Department of Digital Business
**Journal of Artificial Intelligence and Digital Business (RIGGS)**
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 17235- 17242
P-ISSN: 2963-9298, e-ISSN: 2963-914X

# Analisis Komparatif Algoritma *Decision Tree* dan *Random Forest* untuk Prediksi Kelulusan Mahasiswa

Yuce Yuliani
Sistem Informasi, Universitas Gunadarma,
yuce@staff.gunadarma.ac.id

***Abstrak***

*Prediksi kelulusan mahasiswa merupakan salah satu upaya strategis yang dapat dilakukan perguruan tinggi untuk meningkatkan kualitas layanan akademik, mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan studi, serta mengurangi risiko dropout melalui pemberian intervensi akademik secara lebih dini. Pemanfaatan teknik data mining dan machine learning memungkinkan institusi pendidikan menghasilkan model prediksi yang akurat sebagai dasar pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam memprediksi kelulusan mahasiswa menggunakan dataset Predict Students' Dropout and Academic Success yang diperoleh dari UCI Machine Learning Repository. Dataset awal terdiri atas 4.424 data mahasiswa dengan 37 atribut yang mencakup karakteristik akademik, demografis, dan sosial ekonomi. Setelah melalui tahapan preprocessing yang meliputi pembersihan data, seleksi kelas, dan penanganan data yang tidak relevan, diperoleh 3.630 data yang digunakan dalam proses pemodelan. Data kemudian dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki performa yang lebih baik dibandingkan Decision Tree dengan nilai accuracy sebesar 91,46%, sedangkan Decision Tree memperoleh accuracy sebesar 85,81%. Analisis feature importance menunjukkan bahwa variabel akademik, terutama Curricular Units 2nd Sem (Approved) dan Curricular Units 2nd Sem (Grade), merupakan faktor yang paling berpengaruh terhadap kelulusan mahasiswa. Temuan ini menunjukkan bahwa Random Forest merupakan model yang efektif untuk memprediksi kelulusan mahasiswa dan berpotensi diimplementasikan sebagai bagian dari sistem peringatan dini guna mendukung peningkatan keberhasilan studi di perguruan tinggi.*

***Kata Kunci:*** *Random Forest, Decision Tree, Prediksi Kelulusan Mahasiswa, Educational Data Mining, Feature Importance*

## 1. Latar Belakang

Perkembangan transformasi digital telah mendorong perguruan tinggi untuk memanfaatkan data akademik sebagai sumber informasi dalam pengambilan keputusan. Data yang dihasilkan dari berbagai aktivitas pendidikan dapat dianalisis menggunakan teknologi data mining dan machine learning untuk mengidentifikasi pola, mengevaluasi kinerja akademik, serta mendukung peningkatan kualitas pendidikan. Dengan demikian, pemanfaatan data akademik menjadi bagian penting dalam mewujudkan pengelolaan pendidikan tinggi yang lebih efektif dan berbasis data [1], [2].

Kelulusan mahasiswa merupakan salah satu indikator penting dalam menilai keberhasilan penyelenggaraan pendidikan tinggi. Namun, tingginya angka putus studi (dropout) dan keterlambatan penyelesaian studi masih menjadi tantangan yang dihadapi banyak perguruan tinggi. Kondisi tersebut tidak hanya berdampak pada efektivitas institusi pendidikan, tetapi juga memengaruhi perkembangan akademik dan karier mahasiswa. Berbagai faktor, seperti prestasi akademik, kondisi sosial ekonomi, karakteristik individu, serta lingkungan pendidikan, dapat memengaruhi keberhasilan mahasiswa dalam menyelesaikan studi. Oleh karena itu, diperlukan upaya yang mampu mengidentifikasi mahasiswa yang berisiko mengalami dropout atau keterlambatan kelulusan sejak dini sehingga institusi pendidikan dapat menerapkan strategi intervensi yang lebih tepat dan efektif [3], [4].

Prediksi kelulusan mahasiswa menjadi salah satu pendekatan yang penting dalam mendukung pengelolaan pendidikan tinggi berbasis data. Melalui pemanfaatan data akademik, institusi pendidikan dapat mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan atau putus studi sehingga tindakan preventif dapat dilakukan lebih awal. Informasi yang dihasilkan dari model prediksi tidak hanya bermanfaat untuk meningkatkan keberhasilan akademik mahasiswa, tetapi juga membantu perguruan tinggi dalam merancang program pendampingan, mengoptimalkan alokasi sumber daya, serta meningkatkan kualitas layanan pendidikan. Dengan

demikian, kemampuan memprediksi kelulusan mahasiswa secara akurat menjadi faktor penting dalam mendukung pengambilan keputusan yang lebih efektif dan berorientasi pada peningkatan mutu pendidikan [5], [6], [7].

