

## Perbandingan Kinerja Algoritma Decision Tree dan Random Forest dalam Prediksi Kelulusan Siswa

Nur Azizah Harahap<sup>1</sup>, Andika Syahdewa<sup>2</sup>, Fadlan Ihsan Lubis<sup>3</sup>, Hengki Gunawan<sup>4</sup>, Marsini Sibuea<sup>5</sup>,  
Darma Juang<sup>6</sup>, Muhammad Amin<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup> Magister Teknologi Informasi, Universitas Pembangunan Panca Budi

Email: <sup>1</sup>nazizahhrp@gmail.com, <sup>2</sup>andikasadewa0011@gmail.com, <sup>3</sup>fihsanlubis@gmail.com, <sup>4</sup>hengkigunawan1591@gmail.com,  
<sup>5</sup>sibueamarsini09@gmail.com, <sup>6</sup>darma.juang55@gmail.com, <sup>7</sup>mhdamin@dosen.pancabudi.ac.id

Email Penulis Korespondensi: [nazizahhrp@email.com](mailto:nazizahhrp@email.com)

### Abstrak

Kelulusan siswa merupakan salah satu indikator penting dalam evaluasi proses pembelajaran di bidang pendidikan. Pemanfaatan teknik data mining dapat membantu memprediksi kelulusan siswa secara lebih objektif berdasarkan data akademik dan non-akademik. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam memprediksi kelulusan siswa. Dataset yang digunakan adalah *Student Performance Dataset* yang diperoleh dari UCI Machine Learning Repository dengan jumlah 649 data dan 33 variabel. Proses penelitian meliputi tahap prapemrosesan data, pembentukan model klasifikasi, serta evaluasi kinerja model menggunakan metrik accuracy, precision, recall, dan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki nilai akurasi dan precision yang lebih tinggi dibandingkan Decision Tree, sedangkan Decision Tree menunjukkan nilai recall yang lebih baik dalam mendeteksi siswa yang tidak lulus. Temuan ini mengindikasikan bahwa Random Forest lebih unggul dalam menghasilkan prediksi yang tepat, sementara Decision Tree lebih efektif dalam mengenali seluruh kasus ketidaklulusan. Dengan demikian, pemilihan algoritma terbaik perlu disesuaikan dengan kebutuhan analisis dan tujuan penerapan sistem prediksi kelulusan siswa.

**Kata Kunci:** Data Mining, Decision Tree, Random Forest, Prediksi Kelulusan, Siswa

### Abstract

*Student graduation is an important indicator in evaluating the effectiveness of the educational process. The application of data mining techniques can assist in predicting student graduation more objectively based on academic and non-academic factors. This study aims to compare the performance of Decision Tree and Random Forest algorithms in predicting student graduation. The dataset used in this research is the Student Performance Dataset obtained from the UCI Machine Learning Repository, consisting of 649 instances and 33 variables. The research process includes data preprocessing, classification model construction, and performance evaluation using accuracy, precision, recall, and confusion matrix metrics. The results show that the Random Forest algorithm achieves higher accuracy and precision compared to the Decision Tree algorithm, while the Decision Tree demonstrates a higher recall in identifying students who do not graduate. These findings indicate that Random Forest provides more accurate predictions, whereas Decision Tree is more effective in detecting non-graduation cases. Therefore, the selection of the most appropriate algorithm should be adjusted according to the objectives and requirements of student graduation prediction.*

**Keywords:** Data Mining, Decision Tree, Random Forest, Graduation Prediction, Students

## 1. PENDAHULUAN

Kinerja akademik mahasiswa merupakan indikator utama keberhasilan proses pembelajaran di perguruan tinggi. Namun, banyak institusi pendidikan masih mengalami kesulitan dalam mendeteksi mahasiswa yang berpotensi mengalami kegagalan akademik secara dini. Umumnya, evaluasi dilakukan setelah mahasiswa menunjukkan penurunan nilai yang signifikan, sehingga intervensi akademik sering terlambat dilakukan.

