	Lama	Terlambat
...	...	...	...	...	...	...	...	...
518	Mhs 518	PBK	Baik Sekali	Banyak	Tinggi	Sangat Memuaskan	Cepat	Tepat Waktu
519	Mhs 519	PBK	Baik Sekali	Banyak	Rendah	Sangat Memuaskan	Lama	Terlambat
520	Mhs 520	PBK	Baik Sekali	Banyak	Rendah	Sangat Memuaskan	Cepat	Tepat Waktu

Dataset sebanyak 520 mahasiswa telah dibersihkan dan ditransformasikan untuk meningkatkan konsistensi dan kemudahan analisis. Atribut Total SKS, IPK, dan Lama Masa Skripsi dikategorikan ulang, sementara Status Kelulusan diubah menjadi "Label" sebagai target prediksi.

### 2. Visualisasi Data

Visualisasi digunakan untuk menampilkan tingkat kelulusan tepat waktu dan keterlambatan di berbagai program studi, yang memberikan gambaran awal sebagai dasar analisis lebih lanjut.



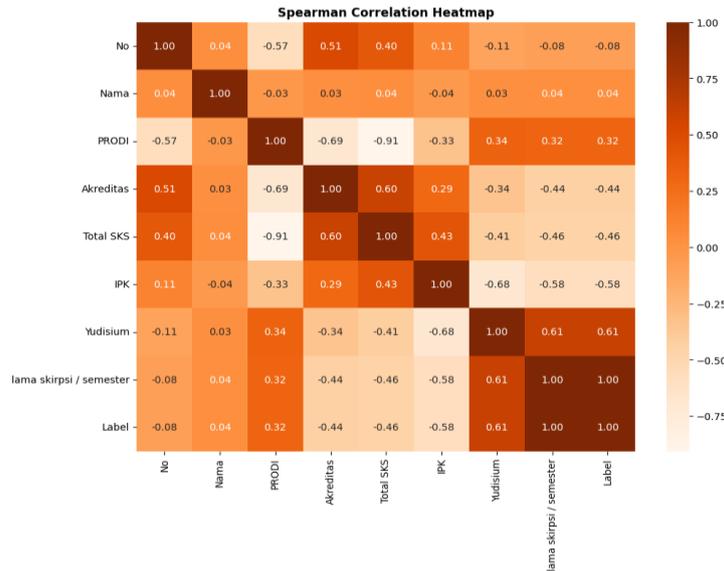
**Gambar 1.** Diagram Tingkat Kelulusan Mahasiswa

Gambar 1. menunjukkan variasi signifikan tingkat kelulusan mahasiswa FTIK antar program studi. PBK memiliki proporsi keterlambatan tertinggi dengan jumlah lulusan tepat waktu dan terlambat hampir seimbang. PTIK mencatat kelulusan tepat waktu tertinggi, sementara PAI memiliki jumlah lulusan terbanyak meskipun masih terdapat keterlambatan. Temuan ini menegaskan pentingnya evaluasi terhadap faktor-faktor penyebab keterlambatan untuk mendukung strategi akademik yang lebih efektif..

### 3. Pemilihan Atribut

Pemilihan atribut dilakukan untuk menentukan faktor yang paling berpengaruh terhadap prediksi kelulusan mahasiswa. Berdasarkan wawancara dengan bidang akademik Fakultas FTIK yaitu Wakil I Dekan FTIK dan hasil analisis korelasi Spearman, ada enam atribut

utama yang dipertimbangkan adalah Program Studi, Akreditasi, Total SKS, IPK, Yudisium, dan Lama Skripsi. Faktor utama yang memengaruhi keterlambatan kelulusan adalah manajemen waktu yang buruk dan kurangnya motivasi mahasiswa. Hasil kolerasi atribut pada Gambar 2.



**Gambar 2.** Kolerasi Data Atribut

Analisis korelasi menunjukkan bahwa Lama Skripsi memiliki hubungan kuat dengan Yudisium dan kelulusan, sementara IPK berkorelasi negatif dengan keterlambatan. Atribut yang tidak relevan dihapus untuk meningkatkan akurasi model prediksi.

C. Pembagian Data Training dan Testing

Sebelum pengolahan model Decision Tree, data kelulusan mahasiswa sebanyak 520 sampel dibagi menjadi 80% data latih (416 sampel) dan 20% data uji (104 sampel), dengan pembagian proporsional per program studi untuk menjaga representasi data. Tujuannya adalah meningkatkan generalisasi model, mengurangi bias, dan mengoptimalkan akurasi prediksi..