*Educational Data Mining* (EDM) merupakan bidang yang berkembang pesat dalam pemanfaatan data pendidikan untuk menghasilkan informasi yang mendukung pengambilan keputusan akademik. EDM memanfaatkan berbagai teknik *data mining*, seperti klasifikasi, klasterisasi, dan *association rule mining*, untuk mengidentifikasi pola serta hubungan yang tersembunyi dalam data pendidikan. Dalam konteks pendidikan tinggi, EDM banyak digunakan untuk memprediksi kinerja akademik, tingkat kelulusan, dan risiko *dropout* mahasiswa. Berbagai penelitian menunjukkan bahwa penerapan teknik klasifikasi mampu menghasilkan model prediksi dengan tingkat akurasi yang tinggi sehingga dapat membantu institusi pendidikan dalam mengidentifikasi mahasiswa berisiko sejak dini dan merancang strategi intervensi yang lebih efektif [8], [9]. Oleh karena itu, EDM menjadi salah satu pendekatan yang relevan dalam mendukung pengembangan sistem prediksi kelulusan mahasiswa berbasis data.

*Machine Learning* merupakan salah satu pendekatan utama dalam *Educational Data Mining* yang mampu mengolah data dalam jumlah besar untuk menghasilkan model prediksi yang akurat. Dalam bidang pendidikan, berbagai algoritma *machine learning* telah dimanfaatkan untuk memprediksi kinerja akademik, tingkat kelulusan, serta risiko *dropout* mahasiswa. Beberapa algoritma yang sering digunakan antara lain *Decision Tree*, *Random Forest*, *Support Vector Machine* (SVM), dan XGBoost. Berbagai penelitian menunjukkan bahwa algoritma tersebut mampu menghasilkan performa prediksi yang baik dengan memanfaatkan data akademik, demografis, maupun karakteristik pembelajaran mahasiswa [10], [11], [12]. Kemampuan *machine learning* dalam mengidentifikasi pola dan hubungan yang kompleks menjadikannya sebagai pendekatan yang efektif untuk mendukung sistem peringatan dini dan pengambilan keputusan di lingkungan pendidikan tinggi.

Algoritma *Decision Tree* telah banyak diterapkan dalam penelitian pendidikan untuk memprediksi kinerja akademik dan kelulusan mahasiswa. Beberapa penelitian menunjukkan bahwa *Decision Tree* mampu menghasilkan tingkat akurasi yang cukup baik dalam mengklasifikasikan keberhasilan akademik mahasiswa. Rukiastiandari et al. (2024) melaporkan bahwa optimalisasi *Decision Tree* menggunakan *Genetic Algorithm* mampu meningkatkan akurasi model dari 86,19% menjadi 87,68%. Sementara itu, Sugitha et al. (2024) menunjukkan bahwa penerapan metode SMOTE pada algoritma C4.5 dapat meningkatkan performa klasifikasi dengan akurasi mencapai 84,4%. Penelitian lain yang dilakukan oleh Co dan Casillano menunjukkan bahwa *Decision Tree* mampu memprediksi kelulusan tepat waktu dengan akurasi sebesar 88,9%. Meskipun memiliki keunggulan dalam interpretabilitas dan kemudahan implementasi, performa *Decision Tree* sering kali dipengaruhi oleh kualitas data dan rentan terhadap *overfitting*, terutama pada dataset yang kompleks [14], [15]. Oleh karena itu, diperlukan evaluasi lebih lanjut terhadap kinerja *Decision Tree* pada berbagai karakteristik data pendidikan.