Seiring dengan meningkatnya ketersediaan data akademik mahasiswa, pendekatan berbasis data mining dan machine learning menjadi alternatif yang menjanjikan untuk memprediksi performa akademik. Teknik klasifikasi memungkinkan pemodelan hubungan antara berbagai faktor akademik dan non-akademik dengan capaian nilai mahasiswa. Dengan model prediksi yang akurat, institusi pendidikan dapat mengidentifikasi mahasiswa berisiko dan memberikan pendampingan lebih awal.

Berbagai penelitian telah menerapkan algoritma machine learning untuk memprediksi performa akademik mahasiswa. Namun, perbedaan karakteristik data dan algoritma menyebabkan variasi hasil yang signifikan, sehingga diperlukan studi komparatif untuk menentukan metode yang paling sesuai. Decision Tree dikenal mampu menghasilkan model yang mudah diinterpretasikan, sedangkan Random Forest memiliki keunggulan dalam meningkatkan akurasi melalui mekanisme ensemble. Selain itu Decision Tree dan Random Forest sudah terbukti digunakan untuk mengukur dan mengklasifikasikan Data diberbagai macam bidang seperti keamanan siber [1]. Kemudian bidang Kesehatan [2], [3], [4], [5] maupun dibidang lainnya [6], [7], [8], [9] termasuk Pendidikan [10].

Berdasarkan hal tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam mengklasifikasikan status kelulusan mahasiswa menggunakan dataset Student Performance dari UCI

Machine Learning Repository[11]. Evaluasi dilakukan menggunakan metrik accuracy, precision, dan recall untuk menilai keandalan masing-masing model.

Hasil penelitian ini diharapkan dapat memberikan rekomendasi algoritma yang paling efektif dalam memprediksi performa akademik mahasiswa, serta memperkuat pemanfaatan machine learning sebagai alat pendukung pengambilan keputusan di bidang pendidikan.

## 2. METODOLOGI PENELITIAN

### 2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, yang bertujuan untuk menguji dan membandingkan kinerja dua algoritma data mining, yaitu Decision Tree dan Random Forest, dalam memprediksi status kelulusan siswa. Pendekatan kuantitatif dipilih karena data yang digunakan bersifat numerik dan kategorikal yang dapat diolah secara statistik serta dianalisis menggunakan algoritma machine learning.

Metode eksperimen dilakukan dengan cara membangun model klasifikasi berdasarkan data historis siswa, kemudian mengevaluasi performa model menggunakan data uji. Hasil dari masing-masing algoritma dibandingkan untuk mengetahui algoritma yang memiliki kinerja terbaik dalam konteks prediksi kelulusan siswa.

### 2.2 Dataset dan Tahapan Pra-pemrosesan Data

Dataset yang digunakan dalam penelitian ini merupakan dataset resmi *Student Performance Dataset* yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi data akademik dan non-akademik siswa pada mata pelajaran Bahasa Portugis dengan total 649 data dan 33 atribut.

Atribut yang digunakan mencakup karakteristik demografis siswa, latar belakang keluarga, kebiasaan belajar, serta nilai akademik. Variabel target dalam penelitian ini adalah nilai akhir siswa (G3) yang kemudian ditransformasikan menjadi kelas biner, yaitu **Lulus** dan **Tidak Lulus**. Siswa dengan nilai  $G3 \geq 10$  dikategorikan sebagai Lulus, sedangkan siswa dengan nilai  $G3 < 10$  dikategorikan sebagai Tidak Lulus.

Tahapan pra-pemrosesan data dilakukan untuk memastikan data siap digunakan dalam proses pemodelan. Data kategorikal seperti jenis kelamin, status keluarga, dan dukungan pendidikan diubah menjadi representasi numerik menggunakan teknik *label encoding*. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan rasio 70:30. Pembagian ini bertujuan untuk melatih model menggunakan sebagian besar data sekaligus menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

