JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Aplikasi Prediksi Tingkat Kelulusan Mahasiswa Berdasarkan Data Akademik dan Demografi Menggunakan Algoritma Klasifikasi Random Forest

Muhammad Arief Rachman[1], Dinar Ridho Maulana[2], Billianto Timothy[3], Riski Annisa[4]

[1,2,3]Program Studi Informatika, Unversitas Bina Sarana Informatika, Pontianak, Indonesia
Email: [1]15230696@bsi.ac.id, [2]15230789@bsi.ac.id, [3]15230580@bsi.ac.id,riski.rnc@bsi.ac.id
Email Penulis Korespondensi: [1]15230696@bsi.ac.id

Abstrak− Ketepatan waktu kelulusan mahasiswa merupakan indikator mutu krusial bagi perguruan tinggi, namun pemanfaatan data akademik untuk langkah preventif masih sangat terbatas. Selain itu, terdapat kesenjangan penelitian berupa kurangnya implementasi praktis model prediksi dalam bentuk antarmuka yang siap digunakan oleh manajemen institusi untuk intervensi harian. Penelitian ini bertujuan untuk mengembangkan Sistem Peringatan Dini Akademik proaktif melalui prediksi tingkat kelulusan mahasiswa menggunakan algoritma Random Forest (RF). Data sekunder yang digunakan mencakup 27 variabel prediktor yang mengintegrasikan fitur akademik dan demografi ke dalam klasifikasi status biner. Hasil pengujian menunjukkan performa model yang sangat efektif dengan pencapaian akurasi sebesar 94,44%, serta nilai presisi (0,97) dan recall (0,97) yang sangat andal. Sebagai kontribusi utama, model prediksi ini diimplementasikan ke dalam aplikasi web interaktif berbasis Streamlit untuk menjamin kegunaan praktis dan intuitif bagi pengambil kebijakan. Penelitian ini menyimpulkan bahwa penyediaan sistem peringatan dini yang terintegrasi secara praktis memungkinkan institusi melakukan intervensi personal secara real-time guna membantu mahasiswa menyelesaikan studi tepat waktu secara humanis.

Kata Kunci: Prediksi Kelulusan, Data Mining, Random Forest, Klasifikasi, Streamlit.

Abstract− Student graduation timeliness is a crucial quality indicator for higher education institutions, yet the utilization of academic data for preventive measures remains very limited. Furthermore, a research gap exists in the lack of practical implementation of predictive models in the form of interfaces ready for use by institutional management for daily interventions. This study aims to develop a proactive Academic Early Warning System through student graduation rate prediction using the Random Forest (RF) algorithm. The secondary data used includes 27 predictor variables that integrate academic and demographic features into a binary status classification. The test results demonstrate highly effective model performance with an accuracy of 94.44%, as well as highly reliable precision (0.97) and recall (0.97). As a key contribution, this predictive model is implemented into a Streamlit-based interactive web application to ensure practical and intuitive usability for policymakers. This study concludes that providing a practical, integrated early warning system enables institutions to conduct real-time personalized interventions to help students complete their studies on time in a humane manner.

Keywords: Graduation Prediction, Data Mining, Random Forest, Classification, Streamlit.

1. PENDAHULUAN

Sangat penting bagi institusi pendidikan tinggi untuk memastikan bahwa mahasiswa menyelesaikan studi mereka tepat waktu. Jika mahasiswa tidak lulus tepat waktu, hal itu akan menyebabkan beban biaya pendidikan yang lebih besar bagi mereka. Selain itu, keterlambatan kelulusan dapat berdampak negatif pada kualitas manajemen akademik dan penilaian akreditasi institusi [1]. Sebagian besar institusi memiliki banyak data akademik, tetapi seringkali mereka hanya menggunakannya untuk pelaporan deskriptif pasca-kejadian daripada untuk intervensi preventif. Oleh karena itu, data mining harus diterapkan dalam manajemen akademik untuk mencegah keterlambatan dan memberikan bantuan yang tepat sasaran [2],[3].

Algoritma klasifikasi telah digunakan dalam penelitian sebelumnya untuk memprediksi kinerja mahasiswa. Pola data akademik sering ditemukan dengan algoritma tunggal seperti Decision Tree, K-Nearest Neighbor (KNN), Naive Bayes, dan Support Vector Machine (SVM)[4]. Meskipun algoritma tunggal seringkali berhasil, mereka seringkali memiliki keterbatasan saat menangani dataset pendidikan yang kompleks dan rentan terhadap *overfitting*. Sebagai solusi, algoritma Random Forest dianggap sebagai metode kelompok belajar yang lebih stabil, generalisasi yang lebih baik, dan efektif untuk menangani berbagai variabel prediktor [5].

Tabel 1 berikut membandingkan penelitian sebelumnya yang berfokus pada pengoptimalan algoritma Random Forest untuk prediksi kelulusan untuk menunjukkan posisi dan kebaruan (novelty) penelitian ini:

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 1. Perbandingan Penelitian Terdahulu dan Kontribusi Penelitian

Peneliti (Tahun)	Metode	Hasil Utama	Fokus & Keterbatasan
Pso dkk. (2023)	Random Forest	Studi perbandingan dan optimasi algoritma untuk mencapai akurasi prediksi kelulusan tepat waktu	Fokus hanya pada optimasi teknis model; *belum meengimplementasi web early warning system* yang dapat diakses oleh *end-user.*
Penelitian Ini (2025)	Random Forest + Streamlit	Akurasi 94,44% dengan integrasi Data Akademik – Demografi	Mengisi celah implementasi dengan membangun Sistem Peringat Dini interaktif berbasis *framework* Streamlit.