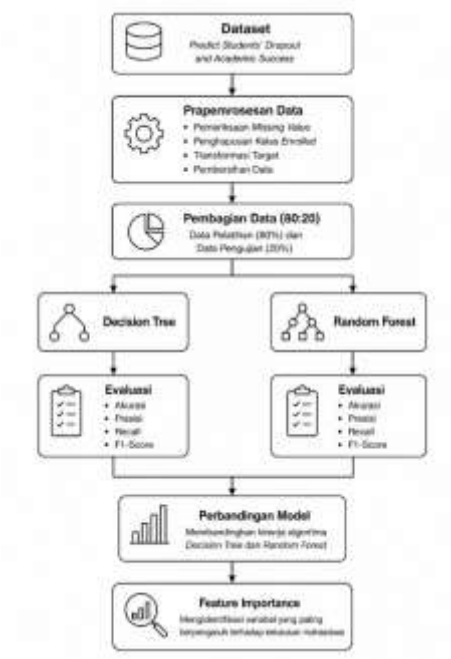
Algoritma *Random Forest* merupakan salah satu metode *ensemble learning* yang banyak digunakan dalam berbagai penelitian klasifikasi karena mampu menghasilkan tingkat akurasi yang tinggi serta memiliki kemampuan generalisasi yang baik. Berbagai penelitian menunjukkan bahwa *Random Forest* sering kali menghasilkan performa yang lebih baik dibandingkan algoritma klasifikasi tunggal, termasuk *Decision Tree*, terutama dalam menangani data yang kompleks dan berukuran besar. Keunggulan tersebut diperoleh melalui mekanisme pembentukan banyak pohon keputusan yang kemudian digabungkan untuk menghasilkan prediksi yang lebih stabil dan akurat. Selain itu, *Random Forest* memiliki ketahanan yang baik terhadap *overfitting* serta mampu mengidentifikasi variabel yang paling berpengaruh melalui analisis *feature importance*. Berbagai studi melaporkan bahwa algoritma ini mampu mencapai tingkat akurasi di atas 90% pada berbagai kasus klasifikasi, sehingga menjadikannya salah satu metode yang banyak direkomendasikan dalam pengembangan model prediksi berbasis *machine learning* [16], [17], [18].

Meskipun berbagai penelitian telah menerapkan algoritma *machine learning* untuk memprediksi keberhasilan akademik dan risiko *dropout* mahasiswa, sebagian besar penelitian masih berfokus pada penggunaan satu algoritma atau membandingkan beberapa model tanpa melakukan analisis yang mendalam terhadap kinerja *Decision Tree* dan *Random Forest* pada dataset yang sama. Selain itu, hasil yang diperoleh pada penelitian terdahulu menunjukkan performa yang bervariasi akibat perbedaan karakteristik data dan metode evaluasi yang digunakan. Oleh karena itu, masih terdapat kebutuhan untuk melakukan analisis komparatif yang lebih spesifik terhadap kedua algoritma tersebut menggunakan dataset *Predict Students' Dropout and Academic Success*. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Decision Tree* dan *Random Forest* dalam memprediksi kelulusan mahasiswa berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, penelitian ini juga menganalisis *feature importance* untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keberhasilan mahasiswa dalam menyelesaikan studi.

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Educational Data Mining* untuk memprediksi kelulusan mahasiswa. Proses penelitian meliputi pengumpulan data, *preprocessing*, pembangunan model klasifikasi menggunakan algoritma *Decision Tree* dan *Random Forest*, serta evaluasi kinerja model. Tahapan penelitian ditunjukkan pada Gambar 1.




Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan proses prapemrosesan data dan pembagian data menjadi data pelatihan serta data pengujian. Selanjutnya dilakukan pembangunan model menggunakan algoritma *Decision Tree* dan *Random Forest* yang kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil evaluasi kedua model dibandingkan untuk menentukan algoritma terbaik, kemudian dilakukan analisis *feature importance* guna mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kelulusan mahasiswa.

Dataset yang digunakan dalam penelitian ini berasal dari *UCI Machine Learning Repository* dengan nama *Predict Students' Dropout and Academic Success*. Dataset tersebut memuat informasi akademik, demografis, dan sosial ekonomi mahasiswa yang digunakan untuk membangun model prediksi kelulusan. Karakteristik dataset penelitian disajikan pada Tabel 1.

Tabel 1. Karakteristik Dataset Penelitian

| Keterangan | Nilai |
|---|---|
| Sumber Data | UCI Machine Learning Repository |
| Nama Dataset | Predict Students' Dropout and Academic Success |
| Jumlah Data Awal | 4.424 |
| Jumlah Data Setelah *Preprocessing* | 3.630 |
| Jumlah Variabel | 37 |
| Variabel Target | *Graduate*, *Dropout* |

Sumber: UCI *Machine Learning Repository* (2026), diolah peneliti.