### 2.3 Pemodelan dan Evaluasi Model

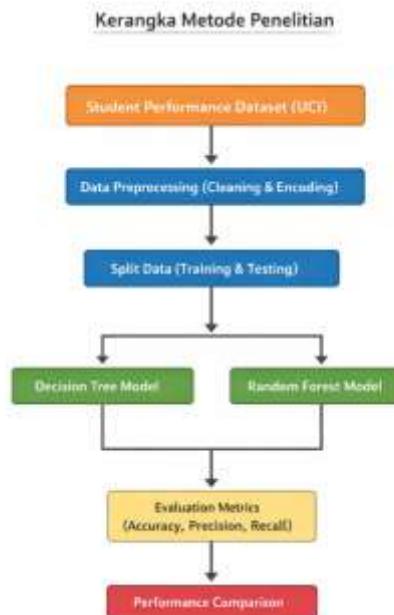
Pada tahap pemodelan, penelitian ini menerapkan dua algoritma klasifikasi, yaitu Decision Tree dan Random Forest. Algoritma Decision Tree bekerja dengan membangun struktur pohon keputusan berdasarkan pemilihan atribut terbaik pada setiap node, yang ditentukan menggunakan kriteria pemisahan seperti *Gini Index*. Proses ini menghasilkan aturan-aturan keputusan yang mudah dipahami dan diinterpretasikan.

Sementara itu, Random Forest merupakan pengembangan dari Decision Tree yang membangun banyak pohon keputusan secara acak. Setiap pohon dilatih menggunakan subset data dan atribut yang berbeda, kemudian hasil prediksi digabungkan menggunakan mekanisme *majority voting*. Pendekatan ini bertujuan untuk meningkatkan akurasi dan mengurangi risiko overfitting yang sering terjadi pada Decision Tree tunggal.

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, dan *recall*. Accuracy mengukur tingkat ketepatan model secara keseluruhan dalam memprediksi kelas kelulusan siswa. Precision digunakan untuk mengetahui ketepatan model dalam memprediksi siswa yang benar-benar lulus, sedangkan recall mengukur kemampuan model dalam menemukan seluruh siswa yang benar-benar lulus dari data yang ada.

Selain itu, confusion matrix digunakan untuk memberikan gambaran yang lebih rinci mengenai hasil klasifikasi, termasuk jumlah prediksi benar dan salah pada masing-masing kelas. Hasil evaluasi dari kedua algoritma kemudian dibandingkan untuk menentukan algoritma yang memiliki kinerja terbaik dalam memprediksi kelulusan siswa.

Kerangka ini memastikan bahwa proses yang dilakukan pada tahap metodologi konsisten dengan hasil yang disajikan pada BAB III.



Gambar 1. Kerangka Penelitian

Diagram ini berfungsi sebagai peta proses yang menunjukkan keterkaitan antara tahap pengolahan data, pemodelan, evaluasi, hingga penarikan kesimpulan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Deskripsi Umum Eksperimen

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *Decision Tree* dan *Random Forest* dalam memprediksi status kelulusan siswa menggunakan *Student Performance Dataset* dari *UCI Machine Learning Repository*. Dataset terdiri dari 649 data siswa dengan 33 atribut, yang mencakup faktor demografis, keluarga, sosial, serta akademik.

Variabel target dalam penelitian ini adalah status kelulusan, yang dibentuk berdasarkan nilai akhir (G3), dengan kriteria:

- a. Lulus, jika nilai G3  $\geq 10$
- b. Tidak Lulus, jika nilai G3  $< 10$

Atribut nilai akademik awal (G1, G2, dan G3) tidak digunakan sebagai fitur masukan untuk menghindari *data leakage*, sehingga model dilatih menggunakan faktor non-akademik dan perilaku siswa.

Dataset dibagi menjadi 70% data latih dan 30% data uji. Evaluasi kinerja model dilakukan menggunakan metrik Accuracy, Precision, Recall, serta Confusion Matrix.

#### 3.2 Hasil Evaluasi Model Klasifikasi

##### 3.2.1 Perbandingan Nilai Accuracy, Precision, dan Recall

Hasil pengujian kedua algoritma ditampilkan pada Tabel 1, yang menunjukkan perbandingan kinerja antara *Decision Tree* dan *Random Forest*.