D. Decision Tree C5.0

Tahap ini menggunakan algoritma Decision Tree C5.0 untuk memproses data latih dan memprediksi kelulusan berdasarkan atribut terpilih.

1. Hitung Entropy Total

Entropy total dihitung untuk mengukur ketidakpastian data terhadap label kelulusan (Tepat Waktu dan Terlambat). Dari 416 data latih, 306 tepat waktu dan 110 terlambat, menghasilkan entropy total sebesar 0,833.

2. Hitung Entropy Masing-Masing Atribut

Setelah entropy totalnya didapatkan selanjutnya entropy dihitung untuk 6 atribut utama: Prodi, Akreditasi, Total SKS, IPK, Yudisium, dan Lama Skripsi. Tujuannya menentukan atribut terbaik dalam membagi data secara optimal, seperti ditampilkan pada Tabel 2.

**Tabel 2.** Hasil Perhitungan Entropy Masing-Masing Atribut

Atribut	Lulus	TW	T	Entropy
Total (Awal)	416	306	110	0,833
<b>PRODI</b>				
PTIK	58	53	5	0,424
PMTK	42	13	29	0,893
....	...	...	...	....
<b>Lama Masa Skripsi (LMS)</b>				
Cepat	357	306	51	0,592
Lama	59	0	59	0,000

Berdasarkan Tabel 2 perhitungan entropy dihitung disetiap atribut yang ada di dalam kategori. Sehingga dapat dilihat atribut mana yang memiliki nilai ketidakpastian, setelah itu baru dilakukan perhitungan *Information Gain*.

### 3. Hitung *Information Gain* Masing-Masing Kategori

Menghitung *Information Gain* (IG) untuk setiap kategori. IG digunakan untuk menentukan kategori terbaik yang dapat digunakan untuk membagi data agar memperoleh klasifikasi yang lebih jelas. Hasil perhitungan pada tabel 3.

**Tabel 3.** Tabel Hasil Perhitungan *Information Gain* Masing-Masing Kategori

Atribut	Lulus	TW	T	Entropy	IG
Total (Awal)	416	306	110	0,833	
PRODI					0,302
PTIK	58	53	5	0,424	
PMTK	42	13	29	0,893	
...	...	...	...	....	
Lama Masa Skripsi (LMS)					0,326
Cepat	306	306	51	0,592	
Lama	59	0	59	0	

Tabel 3. Menunjukkan atribut dengan kategori yudisium besar mengurangi ketidakpastian (*entropy*) dalam data karena bernilai 0,342. Namun itu belum bisa dijadikan sebagai *node*, langkah selanjutnya menghitung *Split Information*.

### 4. Hitung *Split Information* Masing-Masing Atribut

IG saja tidak cukup, maka *Split Information* (SI) dihitung untuk melihat seberapa banyak atribut membagi data menjadi subset yang lebih kecil.

**Tabel 4.** Tabel Hasil Perhitungan *Split Information* Masing-Masing Atribut

Atribut	Lulus	TW	T	Entropy	IG	SI
Total (Awal)	416	306	110	0,833		
PRODI					0,302	
PTIK	58	53	5	0,424		0,396
PMTK	42	13	29	0,893		0,334
...	...	...	...	....	...	....
Lama Masa Skripsi (LMS)					0,326	
Cepat	306	306	51	0,592		0,189
Lama	59	0	59	0		0,400

Hasil Tabel 4 menunjukkan bahwa atribut dengan nilai SI tinggi, seperti Total SKS dan IPK, membagi data ke banyak kategori sehingga Gain Ratio-nya kecil. Sebaliknya, atribut dengan SI rendah seperti Lama Skripsi lebih stabil dan menghasilkan Gain Ratio yang besar.

### 5. Hitung *Gain Ratio* Masing-Masing Kategori

Perhitungan *Gain Ratio* setelah SI membantu algoritma memilih atribut yang memisahkan data secara efektif, bukan karena jumlah kategorinya. Ini menghasilkan pohon keputusan yang lebih optimal dan sederhana. Hasilnya ditampilkan pada Tabel 5.