Dataset yang digunakan dalam penelitian ini adalah *Predict Students' Dropout and Academic Success* yang diperoleh dari UCI Machine Learning Repository. Dataset ini dikembangkan oleh Mohamad Y. Realinho dkk. (2021) dan berisi 4.424 data mahasiswa dengan 37 atribut yang mencakup karakteristik akademik, demografis, serta sosial ekonomi mahasiswa. Dataset tersebut banyak digunakan dalam penelitian *Educational Data Mining* untuk memprediksi keberhasilan akademik dan risiko *dropout* mahasiswa.

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tahap *preprocessing* meliputi pemeriksaan *missing value*, penghapusan data dengan status *Enrolled*, transformasi variabel target ke dalam bentuk numerik, serta pemisahan variabel independen dan dependen. Kategori *Graduate* diberi nilai 1, sedangkan *Dropout* diberi nilai 0. Selanjutnya data dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% menggunakan metode *train-test split*.

Model klasifikasi dibangun menggunakan algoritma *Decision Tree* dan *Random Forest* yang diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka Scikit-Learn. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, analisis *feature importance* pada algoritma *Random Forest* digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kelulusan mahasiswa.

## 3. Hasil dan Diskusi

### 3.1 Distribusi Data Kelulusan Mahasiswa

Sebelum dilakukan proses pemodelan, perlu diketahui distribusi data pada masing-masing kategori target yang digunakan dalam penelitian. Setelah proses *preprocessing* dan penghapusan data dengan status *Enrolled*, diperoleh dua kategori target, yaitu *Graduate* dan *Dropout*. Distribusi data tersebut disajikan pada Tabel 2.

Tabel 2. Distribusi Data Kelulusan Mahasiswa

| Kategori | Jumlah |
|---|---|
| Graduate | 2209 |
| Dropout | 1421 |

Sumber: Hasil pengolahan data (2026).

Berdasarkan Tabel 2, jumlah mahasiswa dengan status *Graduate* sebanyak 2.209 data, sedangkan mahasiswa dengan status *Dropout* sebanyak 1.421 data. Hasil tersebut menunjukkan bahwa jumlah mahasiswa yang berhasil menyelesaikan studi lebih banyak dibandingkan mahasiswa yang mengalami *dropout*. Meskipun demikian, distribusi data pada kedua kelas masih cukup representatif untuk digunakan dalam proses klasifikasi.

### 3.2 Hasil Evaluasi Model Klasifikasi

Evaluasi model dilakukan untuk mengukur kemampuan algoritma dalam memprediksi kelulusan mahasiswa. Pada penelitian ini digunakan dua algoritma klasifikasi, yaitu *Decision Tree* dan *Random Forest*. Pengukuran performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian kedua algoritma disajikan pada Tabel 3.

Tabel 3. Perbandingan Kinerja Algoritma *Decision Tree* dan *Random Forest*

| Algoritma | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Decision Tree | 85,81% | 86% | 85% | 85% |
| Random Forest | 91,46% | 92% | 90% | 91% |

Sumber: Hasil pengolahan data (2026).

Berdasarkan Tabel 3, algoritma *Random Forest* menghasilkan performa yang lebih baik dibandingkan algoritma *Decision Tree*. Algoritma *Random Forest* memperoleh nilai *accuracy* sebesar 91,46%, sedangkan algoritma *Decision Tree* memperoleh nilai *accuracy* sebesar 85,81%. Perbedaan tersebut menunjukkan bahwa pendekatan *ensemble learning* yang digunakan oleh *Random Forest* mampu meningkatkan kemampuan model dalam melakukan klasifikasi data mahasiswa.

Algoritma Random Forest menghasilkan nilai accuracy, precision, recall, dan F1-score yang lebih tinggi dibandingkan Decision Tree. Hasil ini menunjukkan bahwa Random Forest lebih efektif dalam memprediksi status kelulusan mahasiswa karena mampu menghasilkan model yang lebih stabil dan mengurangi risiko overfitting. Rincian performa masing-masing algoritma disajikan pada Tabel 4 dan Tabel 5.

Tabel 4. Hasil Classification Report Algoritma Decision Tree

| Kelas | Precision | Recall | F1-Score |
|---|---|---|---|
| Dropout (0) | 0,83 | 0,81 | 0,82 |
| Graduate (1) | 0,88 | 0,89 | 0,88 |

Sumber: Hasil pengolahan data (2026).