Tabel 1. Perbandingan Kinerja Algoritma Klasifikasi

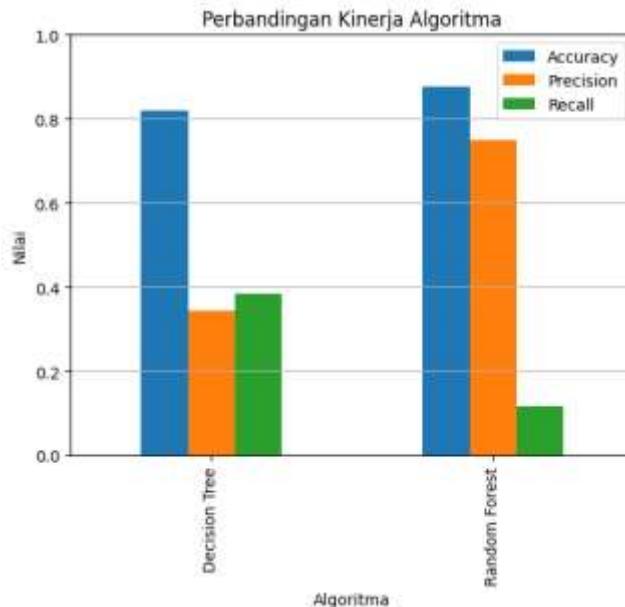
Algoritma	Accuracy	Precision	Recall
Decision Tree	0,821	0,345	0,385
Random Forest	0,877	0,750	0,115

Berdasarkan Tabel 1, algoritma *Random Forest* menghasilkan nilai accuracy tertinggi, yaitu sebesar 87,7%, dibandingkan dengan *Decision Tree* yang memperoleh akurasi 82,1%. Hal ini menunjukkan bahwa *Random Forest* secara umum lebih baik dalam memprediksi status kelulusan siswa.

Namun demikian, perbedaan yang signifikan terlihat pada nilai recall. Decision Tree memiliki nilai recall yang lebih tinggi (38,5%) dibandingkan Random Forest (11,5%). Nilai recall yang lebih tinggi menunjukkan bahwa Decision Tree lebih mampu mendeteksi siswa yang termasuk dalam kelas positif (Lulus) dibandingkan Random Forest.

### 3.2.2 Visualisasi Perbandingan Kinerja Model

Perbandingan kinerja kedua algoritma berdasarkan nilai *accuracy*, *precision*, dan *recall* divisualisasikan dalam bentuk grafik batang pada Gambar 2.



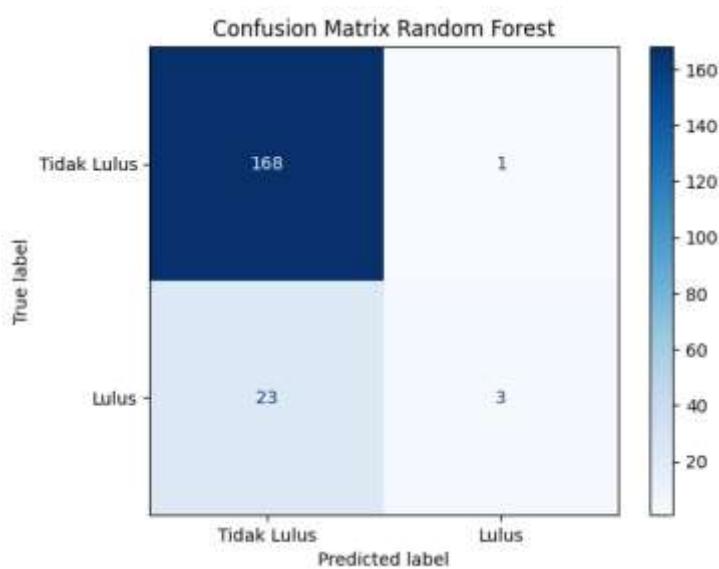
Gambar 2. Perbandingan Kinerja Algoritma *Decision Tree* dan Random Forest

Berdasarkan Gambar 2, dapat diamati bahwa Random Forest unggul dalam hal accuracy dan precision, sementara Decision Tree menunjukkan performa yang lebih baik pada aspek recall. Hal ini mengindikasikan adanya perbedaan karakteristik dalam cara kedua algoritma melakukan klasifikasi..