**Tabel 5.** Tabel Hasil Perhitungan *Gain Ratio* Masing-Masing Kategori

Atribut	L	TW	T	Entropy	IG	SI	GR
Total (Awal)	416	306	110	0,833			
PRODI					0,30		0,132
PTIK	58	53	5	0,424		0,396	
PMTK	42	13	29	0,893		0,334	
...	...	...	...	....	...	...	....
Lama Masa Skripsi (LMS)					0,326		0,553

Cepat	306	306	51	0,592	0,189
Lama	59	0	59	0	0,400

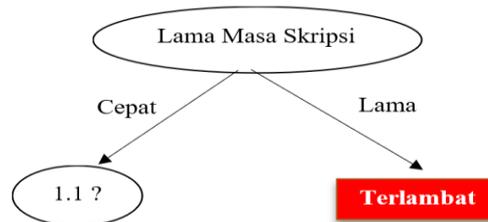
Setelah melakukan perhitungan *GR* dalam 6 kategori yang ada. Selanjutnya memilih *GR* tertinggi.

6. Pemilihan *Gain Ratio* Tertinggi

Berdasarkan Tabel 4, atribut dengan *Gain Ratio* tertinggi adalah "Lama Masa Skripsi (LMS)" sebesar 0,553. Ini menunjukkan bahwa atribut LMS paling informatif dalam membagi data dan menjadi pilihan optimal untuk membentuk pohon keputusan yang akurat dan efisien.

7. Membuat *Node Tree*

Proses pembentukan pohon keputusan dimulai dengan memilih atribut yang memiliki *Gain Ratio* tertinggi sebagai node akar, yaitu "LMS" dengan nilai 0,553, karena memberikan pemisahan data paling optimal. Setelah itu, dataset dibagi ke dalam cabang berdasarkan nilai atribut tersebut. Hasil node akar *tree* Gambar 3.

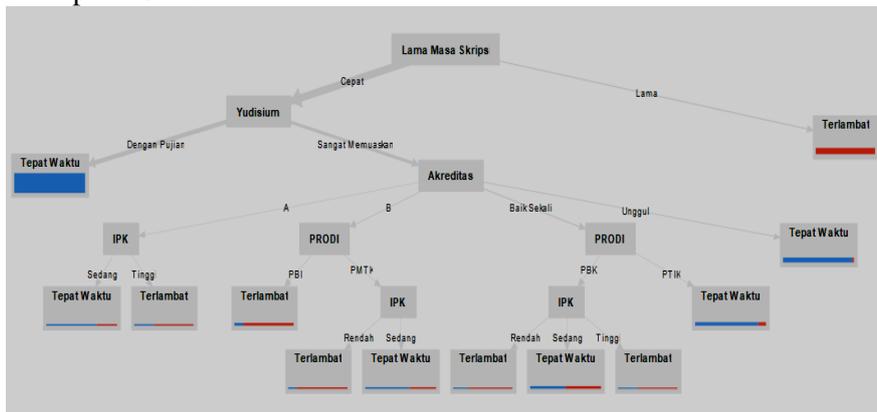


Gambar 3. *Decision Tree* Node 1

Gambar 3 menunjukkan atribut LMS (Lama Masa Skripsi) sebagai akar pohon keputusan karena memiliki *Gain Ratio* tertinggi, dengan dua kategori: Cepat dan Lama. Kategori Lama langsung mengarah pada keputusan "Terlambat", sedangkan kategori Cepat masih memiliki probabilitas, sehingga dilakukan kembali perhitungan Entropy Total dan *Gain Ratio* pada tiap cabang. Proses ini berlanjut dengan memilih atribut terbaik di setiap subset hingga seluruh cabang homogen atau tidak ada atribut signifikan. Pendekatan ini memastikan pohon keputusan tetap optimal dan tidak kompleks.

E. Penerapan Algoritma *Decision Tree* menggunakan RapidMiner 9.10

Perhitungan manual diimplementasikan dalam RapidMiner menggunakan operator *Decision Tree* dengan pemilihan atribut berbasis *Gain Ratio*. Parameter seperti kedalaman dan pruning disesuaikan untuk mengoptimalkan model. Proses ini menggunakan data training dengan metode C5.0 untuk menghasilkan struktur pohon keputusan optimal, yang divisualisasikan pada Gambar 4.