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Berdasarkan Tabel 4, algoritma *Decision Tree* memiliki kemampuan yang cukup baik dalam mengklasifikasikan kedua kategori mahasiswa. Namun, nilai *recall* pada kelas *Dropout* masih relatif lebih rendah dibandingkan kelas *Graduate*, yang menunjukkan adanya sejumlah mahasiswa *dropout* yang belum berhasil diidentifikasi secara optimal oleh model.

Tabel 5. Hasil *Classification Report* Algoritma *Random Forest*

| Kelas | Precision | Recall | F1-Score |
|---|---|---|---|
| Dropout (0) | 0,95 | 0,82 | 0,88 |
| Graduate (1) | 0,90 | 0,97 | 0,93 |

Sumber: Hasil pengolahan data (2026).

Berdasarkan Tabel 5, algoritma *Random Forest* menunjukkan performa yang lebih baik dibandingkan *Decision Tree*. Nilai *precision* sebesar 0,95 pada kelas *Dropout* menunjukkan bahwa model mampu mengidentifikasi mahasiswa yang berpotensi *dropout* dengan tingkat ketepatan yang tinggi. Selain itu, nilai *recall* sebesar 0,97 pada kelas *Graduate* menunjukkan bahwa sebagian besar mahasiswa yang berhasil lulus dapat diprediksi dengan benar oleh model.

### 3.3 Analisis *Feature Importance*

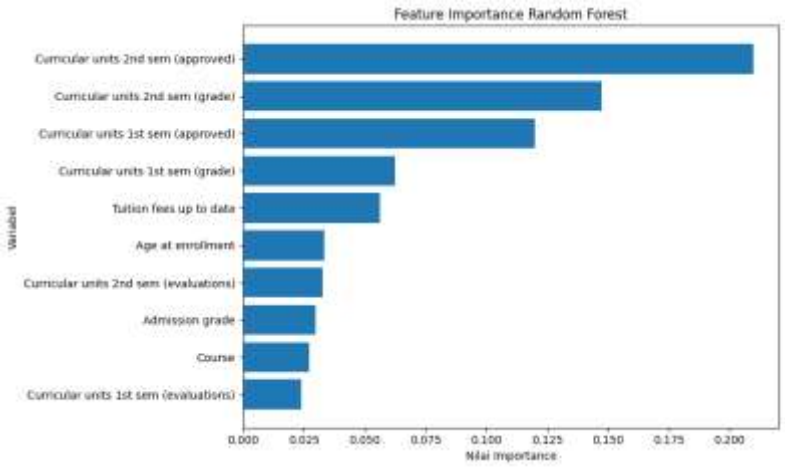
Selain mengevaluasi performa model, penelitian ini juga menganalisis faktor-faktor yang paling berpengaruh terhadap kelulusan mahasiswa menggunakan metode *feature importance* pada algoritma *Random Forest*. Hasil analisis disajikan pada Tabel 6.

Tabel 6. Variabel dengan Nilai *Feature Importance* Tertinggi

| Peringkat | Variabel | Importance |
|---|---|---|
| 1 | Curricular units 2nd sem (approved) | 0,2097 |
| 2 | Curricular units 2nd sem (grade) | 0,1475 |
| 3 | Curricular units 1st sem (approved) | 0,1201 |
| 4 | Curricular units 1st sem (grade) | 0,0626 |
| 5 | Tuition fees up to date | 0,0562 |

Sumber: Hasil pengolahan data (2026).

Untuk memberikan gambaran yang lebih jelas mengenai tingkat kepentingan setiap variabel dalam model, hasil analisis *feature importance* divisualisasikan pada Gambar 2.




Gambar 2. Feature Importance Algoritma Random Forest

Sumber: Hasil pengolahan data (2026).

Berdasarkan Tabel 6 dan Gambar 2, variabel *Curricular units 2nd sem (approved)* memiliki nilai *importance* tertinggi sebesar 0,2097. Variabel tersebut menunjukkan jumlah mata kuliah semester kedua yang

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

berhasil diselesaikan mahasiswa. Hasil ini mengindikasikan bahwa keberhasilan akademik pada tahun pertama perkuliahan merupakan faktor utama yang memengaruhi kelulusan mahasiswa.

