Vol. 5 No. 1 – Tahun 2026
Indonesian Journal Computer Science
ISSN: 2829-3819

Penerapan Kerangka CRISP-DM dalam Evaluasi *Performa Logistic Regression* dan *Decision Tree* untuk Prediksi Kelulusan Mahasiswa

Tifanny Nabarian[1], Imelda Wahyuni[2], Salman El Farisi[3]
Sekolah Tinggi Teknologi Terpadu Nurul Fikri[1,2,3]
nabarian@nurulfikri.ac.id[1], imel22042ti@student.nurulfikri.ac.id[2], salman@nurulfikri.ac.id[3]

Diterima	Direvisi	Disetujui
(17-03-2026)	(05-04-2026)	(15-04-2026)

Abstrak - Ketepatan waktu kelulusan mahasiswa sering digunakan sebagai salah satu indikator dalam menilai efektivitas penyelenggaraan pendidikan di perguruan tinggi. Pemanfaatan data akademik melalui pendekatan *machine learning* dapat membantu mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan secara lebih terukur. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Logistic Regression* dan *Decision Tree* dalam memprediksi kelulusan mahasiswa berdasarkan faktor akademik. Dataset yang digunakan terdiri dari 351 mahasiswa Program Studi Teknik Informatika di STT Terpadu Nurul Fikri, angkatan 2021, dengan variabel input Indeks Prestasi Kumulatif (IPK) serta Indeks Prestasi Semester (IPS) untuk setiap semester. Penelitian ini menggunakan metodologi CRISP-DM dengan pembagian data pelatihan serta pengujian 80:20. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy, precision, recall,* serta *F1-score.* Hasil pengujian menunjukkan bahwa *Logistic Regression* memperoleh *accuracy* 92,96%, *precision* 90,91%, *recall* 97,56%, dan *F1-score* 94,12%, sedangkan *Decision Tree* memperoleh *accuracy* 88,73%, *precision* 88,37%, *recall* 92,68%, serta *F1-score* 90,48%. Hasil tersebut menunjukkan bahwa *Logistic Regression* menunjukkan performa yang lebih optimal dan lebih sesuai dalam memodelkan hubungan antara variabel akademik dan ketepatan kelulusan mahasiswa. Sebagai kontribusi utama, penelitian ini mengintegrasikan model *Logistic Regression* terbaik ke dalam purwarupa sistem deteksi dini (*early warning system*) berbasis web menggunakan Streamlit. Pendekatan ini memberikan keunggulan praktis bagi pemangku kepentingan program studi untuk memitigasi risiko keterlambatan kelulusan secara *real-time* dan mendukung pengambilan keputusan intervensi akademik yang sepenuhnya digerakkan oleh data (*data-drive decision* making).

Kata Kunci: *Machine Learning*, *Logistic Regression*, *Decision Tree*, Prediksi Kelulusan Mahasiswa

Abstract *- Student graduation timeliness is often used as one of the indicators for evaluating the effectiveness of higher education management. The utilization of academic data through a machine learning approach can assist institutions in identifying students who are likely to experience delayed graduation in a more measurable manner. This study aims to compare the performance of the Logistic Regression and Decision Tree algorithms in predicting student graduation based on academic factors. The dataset used in this study consists of records from 351 students of the Informatics Engineering Program at STT Terpadu Nurul Fikri, cohort of 2021, with the input variables including the Cumulative Grade Point Average (GPA) and Semester Grade Point Average (SGPA) for each semester. This research adopts the CRISP-DM methodology with an 80:20 split for training and testing data. Model evaluation was conducted using a confusion matrix with performance metrics including accuracy, precision, recall, and F1-score. The results show that Logistic Regression achieved an accuracy of 92.96%, precision of 90.91%, recall of 97.56%, and an F1-score of 94.12%, while Decision Tree achieved an accuracy of 88.73%, precision of 88.37%, recall of 92.68%, and an F1-score of 90.48%. These results indicate that Logistic Regression performs optimally and is more appropriate in modeling the relationship between academic variables and student graduation accuracy. As a major contribution, this research embeds the superior Logistic Regression model into a web-based early warning system prototype developed with Streamlit. This approach delivers practical benefits for academic stakeholders by enabling real-time mitigation of delayed graduation risks and fostering a fully data-driven decision-making process for academic interventions.*

Keywords: Machine Learning, Logistic Regression, Decision Tree, Student Graduation Prediction

I. PENDAHULUAN

Pendidikan tinggi memiliki peran penting dalam menghasilkan sumber daya manusia yang kompeten dan berdaya saing. Salah satu indikator keberhasilan perguruan tinggi ialah kemampuan mahasiswa menyelesaikan studi sesuai dengan durasi yang sudah ditetapkan. Ketepatan waktu kelulusan mencerminkan efektivitas proses pembelajaran sekaligus kualitas pengelolaan akademik. Sebaliknya, keterlambatan kelulusan dapat memengaruhi evaluasi program studi serta perencanaan akademik institusi secara keseluruhan (Alfaris, 2022). Oleh karena itu, identifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan menjadi aspek penting dalam pengelolaan pendidikan tinggi.

Pemanfaatan teknik *data mining* dalam bidang pendidikan memungkinkan institusi menganalisis data akademik historis untuk membangun model prediksi kelulusan. Teknik klasifikasi telah banyak digunakan untuk memprediksi status kelulusan mahasiswa berdasarkan capaian akademik (Adnyana, 2021). Selain itu, performa model sangat dipengaruhi oleh karakteristik dataset yang digunakan, sehingga pemilihan algoritma yang tepat menjadi faktor penting dalam menghasilkan prediksi yang akurat (Azis, 2024). Model klasifikasi juga berpotensi digunakan sebagai bagian dari sistem pendukung keputusan untuk membantu penyusunan kebijakan akademik berbasis data (Arifin et al., 2024).