Gambar 4. Hasil *Decision Tree* Prediksi Kelulusan Mahasiswa

Hasil pada Gambar 4 menunjukkan bahwa implementasi di RapidMiner sesuai dengan perhitungan manual. Berdasarkan hasil *Decision Tree*, didapatkan 12 *rule* (aturan) yang berfungsi sebagai pola dalam memprediksi tingkat kelulusan mahasiswa. Aturan-aturan ini

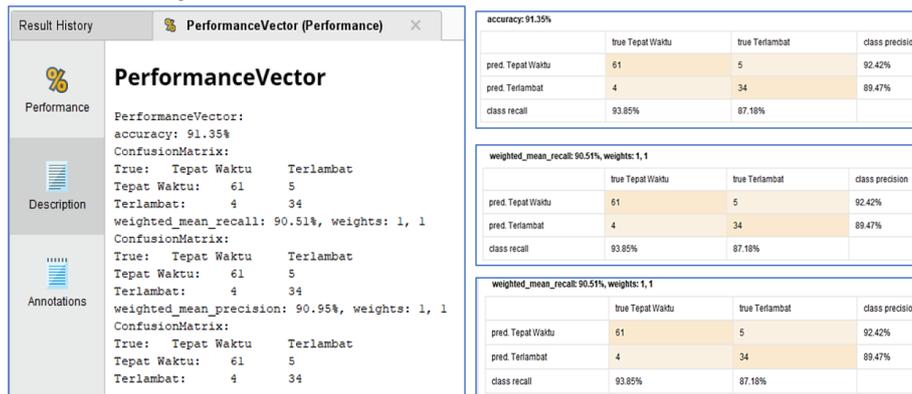
menunjukkan bagaimana kombinasi atribut seperti LMS, Yudisium, Akreditasi, PRODI, dan IPK berpengaruh terhadap ketepatan waktu kelulusan.

**Tabel 5.** Hasil *Rule* Prediksi *Decision Tree* C5.0 Kelulusan Mahasiswa

If (LMS	Yudisium	Akreditasi	PRODI	IPK)	Then (Kelulusan)
Cepat	Dengan Pujian	-(semua)	-(semua)	-(semua)	Tepat Waktu
Cepat	Sangat Memuaskan	A	-(semua)	Sedang	Tepat Waktu
Cepat	Sangat Memuaskan	A	-(semua)	Tinggi	Terlambat
Cepat	Sangat Memuaskan	B	PBI	-(semua)	Terlambat
Cepat	Sangat Memuaskan	B	PMTK	Rendah	Terlambat
Cepat	Sangat Memuaskan	B	PMTK	Sedang	Tepat Waktu
Cepat	Sangat Memuaskan	Baik Sekali	PBK	Rendah	Terlambat
Cepat	Sangat Memuaskan	Baik Sekali	PBK	Sedang	Tepat Waktu
Cepat	Sangat Memuaskan	Baik Sekali	PBK	Tinggi	Terlambat
Cepat	Sangat Memuaskan	Baik Sekali	PTIK	-(semua)	Tepat Waktu
Cepat	Sangat Memuaskan	Unggul	-(semua)	-(semua)	Tepat Waktu
Lama	-(semua)	-(semua)	-(semua)	-(semua)	Terlambat

Aturan-aturan ini membantu mengidentifikasi pola keterlambatan dan memperkirakan kemungkinan mahasiswa lulus sesuai jadwal berdasarkan Lama Masa Skripsi, Yudisium, Akreditasi, Program Studi, dan IPK. Informasi ini dapat dimanfaatkan oleh perguruan tinggi dalam merancang kebijakan akademik yang lebih efektif guna meningkatkan tingkat kelulusan mahasiswa.

Model kemudian diuji menggunakan data *testing* menggunakan *RapidMiner* untuk mengevaluasi akurasinya.



**Gambar 5.** Hasil Evaluasi Model

Performa model *Decision Tree* yang dibangun untuk memprediksi tingkat kelulusan mahasiswa menunjukkan akurasi sebesar 91,35%, yang mengindikasikan bahwa model mampu melakukan klasifikasi dengan tingkat kesalahan yang rendah. Dari *confusion matrix*, model berhasil mengklasifikasikan 61 mahasiswa lulus tepat waktu dan 34 mahasiswa terlambat dengan tingkat *precision* masing-masing 92,42% dan 89,47%. Selain itu, nilai *recall* menunjukkan bahwa model dapat mengidentifikasi mahasiswa lulus tepat waktu dengan 93,85% akurasi dan mahasiswa terlambat dengan 87,18% akurasi. Hasil ini menunjukkan bahwa model memiliki kinerja yang baik dalam memprediksi tingkat kelulusan mahasiswa berdasarkan pola data yang digunakan.