Variabel lain yang juga memberikan kontribusi besar adalah *Curricular units 2nd sem (grade)* dan *Curricular units 1st sem (approved)*. Temuan ini menunjukkan bahwa performa akademik mahasiswa memiliki pengaruh yang lebih dominan dibandingkan faktor lainnya. Selain itu, variabel *Tuition fees up to date* juga memiliki kontribusi yang cukup signifikan, yang mengindikasikan bahwa kondisi ekonomi mahasiswa turut memengaruhi keberhasilan dalam menyelesaikan studi.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma *Random Forest* merupakan model terbaik untuk memprediksi kelulusan mahasiswa pada dataset *Predict Students' Dropout and Academic Success*. Dengan tingkat akurasi sebesar 91,46%, model ini mampu memberikan prediksi yang lebih akurat dibandingkan algoritma *Decision Tree*. Temuan ini juga memperlihatkan bahwa faktor akademik, khususnya pencapaian pada semester pertama dan kedua, menjadi determinan utama keberhasilan mahasiswa dalam menyelesaikan pendidikan tinggi.

**3.4 Diskusi**

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* menghasilkan performa yang lebih baik dibandingkan *Decision Tree* dalam memprediksi kelulusan mahasiswa. Nilai *accuracy* yang diperoleh *Random Forest* mencapai 91,46%, sedangkan *Decision Tree* memperoleh *accuracy* sebesar 85,81%. Perbedaan ini menunjukkan bahwa pendekatan *ensemble learning* pada *Random Forest* mampu meningkatkan kemampuan klasifikasi dengan menggabungkan banyak pohon keputusan sehingga menghasilkan prediksi yang lebih stabil dan akurat.

Temuan penelitian ini menunjukkan bahwa algoritma Random Forest memiliki performa yang lebih baik dibandingkan Decision Tree dalam memprediksi kelulusan mahasiswa. Hasil ini sejalan dengan penelitian Zhao et al. (2024) dan Maindola et al. (2024) yang menyatakan bahwa Random Forest mampu menghasilkan akurasi yang lebih tinggi serta lebih efektif dalam menangani data yang kompleks. Keunggulan tersebut diperoleh melalui pendekatan ensemble learning yang mampu mengurangi risiko overfitting dan meningkatkan stabilitas prediksi.

Selain mengevaluasi performa model, penelitian ini juga menganalisis faktor-faktor yang paling berpengaruh terhadap kelulusan mahasiswa melalui analisis *feature importance*. Hasil penelitian menunjukkan bahwa variabel akademik memiliki kontribusi paling dominan dalam menentukan status kelulusan mahasiswa. Variabel *Curricular Units 2nd Sem (Approved)* dan *Curricular Units 2nd Sem (Grade)* menempati posisi tertinggi dalam model prediksi, yang menunjukkan bahwa jumlah mata kuliah yang berhasil diselesaikan serta capaian nilai pada semester kedua merupakan indikator utama keberhasilan studi mahasiswa. Temuan ini mengindikasikan bahwa performa akademik pada tahun pertama perkuliahan memiliki peran yang sangat penting dalam membentuk keberhasilan akademik pada tahap berikutnya.

Temuan tersebut sejalan dengan penelitian Casanova yang menunjukkan bahwa kinerja akademik pada semester awal memiliki pengaruh signifikan terhadap pencapaian akademik pada semester berikutnya. Mahasiswa yang mampu mempertahankan prestasi akademik sejak awal perkuliahan cenderung memiliki peluang lebih besar untuk menyelesaikan studi tepat waktu dibandingkan mahasiswa yang mengalami kesulitan akademik pada semester awal [20]. Oleh karena itu, capaian akademik pada tahun pertama dapat dijadikan sebagai indikator awal dalam mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan maupun *dropout*

Selain faktor akademik, variabel *Tuition Fees Up to Date* juga menunjukkan tingkat kepentingan yang relatif tinggi dalam model prediksi. Hasil ini menunjukkan bahwa aspek ekonomi masih menjadi salah satu faktor yang memengaruhi keberhasilan mahasiswa dalam menyelesaikan studi. Mahasiswa yang mampu memenuhi kewajiban pembayaran biaya pendidikan secara tepat waktu cenderung memiliki keberlangsungan studi yang lebih baik dibandingkan mahasiswa yang menghadapi kendala finansial. Kondisi tersebut dapat memengaruhi fokus belajar, akses terhadap sumber daya pendidikan, serta keberlanjutan proses akademik selama masa perkuliahan.