Berbagai penelitian telah mengkaji prediksi kelulusan mahasiswa menggunakan pendekatan *machine learning* dengan memanfaatkan data akademik sebagai faktor utama (Pelima et al., 2024; Yuliaty & Pawitan, 2025). *Logistic Regression* banyak digunakan karena mampu memodelkan probabilitas secara langsung serta memberikan interpretasi yang jelas terhadap variabel (Nugroho et al., 2023; Syahranita et al., 2023). Penelitian lain juga menunjukkan bahwa *Logistic Regression* mampu menghasilkan prediksi yang akurat serta memberikan gambaran kuantitatif terhadap peluang kelulusan mahasiswa (Nurmalitasari & Purwanto, 2022).

Selain itu, *Decision Tree* juga banyak digunakan dalam prediksi kelulusan mahasiswa karena mampu menghasilkan aturan keputusan yang mudah dipahami serta mengidentifikasi faktor penting dalam klasifikasi (Al Faruq et al., 2023; Dengen et al., 2020; Setiono & Purwanto, 2025). Beberapa penelitian juga menunjukkan bahwa pengembangan metode *Decision Tree*, seperti C4.5, mampu meningkatkan performa klasifikasi serta memberikan hasil yang mudah diinterpretasikan dalam konteks akademik (Setiono & Purwanto, 2025; Suyanto et al., 2024; Yatimah, 2021).

Meskipun berbagai penelitian telah membahas penggunaan *Logistic Regression* dan *Decision Tree* dalam prediksi kelulusan mahasiswa, sebagian besar dilakukan secara terpisah atau dibandingkan dengan algoritma lain *seperti K-Nearest Neighbor* dan *Random Forest* (Mubarak et al., 2024). Selain itu, perbandingan kedua algoritma pada dataset akademik yang sama masih terbatas dan umumnya hanya berfokus pada evaluasi performa tanpa implementasi dalam sistem praktis. Oleh karena itu, diperlukan penelitian yang tidak hanya membandingkan kinerja algoritma pada kondisi data yang identik, tetapi juga mengimplementasikannya dalam sistem prediksi untuk mendukung monitoring kelulusan mahasiswa.

Berdasarkan permasalahan yang telah diuraikan, penelitian ini melakukan analisis perbandingan antara algoritma *Logistic Regression* dan *Decision Tree* pada dataset akademik yang sama dengan menggunakan variabel IPK dan IPS tiap semester sebagai faktor prediktor kelulusan mahasiswa. Selain mengevaluasi performa model menggunakan metrik klasifikasi, penelitian ini juga mengimplementasikan model terbaik ke dalam sistem prediksi berbasis web menggunakan *framework* Streamlit untuk mendukung proses monitoring dan analisis kelulusan mahasiswa di lingkungan perguruan tinggi.

Penelitian ini memberikan kontribusi sebagai berikut:

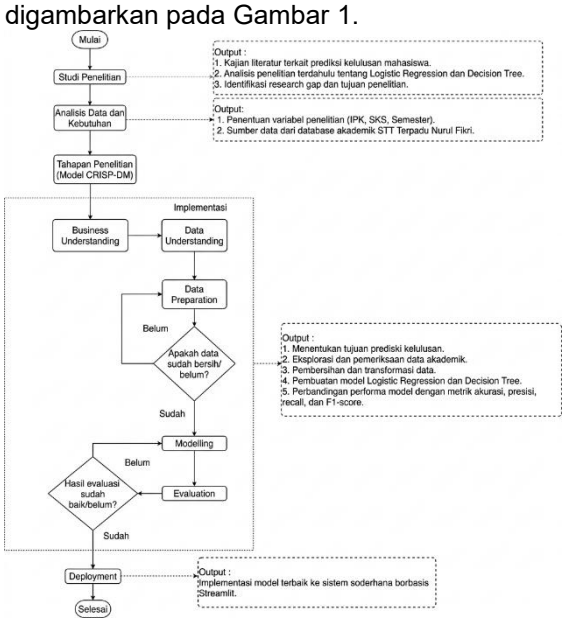
1. Melakukan analisis komparatif antara algoritma *Logistic Regression* dan *Decision Tree* pada dataset akademik yang sama untuk memperoleh perbandingan performa yang lebih objektif.
2. Mengidentifikasi algoritma yang lebih sesuai dalam memodelkan hubungan antara variabel akademik dan ketepatan kelulusan mahasiswa
3. Mengimplementasikan model terbaik ke dalam sistem prediksi berbasis web sebagai bentuk penerapan praktis dalam mendukung pengambilan keputusan akademik.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Logistic Regression* dan *Decision Tree* dalam memprediksi kelulusan mahasiswa berdasarkan data akademik. Hasil penelitian ini diharapkan dapat memberikan dasar empiris dalam menentukan algoritma

yang lebih optimal untuk diterapkan dalam sistem prediksi kelulusan mahasiswa, serta mendukung pengembangan sistem pendukung keputusan dalam monitoring perkembangan studi mahasiswa secara lebih sistematis.

II. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode komparatif dengan pendekatan kuantitatif untuk membandingkan kinerja algoritma *Logistic Regression* dan *Decision Tree* dalam memprediksi kelulusan mahasiswa berdasarkan variabel akademik. Tahapan penelitian dilakukan secara terstruktur dan digambarkan pada Gambar 1.



Sumber : Olahan Peneliti (2026)

Gambar 1. *Flowchart* Alur Penelitian

1. Studi Literatur

Tahap awal penelitian dimulai dengan studi penelitian untuk memahami teori, konsep, dan hasil penelitian sebelumnya yang berkaitan dengan prediksi kelulusan mahasiswa. Kajian literatur mencakup pembahasan mengenai faktor akademik yang memengaruhi kelulusan serta penerapan algoritma klasifikasi seperti *Logistic Regression* dan *Decision Tree* dalam konteks pendidikan tinggi (Anugrawati et al., 2023; Dengen et al., 2020; Yatimah, 2021). Hasil studi literatur digunakan sebagai dasar dalam menentukan variabel penelitian serta metode analisis yang digunakan.