### 3.3 Analisis Confusion Matrix

#### 3.3.1 Confusion Matrix Algoritma Random Forest

Untuk memahami pola kesalahan klasifikasi, dilakukan analisis menggunakan *confusion matrix*, yang ditunjukkan pada Gambar 3.



Gambar 3. Confusion Matrix Algoritma Random Forest

Berdasarkan confusion matrix Random Forest, diperoleh hasil sebagai berikut:

- a) 168 data siswa tidak lulus berhasil diklasifikasikan dengan benar (*true negative*)
- b) 1 data siswa tidak lulus salah diklasifikasikan sebagai lulus (*false positive*)
- c) 3 data siswa lulus berhasil diklasifikasikan dengan benar (*true positive*)
- d) 23 data siswa lulus salah diklasifikasikan sebagai tidak lulus (*false negative*)

Hasil ini menunjukkan bahwa Random Forest sangat berhati-hati dalam memprediksi kelas *Lulus*, yang tercermin dari jumlah *false positive* yang sangat kecil. Namun, konsekuensinya adalah tingginya jumlah *false negative*, yang menyebabkan nilai recall menjadi rendah.

### 3.4 Pembahasan Hasil

Hasil eksperimen menunjukkan bahwa Random Forest memiliki kinerja yang lebih baik secara keseluruhan berdasarkan nilai accuracy dan precision. Hal ini sejalan dengan karakteristik Random Forest yang merupakan metode *ensemble*, sehingga mampu mengurangi overfitting dan meningkatkan stabilitas prediksi.

Namun, nilai recall yang rendah pada Random Forest menunjukkan bahwa model ini kurang sensitif dalam mendeteksi siswa yang lulus, karena cenderung mengklasifikasikan sebagian besar data ke dalam kelas Tidak Lulus. Kondisi ini dapat disebabkan oleh ketidakseimbangan kelas (class imbalance) pada dataset, di mana jumlah siswa tidak lulus lebih dominan dibandingkan siswa lulus.

Sebaliknya, Decision Tree memiliki nilai recall yang lebih tinggi, yang menunjukkan bahwa algoritma ini lebih sensitif dalam mengidentifikasi siswa yang lulus, meskipun dengan tingkat akurasi dan precision yang lebih rendah. Hal ini menunjukkan adanya trade-off antara kemampuan model dalam meminimalkan kesalahan klasifikasi secara keseluruhan dan kemampuan mendeteksi kelas tertentu.

Dengan demikian, pemilihan algoritma terbaik sangat bergantung pada tujuan penggunaan model. Jika tujuan utama adalah memperoleh prediksi yang stabil dan akurat secara umum, maka Random Forest lebih sesuai. Namun, jika tujuan penelitian lebih menekankan pada deteksi siswa yang lulus, maka Decision Tree dapat menjadi alternatif yang lebih baik.

## 4. KESIMPULAN

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam memprediksi kelulusan siswa menggunakan dataset *Student Performance* dari UCI Machine Learning Repository. Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada BAB IV, dapat ditarik beberapa kesimpulan sebagai berikut.

Pertama, kedua algoritma yang digunakan, yaitu Decision Tree dan Random Forest, mampu melakukan klasifikasi status kelulusan siswa dengan tingkat akurasi yang cukup baik. Decision Tree menghasilkan nilai akurasi sebesar 82,05%, sedangkan Random Forest menunjukkan kinerja yang lebih tinggi dengan akurasi sebesar 87,69%. Hal ini menunjukkan bahwa pendekatan ensemble pada Random Forest lebih efektif dalam meningkatkan akurasi prediksi dibandingkan model pohon keputusan tunggal.

Kedua, dari sisi metrik evaluasi lainnya, Decision Tree memiliki nilai recall yang lebih tinggi dibandingkan Random Forest. Hal ini menunjukkan bahwa Decision Tree lebih baik dalam mengenali siswa yang tidak lulus. Sebaliknya, Random Forest memiliki nilai precision yang lebih tinggi, yang mengindikasikan bahwa prediksi kelulusan yang dihasilkan cenderung lebih tepat, meskipun masih terdapat keterbatasan dalam mendeteksi seluruh siswa yang tidak lulus.