Secara praktis, hasil penelitian ini memberikan implikasi penting bagi perguruan tinggi dalam mengembangkan sistem peringatan dini (*early warning system*) berbasis data. Dengan memanfaatkan model *Random Forest* yang memiliki tingkat akurasi tinggi serta informasi mengenai faktor-faktor utama yang memengaruhi kelulusan, institusi pendidikan dapat mengidentifikasi mahasiswa berisiko sejak semester awal. Selanjutnya, perguruan tinggi dapat memberikan intervensi yang lebih tepat sasaran, seperti program pendampingan akademik, layanan konseling, monitoring kemajuan studi, maupun dukungan finansial bagi mahasiswa yang membutuhkan. Langkah tersebut diharapkan dapat meningkatkan tingkat retensi mahasiswa, mengurangi angka *dropout*, dan meningkatkan persentase kelulusan tepat waktu.

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

## 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Random Forest* menunjukkan kinerja yang lebih baik dibandingkan *Decision Tree* dalam memprediksi kelulusan mahasiswa pada dataset *Predict Students' Dropout and Academic Success*. Algoritma *Random Forest* memperoleh nilai *accuracy* sebesar 91,46%, sedangkan *Decision Tree* memperoleh nilai *accuracy* sebesar 85,81%. Hasil ini menunjukkan bahwa pendekatan *ensemble learning* pada *Random Forest* lebih efektif dalam mengklasifikasikan status kelulusan mahasiswa. Selain itu, analisis *feature importance* menunjukkan bahwa variabel akademik, khususnya *Curricular Units 2nd Sem (Approved)*, *Curricular Units 2nd Sem (Grade)*, dan *Curricular Units 1st Sem (Approved)*, merupakan faktor yang paling berpengaruh terhadap keberhasilan mahasiswa dalam menyelesaikan studi. Temuan ini mengindikasikan bahwa capaian akademik pada tahun pertama perkuliahan memiliki peran penting dalam menentukan kelulusan mahasiswa. Hasil penelitian ini dapat dimanfaatkan sebagai dasar pengembangan sistem peringatan dini (*early warning system*) untuk mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan kelulusan atau *dropout*. Penelitian selanjutnya dapat mengembangkan model prediksi dengan menggunakan algoritma lain atau teknik optimasi untuk meningkatkan performa klasifikasi serta mengeksplorasi dataset pendidikan yang lebih beragam.

## Referensi

[1] A. Nazyrova, M. Miłosz, G. Bekmanova, A. Omarbekova, G. Aimicheva, and Y. Kadyr, "The Digital Transformation of Higher Education in the Context of an AI-Driven Future," *Sustainability*, vol. 17, no. 22, p. 9927, Nov. 2025, doi: 10.3390/su17229927.

[2] Y. Liu, N. Tuerdahong, and B. Wang, "Research on the Optimization Path of College Education and Teaching Based on Big Data Analysis," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-2862.

[3] P. Basavaraj and I. Garibay, "Dropout vs. Time to Degree," in *Proceedings of the 20th Annual SIG Conference on Information Technology Education*, New York, NY, USA: ACM, Sep. 2019, pp. 154–154. doi: 10.1145/3349266.3351374.

[4] A. Viloria, A. Senior Naveda, H. Hernández Palma, W. Niebles Núñez, and L. Niebles Núñez, "Using Big Data to Determine Potential Dropouts in Higher Education," *J. Phys. Conf. Ser.*, vol. 1432, no. 1, p. 012077, Jan. 2020, doi: 10.1088/1742-6596/1432/1/012077.

[5] F. Irhamna Rachman, S. Mujadilah, T. Wahyuni, and L. Anas, "Prediksi Tingkat Kelulusan Menggunakan K-Means Pada Program Studi Informatika Unismuh Makassar," *JURNAL FASILKOM*, vol. 13, no. 3, pp. 504–510, Dec. 2023, doi: 10.37859/jf.v13i3.6061.

[6] S. Abdillah, G. Juni Yanris, and V. Sihombing, "Implementation of the Support Vector Machine Method in Predicting Student Graduation," *International Journal of Science, Technology & Management*, vol. 5, no. 4, pp. 1036–1043, Aug. 2024, doi: 10.46729/ijstm.v5i4.1140.