2. Analisis Data dan Kebutuhan

Data yang digunakan dalam penelitian ini merupakan data akademik dari 351 mahasiswa Program Studi Teknik Informatika di STT Terpadu Nurul Fikri angkatan 2021, yang diperoleh melalui sistem akademik internal institusi. Data telah diolah sedemikian rupa untuk melindungi kerahasiaan mahasiswa. Variabel yang digunakan dalam penelitian ini meliputi IPK dan IPS pada setiap semester sebagai variabel prediktor, sedangkan status kelulusan berfungsi sebagai variabel target. Status kelulusan dibedakan menjadi dua kategori yaitu lulus tepat waktu serta tidak lulus tepat waktu, yang ditentukan berdasarkan durasi studi maksimum selama delapan semester.

3. Penerapan Metodologi CRISP-DM

Proses pengolahan data mengacu pada metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang tersusun dari enam tahapan penting, yakni *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,* serta *Deployment* (Rianti et al., 2023).

a. *Business Understanding*
Tahapan ini bertujuan untuk memahami permasalahan keterlambatan kelulusan mahasiswa dan menentukan tujuan penelitian, yakni membangun model klasifikasi yang dapat memprediksi status kelulusan mahasiswa berdasarkan faktor akademik.

b. *Data Understanding*
Tahap ini dilakukan dengan menganalisis karakteristik dataset akademik mahasiswa yang digunakan dalam penelitian. Proses ini mencakup identifikasi atribut yang tersedia seperti IPK tiap semester, IPS, serta distribusi status kelulusan mahasiswa.

c. *Data Preparation*
Pada tahap *data preparation*, dilakukan pembersihan data untuk memastikan tidak terdapat *missing values* maupun duplikasi. Variabel target ditransformasikan menjadi klasifikasi biner menggunakan teknik encoding, dengan nilai 1 untuk "Lulus Tepat Waktu" dan 0 untuk "Tidak Lulus Tepat Waktu". Fitur prediktor yang digunakan meliputi nilai IPS dari semester 1 hingga 8 dan IPK akhir. Terkait tahap *praprocess*, proses normalisasi atau *scaling* tidak dilakukan. Hal ini didasarkan pada karakteristik natural variabel prediktor (IPS dan IPK) yang seluruhnya telah berada pada rentang skala ukur yang seragam (0,00 hingga 4,00). Keseragaman ini mengeliminasi risiko dominasi fitur tertentu pada saat pelatihan model. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20.

d. *Modeling*
Tahap pemodelan mengeksekusi algoritma *Logistic Regression* dan *Decision Tree* untuk mengklasifikasikan status kelulusan. Untuk mengoptimalkan kinerja dan stabilitas, dilakukan pengaturan *hyperparameter* pada masing-masing model. Pada *Logistic Regression*, parameter *max_iter* diatur sebesar 1000 guna memastikan fungsi optimasi memiliki iterasi yang cukup untuk mencapai konvergensi yang optimal. Pada *Decision Tree*, pengaturan *hyperparameter* difokuskan pada pembatasan struktur pohon melalui parameter *max_depth* = 5. Pembatasan ini berfungsi sebagai strategi *pruning* awal untuk mengendalikan kompleksitas model dan secara efektif memitigasi risiko *overfitting*. Selain itu, parameter *random_state* ditetapkan pada kedua algoritma untuk memastikan hasil pemodelan tetap konsisten dan dapat direplikasi (*reproducible*).

e. *Evaluation*
Pada tahap ini, evaluasi kinerja model dilakukan menggunakan metrik evaluasi berupa *accuracy, precision, recall,* dan *F1-score* untuk menentukan tingkat akurasi klasifikasi yang dihasilkan oleh model. Selain menggunakan pembagian data pelatihan dan pengujian, penelitian ini juga menerapkan teknik *5-fold cross-validation* untuk mengevaluasi kestabilan performa model. Teknik ini digunakan untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik terhadap variasi data serta tidak hanya bergantung pada satu pembagian data.

f. *Deployment*
Tahap *deployment* dilakukan dengan mengimplementasikan model terbaik ke dalam sistem prediksi sederhana berbasis web menggunakan *framework* Streamlit sebagai simulasi penerapan model dalam lingkungan akademik.

4. Pembangunan Model Klasifikasi

Pada tahap pemodelan dilakukan pembangunan dua model klasifikasi menggunakan algoritma *Logistic Regression* serta *Decision Tree*. *Logistic Regression* digunakan untuk memodelkan probabilitas kelulusan mahasiswa berdasarkan variabel akademik yang tersedia (Nurmalitasari & Purwanto, 2022; Sihotang, 2023). Sementara itu, *Decision Tree* digunakan untuk membentuk aturan keputusan berdasarkan atribut akademik yang paling informatif dalam memisahkan kelas kelulusan (Dengen et al., 2020; Suyanto et al., 2024). Kedua model dibangun menggunakan dataset pelatihan yang sama sehingga proses perbandingan performa dapat dilakukan secara adil dan konsisten pada kondisi data yang identik.