#### F. Analisis Hasil dan Pembahasan

Hasil perhitungan manual dan implementasi model *Decision Tree* C5.0 dalam RapidMiner menunjukkan kesesuaian tinggi, menegaskan validitas metode yang digunakan. Hasil Model telah diuji menggunakan 104 data *testing*.

**Tabel 6.** Hasil Prediksi Kelulusan Mahasiswa

No	NIM	PRODI	Label	Prediction	Confidence (Tepat Waktu)	Confidence (Terlambat)	Hasil Prediksi
1	21***38	PAI	Terlambat	Tepat Waktu	0,977	0,023	Salah
2	26***33	PBK	Tepat Waktu	Terlambat	0,444	0,556	Salah
3	24***47	PMTK	Terlambat	Terlambat	0,257	0,743	Benar
4	21***53	PAI	Terlambat	Terlambat	0,000	1,000	Benar
5	22***72	PBA	Tepat Waktu	Tepat Waktu	1,000	0,000	Benar
...	...	...	...	...	...	...	...
101	25***70	PTIK	Tepat Waktu	Tepat Waktu	1,000	0,000	Benar
102	25***96	PTIK	Tepat Waktu	Tepat Waktu	1,000	0,000	Benar
103	26***30	PBK	Tepat Waktu	Tepat Waktu	1,000	0,000	Benar
104	21***47	PAI	Tepat Waktu	Tepat Waktu	0,977	0,023	Benar

Analisis pohon keputusan menunjukkan bahwa lama pengerjaan skripsi merupakan faktor utama penentu kelulusan mahasiswa, diikuti yudisium dan akreditasi program studi. Mahasiswa dengan predikat "Dengan Pujian" hampir selalu lulus tepat waktu, sementara yang meraih "Sangat Memuaskan" masih dipengaruhi oleh akreditasi. Program studi berakreditasi unggul cenderung menghasilkan lulusan tepat waktu, sedangkan pada akreditasi lebih rendah, faktor IPK dan jenis program studi menjadi penentu. Misalnya, PBI cenderung meluluskan mahasiswa terlambat, sedangkan PMTK dan PBK dipengaruhi oleh IPK, di mana IPK rendah meningkatkan risiko keterlambatan. Meski demikian, IPK tinggi tidak selalu menjamin kelulusan tepat waktu karena faktor lain seperti cuti akademik atau perbaikan nilai. Atribut "Total SKS" tidak masuk ke dalam model, kemungkinan telah tercakup dalam akreditasi program studi.

Model pohon keputusan yang digunakan menghasilkan akurasi sebesar 91,35%, dengan precision 92,42% untuk kelas "Tepat Waktu" dan 89,47% untuk "Terlambat", serta recall masing-masing 93,85% dan 87,18%. Temuan ini dapat dimanfaatkan institusi untuk merancang kebijakan akademik yang lebih tepat, seperti mempercepat penyampaian mata kuliah metodologi penelitian, meningkatkan kualitas bimbingan skripsi, memperkuat standar akreditasi, dan memantau IPK mahasiswa sebagai indikator dini keterlambatan. Dengan strategi intervensi yang lebih cepat dan terarah, peluang mahasiswa lulus tepat waktu dapat ditingkatkan secara signifikan..

## KESIMPULAN

Model *Decision Tree* C5.0 menunjukkan akurasi 91,35%, membuktikan efektivitasnya dalam memprediksi kelulusan mahasiswa. Faktor utama yang berpengaruh adalah lama masa skripsi, diikuti oleh yudisium, akreditasi program studi, program studi, dan IPK. Mahasiswa yang menyelesaikan skripsi lebih cepat cenderung lulus tepat waktu.

1. Hasil penelitian ini dapat menjadi dasar bagi institusi dalam merancang kebijakan akademik yang lebih efektif, terutama dalam penyusunan kurikulum dan sistem pembelajaran.
2. Institusi dapat meningkatkan kualitas bimbingan akademik, memperkuat sistem evaluasi skripsi, serta menyesuaikan kebijakan berdasarkan faktor-faktor utama yang berpengaruh terhadap kelulusan mahasiswa.
3. Pendekatan berbasis data, institusi dapat merancang strategi yang lebih tepat untuk mempercepat kelulusan mahasiswa serta mengurangi risiko keterlambatan studi.