Ketiga, perbedaan hasil antara kedua algoritma menunjukkan adanya trade-off antara accuracy, precision, dan recall. Oleh karena itu, pemilihan algoritma terbaik tidak hanya bergantung pada akurasi semata, tetapi juga harus disesuaikan dengan tujuan penelitian atau kebutuhan pengguna, misalnya apakah lebih mengutamakan ketepatan prediksi atau kemampuan mendeteksi kasus tertentu.

Secara keseluruhan, hasil penelitian membuktikan bahwa Random Forest memiliki kinerja yang lebih unggul dalam hal akurasi dan precision, sedangkan Decision Tree lebih baik dalam hal interpretabilitas dan recall. Dengan demikian, kedua algoritma memiliki kelebihan dan kekurangan masing-masing dalam konteks prediksi kelulusan siswa.

## UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada seluruh anggota kelompok atas kerja sama, komitmen, dan kontribusi yang diberikan selama proses penelitian dan penyusunan jurnal ini. Kerja sama yang baik dari seluruh anggota sangat membantu dalam penyelesaian jurnal ini dengan optimal.

## DAFTAR PUSTAKA

- [1] A. Budi Aulianoor and M. Koprawi, “Comparative Analysis of the Performance of Decision Tree and Random Forest Algorithms in SQL Injection Attack Detection,” 2024. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [2] H. Oktavianto, H. W. Sulistyо, G. Wijaya, D. Irawan, and G. Abdurrahman, “Analisis Komparasi Kinerja Metode Decision Tree dan Random Forest dalam Klasifikasi Teks Data Kesehatan,” *Bina Insani ICT Journal*, vol. 11, no. 1, pp. 56–65, 2024, [Online]. Available: <https://www.kaggle.com/datasets/falgunipatel19/biomedical-text-publication-classification>.

- [3] N. I. Yaman, A. R. Juwita, S. A. P. Lestari, and S. Faisal, “Perbandingan Kinerja Algoritma Decision Tree dan Random Forest untuk Klasifikasi Nutrisi pada Makanan Cepat Saji,” *Jurnal Algoritma*, vol. 21, no. 2, pp. 184–196, Dec. 2024, doi: 10.33364/algoritma.v.21-2.1649.
- [4] M. Y. Iskandar and H. W. Nugroho, “Comparative Evaluation of Decision Tree and Random Forest for Lung Cancer Prediction Based on Computational Efficiency and Predictive Accuracy,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3392–3404, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.4877.
- [5] A. Ryan Dana, R. Valentino Kristananda, M. Bagas Satrio Wibowo, and D. Arman Prasetya, “Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke,” 2024.
- [6] P. Ayu Firnanda *et al.*, “Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket,” *Emerging Statistics and Data Science Journal*, vol. 3, no. 1, 2025.
- [7] A. Irma Purnamasari and I. Ali, “PERBANDINGAN TINGKAT AKURASI ALGORITMA DECISION TREE DAN RANDOM FOREST DALAM MENKLASIFIKASI PENERIMA BANTUAN SOSIAL BPNT DI DESA SLANGIT,” 2024.
- [8] A. Nugroho, “Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan,” *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, vol. 1, no. 1, pp. 41–49, Dec. 2022, doi: 10.53624/jsitik.v1i1.154.
- [9] R. A. Silvana, N. Anggiani, A. Labib, R. L. Pratiwi, and E. Widanegsih, “Perbandingan Kinerja Algoritma Decision Tree dan Random Forest dalam Memprediksi Kepuasan Penumpang Maskapai,” *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 3, 2025, doi: 10.55606/juisik.v5i3.1759.
- [10] M. Amin, C. Rizal, and I. Muslem, “Instal : Jurnal Komputer Development of Multimodal Generative AI Models for Adaptive Education Personalization in the Era of Quantum Machine Learning.” 2025, doi: 10.54209/jurnalinstall.v17i09.437.
- [11] P. Cortez., “Student Performance,” UCI Machine Learning Repository. Accessed: Jan. 08, 2026. [Online]. Available: <https://archive.ics.uci.edu/dataset/320/student+performance>
- [12] Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6), 601–618.