5. Evaluasi Model

Evaluasi performa model dilakukan menggunakan *confusion matrix* untuk mengukur tingkat ketepatan klasifikasi yang dihasilkan oleh model. *Confusion matrix* merupakan tabel evaluasi yang menunjukkan perbandingan antara hasil prediksi model dengan kondisi sebenarnya sehingga dapat diketahui jumlah prediksi yang benar maupun salah untuk setiap kelas (Mubarak et al., 2024). Struktur *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

Actual/Predicted	*Positif*	*Negatif*
Positif	*True Positive (TP)*	*False Negative (FN)*
Negatif	*False Positive (FP)*	*True Negative (TN)*

Sumber : Hasil Penelitian (2026)

Berdasarkan nilai pada *confusion matrix,* beberapa metrik evaluasi digunakan untuk mengukur kinerja model klasifikasi, yaitu *accuracy, precision, recall*, dan *F1-score* (Ridwan et al., 2024; Wulan Yekti Rahayu et al., 2025).

a. *Accuracy*
Accuracy merupakan metrik yang digunakan untuk mengukur ketepatan model melalui perbandingan jumlah prediksi benar terhadap total data (Ridwan et al., 2024).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

b. *Precision*
Precision mengukur ketepatan prediksi positif *melalui* perbandingan prediksi positif yang benar terhadap seluruh prediksi positif (Rahayu et al., 2025).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall*
Recall mengukur kemampuan model dalam mendeteksi data positif yang berhasil dikenali (Wulan Yekti Rahayu et al., 2025).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. *F1-score*
F1-score mengukur kinerja model secara seimbang melalui rata-rata harmonik precision dan recall, terutama pada data tidak seimbang (Mubarak et al., 2024; Ridwan et

al., 2024)

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Nilai dari metrik evaluasi tersebut kemudian digunakan untuk mengevaluasi perbedaan kinerja algoritma *Logistic Regression* dan *Decision Tree* dalam proses prediksi kelulusan mahasiswa.

6. Implementasi Model Terbaik

Model dengan performa terbaik selanjutnya diimplementasikan dalam sistem sederhana berbasis web menggunakan *framework* Streamlit. Sistem ini dirancang untuk mensimulasikan proses prediksi kelulusan mahasiswa berdasarkan input nilai IPK dan IPS. *Framework* Streamlit memungkinkan integrasi model *machine learning* dengan antarmuka web secara sederhana sehingga pengguna dapat melakukan prediksi secara interaktif (Yatimah, 2021). Implementasi ini bertujuan untuk menunjukkan potensi penerapan model dalam mendukung proses monitoring dan pengambilan keputusan akademik secara praktis.

7. Tools dan Lingkungan Pengembangan

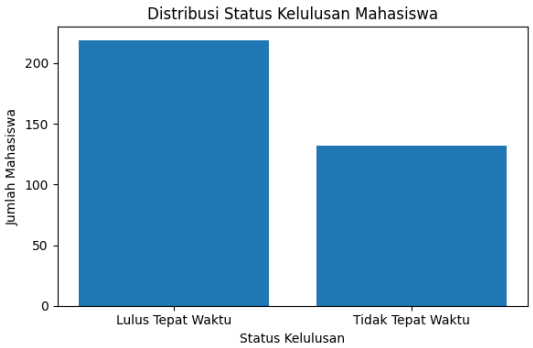
Implementasi model dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan pustaka Pandas dan NumPy untuk pengolahan data, Scikit-learn untuk pemodelan dan evaluasi *machine learning*, serta *Matplotlib* dan *Seaborn* untuk visualisasi data. Seluruh proses pengembangan dan eksperimen dilakukan menggunakan Google Colab guna memudahkan pengolahan data, pengujian model, serta memastikan efisiensi dan reproduksibilitas penelitian.

III. HASIL DAN PEMBAHASAN

1. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data akademik mahasiswa Program Studi Teknik Informatika di STT Terpadu Nurul Fikri angkatan 2021. Dataset terdiri dari 351 data mahasiswa yang memuat informasi capaian akademik selama masa studi. Variabel yang digunakan dalam proses pemodelan meliputi Indeks Prestasi Kumulatif (IPK) dan Indeks Prestasi Semester (IPS) dari semester pertama hingga semester kedelapan.

Atribut yang tidak memiliki pengaruh langsung terhadap proses prediksi seperti nama mahasiswa dan nomor induk mahasiswa tidak digunakan dalam proses pemodelan. Hal ini bertujuan agar model hanya memanfaatkan atribut akademik yang relevan dalam menentukan status kelulusan mahasiswa. Status kelulusan mahasiswa digunakan sebagai variabel target dalam penelitian ini. Mahasiswa diberikan kelulusan tepat waktu jika mereka menyelesaikan studi mereka dalam jangka waktu maksimal delapan semester dengan jumlah SKS yang telah ditentukan. Mahasiswa yang menyelesaikan studi melebihi batas waktu tersebut dikategorikan sebagai tidak lulus tepat waktu. Distribusi data kelulusan mahasiswa pada dataset penelitian ditunjukkan pada Gambar 2.



Sumber : Hasil Penelitian (2026)

Gambar 2. Distribusi Status Kelulusan Mahasiswa

Berdasarkan distribusi tersebut, terdapat 219 mahasiswa (62,39%) yang termasuk dalam kategori lulus tepat waktu dan 132 mahasiswa (37,61%) yang termasuk dalam kategori tidak lulus tepat waktu. Distribusi ini menunjukkan bahwa sebagian besar mahasiswa pada dataset mampu menyelesaikan studi sesuai dengan durasi yang telah ditetapkan.

2. Hasil Pengujian *Logistic Regression*

Model *Logistic Regression* dibangun menggunakan data pelatihan yang didapatkan melalui pemisahan dataset dengan perbandingan 80% untuk data pelatihan serta 20% untuk data pengujian. Model tersebut kemudian diuji menggunakan data pengujian guna mengidentifikasi kemampuan model dalam memprediksi status kelulusan mahasiswa.
Hasil pengujian model *Logistic Regression* ditunjukkan pada Tabel 2.