## SARAN

Penelitian selanjutnya dapat membandingkan algoritma lain untuk meningkatkan akurasi prediksi serta menambahkan variabel baru, seperti faktor sosial-ekonomi. Selain itu, analisis longitudinal dan implementasi model dalam sistem akademik dapat menguji efektivitasnya dalam mendukung kebijakan pendidikan

## DAFTAR PUSTAKA

- Anggriani, G., Amiruddin, & Zulfrianto. (2024). Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode Decision Tree. *Jurnal Ilmiah Ilmu Komputer*, 3(1), 31–36.
- Do Nascimento, D. V., Sousa, R. T., Costa Silva, D. F., Do Prado Pagotto, D., Coelho, C. J., & Galvao Filho, A. R. (2023). Live Birth Forecasting in Brazillian Health Regions with Tree-based Machine Learning Models. *Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2023-June*, 85–90. <https://doi.org/10.1109/CBMS58004.2023.00197>
- Fan, W., Liu, K., Liu, H., Ge, Y., Xiong, H., & Fu, Y. (2024). Interactive Reinforcement Learning for Feature Selection with Decision Tree in the Loop. *IEEE Transactions on Knowledge and Data Engineering*, 35(2), 1624–1636. <https://doi.org/10.1109/TKDE.2021.3102120>
- Firman Akbar, & Rahmaddeni. (2024). Komparasi Algoritma Machine Learning Untuk Memprediksi Penyakit Alzheimer. *Jurnal Komputer Terapan*, 8(2), 236–245. <https://doi.org/10.35143/jkt.v8i2.5713>
- Madhyastha, M., Budavari, T., Braverman, V., Vogelstein, J., & Burns, R. (2024). T-Rex (Tree-Rectangles): Reformulating Decision Tree Traversal as Hyperrectangle Enclosure. *Proceedings - International Conference on Data Engineering*, 1792–1804. <https://doi.org/10.1109/ICDE60146.2024.00145>
- Menteri Pendidikan, kebudayaan, R. dan teknologi R. I. (2023). Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 53 Tahun 2023 Tentang Penjaminan Mutu Pendidikan Tinggi. In [jdih.kemdikbud.go.id](http://jdih.kemdikbud.go.id).
- Mukrimaa, S. S., Nurdyansyah, Fahyuni, E. F., YULIA CITRA, A., Schulz, N. D., د، غسان، Taniredja, T., Faridli, E. M., & Harmianto, S. (2024). Machine Learning Teori, Studi Kasus dan Implementasi Menggunakan Python. In *Jurnal Penelitian Pendidikan Guru Sekolah Dasar* (Vol. 6, Issue August, p. 128).
- Mundok, G. A., Amiruddin, & Zulfrianto. (2024). Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode Decision Tree. *Jurnal Ilmiah Ilmu Komputer*, 3(1), 31–36.
- Presiden Revublik Indonesia. (2022). *Peraturan Presiden Republik Indonesia Nomor 85 Tahun 2022 Tentang Universitas Islam Negeri Sjech M.Djamil Djambek Bukittinggi*.
- Qisthiano, M. R., Prayesy, P. A., & Ruswita, I. (2023). Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa. *G-Tech: Jurnal Teknologi Terapan*, 7(1), 21–28. <https://doi.org/10.33379/gtech.v7i1.1850>
- Saputra, I., & Kristiyanti, D. A. (2024). *Machine Learning*. Informatika Bandung.
- Suyanto, R. V. A., Eduard Rusdianto, & Ernawati. (2024). Penerapan Algoritma Decision Tree C4.5 dan Metode AdaBoost Untuk Prediksi Kelulusan Mahasiswa. *Jurnal Informatika Atma Jogja*, 5(1), 75–86. <https://doi.org/10.24002/jiaj.v5i1.8646>
- Umar, N., Susilo, A., Trianto, & Bagus, R. (2024). Prediksi Tingkat Kelulusan Mahasiswa S1 Universitas an Nuur Dengan Metode DECISION TREE C4.5. *Julia: Jurnal Ilmu Komputer An Nuur*, 4(Rohmawan 2018), 21–29.
- Zebua, F. J., Br Manalu, R. P., & Nababan, M. N. K. (2024). Prediksi Kelulusan Mahasiswa Menggunakan Perbandingan Algoritma C5.0 Dengan Regression Linear. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 4(2), 230. <https://doi.org/10.37600/tekinkom.v4i2.400>
- Zhong, G., Xiang, K., & Mo, Q. (2024). Weather Forecast Model Based on Decision Tree and Neural Network. *Proceedings - 2024 4th Asia-Pacific Conference on Communications Technology and Computer Science, ACCTCS 2024*, 509–512. <https://doi.org/10.1109/ACCTCS61748.2024.00096>