[7] M. C. Dinh, V. T. Ha, and D. H. Dinh, "A Comparative Analysis of Machine Learning and Deep Learning Models for Student On-Time Graduation Prediction," in *2025 IEEE International Conference on Emerging Trends in Engineering and Computing (ETECOM)*, IEEE, Oct. 2025, pp. 1–7. doi: 10.1109/ETECOM66111.2025.11319063.

[8] P. Ghosh, R. Roy, S. Mandal, M. Chowdhary, and S. Bokshi, "Data mining approach to predict academic performanceof students," *BOHR International Journal of Smart Computing and Information Technology*, vol. 4, no. 1, pp. 39–49, 2023, doi: 10.54646/bijscit.2023.35.

[9] S. Alturki, I. Hulpuș, and H. Stuckenschmidt, "Predicting Academic Outcomes: A Survey from 2007 Till 2018," *Technology, Knowledge and Learning*, vol. 27, no. 1, pp. 275–307, Mar. 2022, doi: 10.1007/s10758-020-09476-0.

[10] K. Alalawi, R. Athauda, and R. Chiong, "Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review," *Engineering Reports*, vol. 5, no. 12, Dec. 2023, doi: 10.1002/eng2.12699.

[11] R. Qureshi and P. S. Lokhande, "A Comprehensive Review of Machine Learning techniques used for Designing An Academic Result Predictor And Identifying The Multi-Dimensional Factors Affecting Student's Academic Results," in *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/IDICAIEI61867.2024.10842901.

[12] W. Yustanti, R. E. Putra, I. Gusti Putu Asto Buditjahjanto, and A. I. Nurhidayat, "Exploring Ensemble Classifiers and Filter-Based Feature Selection for Predicting On-Time Graduation Using Multidimensional Student Data," in *2025 Eight International Conference on Vocational Education and Electrical Engineering (ICVEE)*, IEEE, Sep. 2025, pp. 246–251. doi: 10.1109/ICVEE66651.2025.11281493.

[13] I. P. Y. T. Sugitha, F. A. Bachtiar, and S. A. Wicaksono, "Application of Students Graduation Prediction Model Using Decision Tree C4.5 Algorithm and Synthetic Minority Oversampling Technique (SMOTE)," in *2024 Seventh International Conference on Vocational Education and Electrical Engineering (ICVEE)*, IEEE, Oct. 2024, pp. 175–181. doi: 10.1109/ICVEE63912.2024.10823806.

[14] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, "Data driven decisions in education using a comprehensive machine learning framework for student performance prediction," *Discover Computing*, vol. 28, no. 1, p. 153, Jul. 2025, doi: 10.1007/s10791-025-09585-3.

[15] S. Neththikumara and M. W. P. Maduranga, "Recent Results of Machine Learning in Predicting Student Performance: A Systematic Literature Review," in *2025 IEEE Digital Education and MOOCS Conference (DEMOcon)*, IEEE, Oct. 2025, pp. 1–5. doi: 10.1109/DEMOcon65705.2025.11282820.

[16] J. Zhao, C.-D. Lee, G. Chen, and J. Zhang, "Research on the Prediction Application of Multiple Classification Datasets Based on Random Forest Model," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, IEEE, Jul. 2024, pp. 156–161. doi: 10.1109/ICPICS62053.2024.10795875.

[17] Z. Feng, Y. Shi, D. Zhou, and L. Mo, "Research on Human Activity Recognition Based on Random Forest Classifier," in *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*, IEEE, Apr. 2023, pp. 1507–1513. doi: 10.1109/ICCECT57938.2023.10140545.

[18] M. Maindola *et al.*, "Utilizing Random Forests for High-Accuracy Classification in Medical Diagnostics," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, Sep. 2024, pp. 1679–1685. doi: 10.1109/IC3I61595.2024.10828609.

[19] J. Zhao, C.-D. Lee, G. Chen, and J. Zhang, "Research on the Prediction Application of Multiple Classification Datasets Based on Random Forest Model," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, IEEE, Jul. 2024, pp. 156–161. doi: 10.1109/ICPICS62053.2024.10795875.

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

[20] J. Casanova, J. Sinval, and L. Almeida, "Academic success, engagement and self-efficacy of first-year university students: personal variables and first-semester performance," *Anales de Psicología*, vol. 40, no. 1, pp. 44–53, Jan. 2024, doi: 10.6018/analesps.479151.

DOI: https://doi.org/10.31004/riggs.v5i2.10534
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)