Tabel 2. *Confusion Matrix Logistic Regression*

Actual/ Predicted	Lulus Tepat Waktu	Tidak Lulus Tepat Waktu
Lulus Tepat Waktu	40	1
Tidak Lulus Tepat Waktu	4	26

Sumber: Hasil Penelitian (2026)

Berdasarkan hasil *confusion matrix* pada Tabel 2, model *Logistic Regression* menunjukkan kemampuan yang cukup baik dalam mengklasifikasikan mayoritas data secara tepat. Model tersebut berhasil mengklasifikasikan dengan benar sebanyak 40 mahasiswa yang menyelesaikan studi tepat waktu serta 26 mahasiswa yang tidak mampu lulus sesuai waktu yang ditentukan. Namun, masih ada beberapa kesalahan klasifikasi yaitu terdapat 4 mahasiswa yang sebenarnya tidak lulus tepat waktu tetapi diprediksi oleh model sebagai lulus tepat waktu, serta 1 mahasiswa yang sebenarnya lulus tepat waktu namun diprediksi sebagai tidak lulus tepat waktu.

Untuk memperoleh gambaran yang lebih menyeluruh mengenai performa model, dilakukan evaluasi tambahan dengan memanfaatkan sejumlah indikator pengukuran kinerja klasifikasi, yaitu *accuracy, precision, recall*, serta *F1-score*. Hasil pengukuran kinerja model *Logistic Regression* tersebut disajikan secara sistematis pada Tabel 3.

Tabel 3. Hasil Evaluasi *Logistic Regression*

Accuracy	*Precision*	*Recall*	*F1-score*
0.9296	0.9091	0.9756	0.9412

Sumber: Hasil Penelitian (2026)

Berdasarkan hasil evaluasi pada Tabel 3, model *Logistic Regression* memperoleh tingkat *accuracy* sebesar 92,96%, yang menunjukkan bahwa model ini mampu mengklasifikasikan sebagian besar data pengujian dengan tepat. Tingkat *precision* sebesar 90,91% menunjukkan bahwa mayoritas mahasiswa yang diprediksi oleh model untuk lulus tepat waktu adalah akurat. Selain itu, nilai *recall* yang mencapai 97,56% menunjukkan bahwa model tersebut dapat mengenali hampir seluruh mahasiswa yang benar-benar lulus tepat waktu. Nilai F1-score sebesar 94,12% menunjukkan adanya keseimbangan yang baik antara *precision* dan *recall* dalam proses klasifikasi.

3. Hasil Pengujian *Decision Tree*

Selain *Logistic Regression,* penelitian ini juga menggunakan algoritma *Decision Tree* untuk membangun model klasifikasi kelulusan mahasiswa. Model dilatih menggunakan dataset yang sama sehingga proses perbandingan performa dapat dilakukan secara konsisten.
Hasil pengujian model *Decision Tree* ditunjukkan pada Tabel 4.

Tabel 4. *Confusion Matrix Decision Tree*

Actual/ Predicted	Lulus Tepat Waktu	Tidak Lulus Tepat Waktu
Lulus Tepat Waktu	25	5
Tidak Lulus Tepat Waktu	3	38

Sumber: Hasil Penelitian (2026)

Berdasarkan *confusion matrix* pada Tabel 4, model *Decision Tree* mampu mengklasifikasikan sebagian besar data dengan benar. Model tersebut berhasil memprediksi 25 mahasiswa yang lulus tepat waktu dan 38 mahasiswa yang tidak lulus tepat waktu. Namun, masih ada beberapa kesalahan klasifikasi yaitu 5 mahasiswa yang lulus tepat waktu diprediksi tidak lulus tepat waktu, dan 3 mahasiswa yang tidak lulus tepat waktu diprediksi lulus tepat waktu.

Hasil evaluasi kinerja model *Decision Tree* ditunjukkan pada Tabel 5.

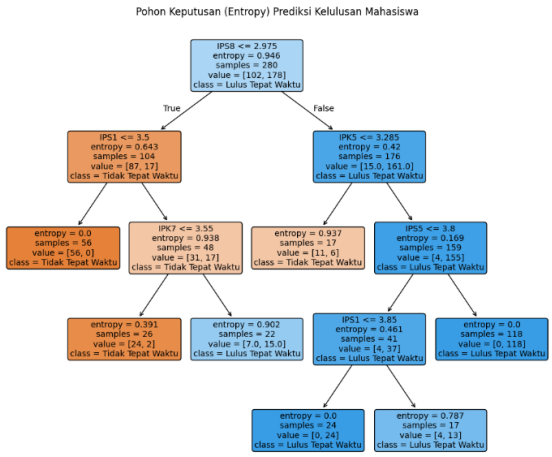
Tabel 5. Hasil Evaluasi *Decision Tree*

Accuracy	*Precision*	*Recall*	*F1-score*
0.8873	0.8837	0.9268	0.9048

Sumber: Hasil Penelitian (2026)

Berdasarkan hasil evaluasi pada Tabel 5, model *Decision Tree* memperoleh nilai *accuracy* 88,73%, yang menunjukkan bahwa model ini dapat memprediksi sebagian besar data uji dengan baik. Nilai *precision* sebesar 88,37% menunjukkan tingkat ketepatan prediksi *positif* dari model ini. Nilai *recall* yang mencapai 92,68% menunjukkan bahwa model tersebut berhasil mengidentifikasi mayoritas mahasiswa yang benar-benar lulus tepat waktu. Nilai *F1-score* yang mencapai 90,48% menunjukkan adanya keseimbangan yang baik antara *precision* dan *recall* dalam proses klasifikasi.

Selain evaluasi kuantitatif, interpretasi model *Decision Tree* juga dapat dilihat melalui visualisasi struktur pohon keputusan. Visualisasi tersebut menunjukkan bagaimana algoritma membentuk aturan klasifikasi berdasarkan variabel akademik yang digunakan dalam penelitian. Struktur pohon keputusan hasil pelatihan model ditampilkan pada Gambar 3.



Sumber : Hasil Penelitian (2026)

Gambar 3. Pohon Keputusan *Decision Tree*

Berdasarkan visualisasi pada Gambar 3, atribut IPS semester 8 (IPS8) menjadi node akar yang pertama kali digunakan dalam proses pemisahan data. Dalam *algoritma Decision Tree*, node akar dipilih berdasarkan atribut yang memiliki kemampuan terbaik dalam memisahkan kelas data, yang diukur menggunakan nilai entropy atau information gain. Pemilihan IPS8 sebagai node akar menunjukkan bahwa variabel tersebut memiliki kontribusi paling besar dalam membedakan mahasiswa yang lulus tepat waktu serta tidak lulus tepat waktu pada dataset penelitian.

Hal ini menunjukkan bahwa nilai IPS pada semester akhir mempunyai dampak yang signifikan pada prediksi status kelulusan mahasiswa. Mahasiswa dengan nilai IPS8 di atas batas tertentu cenderung diklasifikasikan sebagai lulus tepat waktu, sedangkan mahasiswa dengan nilai IPS8 yang lebih rendah lebih banyak dikategorikan sebagai tidak lulus tepat waktu. Temuan ini juga menunjukkan bahwa IPS semester 8 dapat menjadi indikator penting dalam menentukan kemungkinan kelulusan mahasiswa, karena semester akhir merupakan fase penentuan penyelesaian studi seperti penyelesaian tugas akhir maupun pemenuhan beban SKS. Dengan demikian, capaian akademik pada semester ini dapat dianggap sebagai titik kritis *(bottleneck)* yang memengaruhi keberhasilan mahasiswa dalam menyelesaikan studi tepat waktu.

Pada tingkat percabangan berikutnya, model menggunakan variabel akademik lain seperti IPK dan IPS pada semester sebelumnya untuk memperjelas proses klasifikasi. Setiap node pada pohon keputusan menampilkan nilai entropy, jumlah sampel, serta distribusi kelas yang digunakan untuk menentukan kategori mayoritas pada setiap cabang. Nilai entropy yang semakin kecil menunjukkan bahwa data pada node tersebut semakin homogen sehingga keputusan klasifikasi yang dihasilkan menjadi lebih pasti. Hasil visualisasi pohon keputusan ini menunjukkan bahwa informasi capaian akademik mahasiswa, khususnya pada semester akhir, dapat dimanfaatkan sebagai indikator awal *(early warning)* dalam memantau potensi keterlambatan kelulusan mahasiswa.

4. Perbandingan Kinerja Algoritma

Setelah dilakukan pengujian menggunakan algoritma *Logistic Regression* dan *Decision Tree*, tahap selanjutnya adalah membandingkan kinerja kedua model berdasarkan metrik evaluasi klasifikasi. Evaluasi dilakukan dengan menggunakan nilai *accuracy, precision, recall,* dan *F1-score* yang dihitung dari hasil *confusion matrix* pada data uji.

Tabel 6. Hasil Evaluasi Model *Logistic Regression*

Metric	Nilai
Accuracy	0.9296
Precision	0.9091
Recall	0.9756
F1-score	0.9412

Sumber: Hasil Penelitian (2026)

Tabel 7. Hasil Evaluasi Model *Decision Tree*

Metric	Nilai
Accuracy	0.8873
Precision	0.8837
Recall	0.9268
F1-score	0.9048

Sumber: Hasil Penelitian (2026)

Berdasarkan hasil evaluasi pada Tabel 6 dan Tabel 7, model *Logistic Regression* menunjukkan performa yang lebih baik dibandingkan *Decision Tree* pada seluruh metrik. *Logistic Regression* memperoleh *accuracy* 92,96%, *precision* 90,91%, *recall* 97,56%, dan *F1-score* 94,12%, sedangkan *Decision Tree* memperoleh *accuracy* 88,73%, *precision* 88,37%, *recall* 92,68%, dan *F1-score* 90,48%. Nilai *recall* yang lebih tinggi pada *Logistic Regression* menunjukkan kemampuannya dalam mengidentifikasi mahasiswa yang lulus tepat waktu secara lebih optimal dibandingkan *Decision Tree*.

Secara keseluruhan, hasil perbandingan menunjukkan bahwa *Logistic Regression* mempunyai kemampuan klasifikasi yang lebih

stabil pada dataset kajian ini. Hal ini menunjukkan bahwa hubungan antara variabel akademik seperti IPK dan IPS per semester dengan status kelulusan mahasiswa cenderung mengikuti pola yang relatif linier, sehingga lebih cocok untuk dimodelkan menggunakan *Logistic Regression* daripada metode berbasis pohon keputusan seperti *Decision Tree*.

Kondisi ini juga dapat dikaitkan dengan karakteristik proses pembelajaran pada program studi Teknik Informatika yang bersifat bertahap dan terstruktur antar semester. Capaian nilai IPK pada setiap semester serta nilai IPS sebagai akumulasi performa akademik mahasiswa umumnya mencerminkan konsistensi kemampuan mahasiswa dalam mengikuti perkuliahan dan menyelesaikan tugas akademik. Dengan demikian, pola peningkatan atau penurunan performa akademik mahasiswa dari semester ke semester cenderung membentuk hubungan yang relatif konsisten terhadap ketepatan waktu kelulusan mahasiswa.

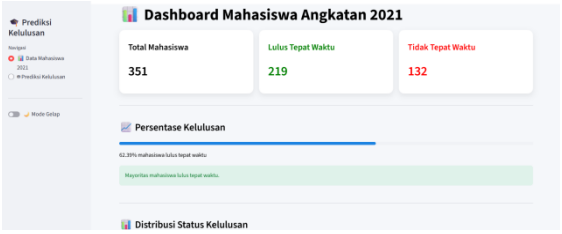
Sementara itu, implementasi pendekatan pembelajaran berbasis proyek *(project-based learning)* di beberapa mata kuliah inti juga menjadikan prestasi akademik mahasiswa sebagai indikator yang cukup representatif terhadap penguasaan kompetensi. Oleh karena itu, hubungan antara variabel akademik seperti IPK dan IPS dengan status kelulusan mahasiswa pada dataset penelitian ini lebih mudah dimodelkan menggunakan pendekatan *Logistic Regression* yang menangkap pola hubungan linier antar variabel.

Untuk memastikan konsistensi performa model, dilakukan evaluasi tambahan menggunakan teknik *5-fold cross-validation*. Hasil pengujian menunjukkan bahwa *Logistic Regression* memperoleh rata-rata akurasi sebesar 90,30% dengan standar deviasi sebesar 0,083, sedangkan *Decision Tree* memperoleh rata-rata akurasi sebesar 88,89% dengan standar deviasi sebesar 0,104. Nilai standar deviasi yang lebih rendah pada *Logistic Regression* menunjukkan bahwa model ini memiliki performa yang lebih stabil dan konsisten terhadap variasi data dibandingkan *Decision Tree.*

Perbedaan tingkat stabilitas ini menunjukkan bahwa *Logistic Regression* memiliki kemampuan generalisasi yang lebih baik dibandingkan *Decision Tree*. Model yang memiliki standar deviasi rendah cenderung tidak terlalu sensitif terhadap perubahan data pelatihan, sehingga hasil prediksi yang dihasilkan lebih dapat diandalkan ketika diterapkan pada data baru. Sebaliknya, nilai standar deviasi yang lebih tinggi pada *Decision Tree* mengindikasikan bahwa model tersebut lebih bergantung pada pembagian data tertentu. Kondisi ini mengarah pada potensi *overfitting* pada model *Decision Tree*, di mana model cenderung mempelajari pola yang terlalu spesifik terhadap data pelatihan. Hal ini dapat menyebabkan performa model menurun ketika dihadapkan pada data yang berbeda. Dengan demikian, selain memiliki nilai *accuracy* yang lebih tinggi, *Logistic Regression* juga menunjukkan performa yang lebih stabil dan kemampuan generalisasi yang lebih baik dalam memprediksi kelulusan mahasiswa pada dataset penelitian ini.

5. Implementasi Sistem Prediksi

Berdasarkan hasil evaluasi, *Logistic Regression* terpilih sebagai model terbaik karena memiliki performa lebih tinggi dibandingkan *Decision Tree* pada seluruh metrik. Model ini kemudian diimplementasikan dalam sistem prediksi berbasis web menggunakan Streamlit yang memungkinkan pengguna melakukan prediksi kelulusan secara interaktif.



Sumber: Hasil Penelitian (2026)

Gambar 4. Tampilan Halaman Utama Sistem Prediksi

Gambar 4 menampilkan halaman dashboard sistem prediksi kelulusan mahasiswa yang dikembangkan menggunakan *framework* Streamlit. Pada halaman ini ditampilkan ringkasan dataset penelitian yang meliputi jumlah mahasiswa serta distribusi status kelulusan mahasiswa. Selain itu, sistem juga menyajikan visualisasi data berupa grafik distribusi status kelulusan mahasiswa dan grafik rata-rata IPK berdasarkan kategori kelulusan. Dashboard juga menyediakan tabel data mahasiswa yang berisi informasi akademik seperti nilai IPK dan IPS per semester yang digunakan dalam proses pemodelan. Visualisasi dan tabel data tersebut memberikan gambaran umum mengenai karakteristik dataset yang dianalisis dalam penelitian ini.

Dalam sistem yang dikembangkan,

pengguna juga dapat melakukan prediksi dengan memasukkan data akademik mahasiswa melalui form input yang tersedia pada halaman prediksi. Data yang dimasukkan meliputi nilai IPK dan IPS dari semester pertama hingga semester kedelapan. Data tersebut kemudian diproses oleh model *Logistic Regression* yang telah dilatih pada tahap sebelumnya.

Sumber : Hasil Penelitian (2026)

Gambar 5. Hasil Prediksi Status Kelulusan

Gambar 5 menampilkan hasil prediksi status kelulusan mahasiswa berdasarkan data akademik yang dimasukkan ke dalam sistem. Setelah pengguna mengisi data pada form input, sistem akan memproses data tersebut menggunakan model *Logistic Regression* dan menampilkan hasil prediksi dalam bentuk kategori "Lulus Tepat Waktu" atau "Tidak Lulus Tepat Waktu". Selain itu, sistem juga menampilkan probabilitas prediksi yang dihasilkan oleh model.

Secara keseluruhan, implementasi sistem ini menunjukkan bahwa model *Logistic Regression* yang dihasilkan dalam penelitian dapat diterapkan dalam bentuk aplikasi berbasis web yang mampu membantu proses analisis dan prediksi kelulusan mahasiswa secara lebih sistematis.

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *Logistic Regression* serta *Decision Tree* dalam memprediksi kelulusan mahasiswa berdasarkan faktor akademik. Berdasarkan hasil pengujian, kedua algoritma mampu melakukan klasifikasi dengan baik menggunakan variabel akademik berupa IPK dan IPS pada tiap semester. Namun, berdasarkan hasil evaluasi menggunakan metrik *accuracy, precision, recall,* dan *F1-score,* algoritma *Logistic Regression* berkinerja lebih baik daripada *Decision Tree. Logistic Regression* mencapai *accuracy* 92,96%, *precision* 90,91%, *recall* 97,56%, dan *F1-score* 94,12%, sedangkan *Decision Tree* mencapai *accuracy* 88,73% dengan nilai *precision, recall,* dan *F1-score* yang sedikit lebih rendah. Hasil ini menunjukkan bahwa hubungan antara variabel akademik dan status kelulusan mahasiswa dalam dataset penelitian lebih tepat dimodelkan menggunakan pendekatan *Logistic Regression.* Model terbaik kemudian diimplementasikan dalam sebuah sistem prediksi sederhana berbasis Streamlit untuk menunjukkan aplikasi praktis dari temuan penelitian. Sistem yang dikembangkan mampu memprediksi status kelulusan mahasiswa berdasarkan data akademik yang dimasukkan pengguna. Dengan demikian, penelitian ini menunjukkan bahwa pemanfaatan data akademik melalui pendekatan *machine learning* dapat digunakan sebagai alat untuk mendukung proses evaluasi akademik serta membantu mengidentifikasi kemungkinan kelulusan mahasiswa secara lebih sistematis.

V. REFERENSI

Adnyana, I. M. B. (2021). Penerapan Teknik Klasifikasi untuk Prediksi Kelulusan Mahasiswa Berdasarkan Nilai Akademik. *JUTIK : Jurnal Teknologi Informasi dan Komputer*, *7*(3).

Al Faruq, U., Fauzi, M. A. N., Fatayasya, I., Daniati, E., & Ristyawan, A. (2023). Prediksi Data Kelulusan Mahasiswa Dengan Metode *Decision Tree* menggunakan Rapidminer. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, *7*, 2549–7952. https://proceeding.unpkediri.ac.id/index.php/inotek/

Alfaris, S. (2022). Faktor Penghambat Keterlambatan Penyelesaian Studi Mahasiswa Prodi Pendidikan Teknik Mesin FKIP Undana. *Haumenni Journal of Education*, *2*(2), 1–8.

Anugrawati, S. D., Nurhikma, Saputri, I. W., & Nurfadilah, K. (2023). Analisis Regresi Logistik Biner dalam Penentuan Faktor-Faktor yang Mempengaruhi Ketepatan Waktu Lulus Mahasiswa UIN Alauddin Makassar. *Journal of Mathematics: Theory and Applications*, *5*(1), 11–16. https://doi.org/10.31605/jomta.v5i1.2401

Arifin, M., Helmi, F., & Hikmawansyah, R. Bagus. (2024). Analisis Metode dan Algoritma dalam Sistem Pendukung Keputusan untuk Memprediksi Kelulusan. *Jurnal Advance Research Informatika*, *3*(1), 73–80. https://www.ejournalwiraraja.com/index.php/JARS

Azis, A. R. (2024). Analisis Komparasi Algoritma *Machine Learning* dalam Prediksi

Performa Akademik Mahasiswa: Literature Review. *Jurnal Ilmu Komputer dan Informatika (JIKI)*, *4*(2), 143–150. https://doi.org/10.54082/jiki.212

Dengen, C. N., Kusrini, & Luthfi, E. T. (2020). Implementasi *Decision Tree* Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Sisfotenika*, *10*(1), 1. https://doi.org/10.30700/jst.v10i1.484

Mubarak, R., Hanafi, M., & Sasongko, D. (2024). Komparasi Performa *Naive Bayes Gaussian* dan *K-NN* Untuk Prediksi Kelulusan Mahasiswa dengan CRISP-DM. *KLIK: Kajian Ilmiah Informatika dan Komputer*, *4*(6), 2982–2991. https://doi.org/10.30865/klik.v4i6.1924

Nugroho, D. R., Idris, N., Kurniawan, K., Fitrianah, D., Irwansyah, E., & Kusuma, G. P. (2023). *Logistic Regression and Random Forest Comparison in Predicting Students' Qualification Based on Students' Half-Semester Performance*. *International Conference on Information and Communication Technology, ICoICT*, *2023-August*, 214–219. https://doi.org/10.1109/ICoICT58202.2023.10262783

Nurmalitasari, & Purwanto, E. (2022). Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik. *Jurnal Derivat*, *9*(2).

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). *Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review*. *IEEE Access*, *1*(1), 99. https://doi.org/10.1109/ACCESS.2024.3361479

Rahayu, D. W. Y., Umam, K., & Handayani, M. R. (2025). *Performance of Machine Learning Algorithms on Imbalanced Sentiment Datasets Without Balancing Techniques. Journal of Applied Informatics and Computing (JAIC)*, *9*(3), 998–1005. http://jurnal.polibatam.ac.id/index.php/JAIC

Rianti, A., Majid, N. W. A., & Fauzi, A. (2023). CRISP-DM: Metodologi Proyek Data Science. *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*.

Ridwan, Hermaliani, E. H., & Ernawati, M. (2024). Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian. *Computer Science (CO-SCIENCE)*, *4*(1). http://jurnal.bsi.ac.id/index.php/co-science

Setiono, S. A., & Purwanto, E. (2025). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Decision Tree*. *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*.

Sihotang, S. F. (2023). Analisis Regresi Logistik Biner untuk Memprediksi Probabilitas Kelulusan Ujian Akhir Semester Mahasiswa yang Mengambil Mata Kuliah Matematika Farmasi. *Journal of Mathematics Education and Science*, *8*(2). https://jurnal.uisu.ac.id/index.php/mesuisu

Suyanto, R. V. A., Rusdianto, E., & Ernawati. (2024). Penerapan Algoritma *Decision Tree* C4.5 dan Metode AdaBoost Untuk Prediksi Kelulusan Mahasiswa. *Jurnal Informatika Atma Jogja*, *5*(1), 75–86.

Syahranita, R., Suhartono, & Zaman, S. (2023). Regresi Logistik Multinomial untuk Prediksi Kategori Kelulusan Mahasiswa. *Jurnal Informatika Sunan Kalijaga (JISKA)*, *8*(2), 102–111.

Yatimah, M. N. (2021). Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan *Decision Tree* C4.5. *JUMANJI*, *5*(2), 89–98.

Yuliaty, T., & Pawitan, G. (2025). *Developing a Predictive System for On-Time Graduation Using Logistic Regression. MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(4), 1253–1265. https://doi.org/10.57152/malcom.v5i4.2142