Tabel 1 menunjukkan bahwa ada perbedaan (gap) yang signifikan dalam penelitian sebelumnya [6], biasanya berkonsentrasi pada eksperimen yang meningkatkan akurasi atau tinjauan teknis model. Model prediksi, meskipun sangat akurat, menjadi kurang optimal jika tidak dikemas ke dalam antarmuka yang mudah digunakan oleh pengambil keputusan institusi. Kurangnya implementasi model prediksi menjadi sistem yang siap pakai dan mudah digunakan adalah gap analysis penelitian ini.

Oleh karena itu, penelitian ini bertujuan untuk mengisi kekurangan tersebut dengan mengembangkan aplikasi prediksi kelulusan yang menggunakan algoritma Random Forest. Algoritma ini diintegrasikan dengan *framework* Streamlit karena mampu mengubah model pembelajaran mesin menjadi aplikasi *web interaktif* yang ringan dan mudah digunakan [7]. Penemuan baru dari penelitian ini adalah sistem peringatan dini yang praktis dan *real-time* yang memungkinkan institusi untuk memprediksi dan melakukan intervensi humanis berdasarkan data untuk meningkatkan persentase kelulusan tepat waktu.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Metode *data mining*, yang digunakan dalam penelitian ini, adalah pendekatan kuantitatif yang dipilih[8],[9]. Kemampuannya untuk mengolah sejumlah besar data dan menghasilkan hasil yang dapat diukur secara statistik merupakan dasar dari keputusan untuk menggunakan metode ini.

Dengan menggunakan teknik *data mining*, terutama klasifikasi, model dibuat untuk memprediksi status kelulusan mahasiswa berdasarkan atribut sosial dan akademik mereka [10]. Metode *data mining*, khususnya klasifikasi, digunakan untuk membuat model prediksi status kelulusan mahasiswa berdasarkan nilai sosial dan akademik mereka [11]. Tujuan utama penelitian ini adalah membuat model yang dapat melakukan prediksi dengan akurat untuk membantu lembaga pendidikan membuat keputusan tentang memberikan intervensi dini kepada mahasiswa.

2.2 Sumber data dan Variabel Penelitian

2.2.1 Sumber Data

Data yang digunakan adalah data sekunder berupa dataset publik yang berfokus pada pencapaian mahasiswa di Portugis, yang bersumber dari platform Kaggle (https://www.kaggle.com/datasets/whenamancodes/student-performance). Dataset ini berasal dari laporan sekolah dan kuesioner, dan secara spesifik mencakup 395 catatan mahasiswa.

2.2.2 Variabel Penelitian

Penelitian ini menggunakan 30 variabel prediktor dan satu variabel target.

A. Variabel Prediktor (Independent Variables)

Variabel prediktor terdiri dari 30 fitur yang mencakup data demografi, sosial, dan akademik mahasiswa.

Tabel 2. Variabel Prediktor

Kategori Variabel	Variabel (Contoh)	Tipe Data	Keterangan/Skala
Akademik Inti	*failures*, *absences*	Numerik	Jumlah kegagalan kelas sebelumnya, jumlah absen.
Dukungan Edukasi	*Medu* (Pendidikan Ibu), *Fedu* (Pendidikan Ayah), *studytime*, *schoolsup*, *famsup*, *higher*	Numerik/Kategorikal	Skala 0-4 untuk pendidikan; waktu belajar (1-4); dukungan (ya/tidak).
Sosial & Demografi	*age*, *sex*, *address*, *famrel*,	Numerik/Kategorikal	Usia, jenis kelamin, kualitas

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

	freetime, *goout*, *Dalc*, *Walc*, *romantic*	hubungan keluarga (1-5), konsumsi alkohol (1-5).

Variabel prediktor meliputi informasi demografis, sosial, dan akademik mahasiswa seperti *school*, *sex*, *age*, *address*, *studytime*, *failures*, *Medu*, *Fedu*, dan variabel lain yang relevan dengan lingkungan pendidikan.
Beberapa variabel diketahui memiliki pengaruh signifikan terhadap kelulusan, seperti:

a. Pendidikan orang tua (*Medu* dan *Fedu*)
b. Waktu belajar mingguan (*studytime*)
c. Jumlah kegagalan akademik sebelumnya (*failures*) [10]

B. Variabel Target (Dependent Variables)

Variabel target adalah Kelulusan, yang merupakan hasil transformasi nilai akhir (*G3*) menjadi bentuk biner.

Tabel 3. Variabel Target

Variabel	Keterangan	Nilai	Distribusi
Kelulusan	Status kelulusan mahasiswa	1 (Lulus) jika G3 ≥ 10 o (Tidak Lulus) jika G3 < 10.	≈ 67.09 % Lulus ≈ 32.91% Tidak Lulus

2.3 Metode Data Mining : Algoritma Random Forest

Penelitian ini menggunakan algoritma Random Forest (RF) untuk klasifikasi [5].

2.3.1 Prinsip Kerja Random Forest

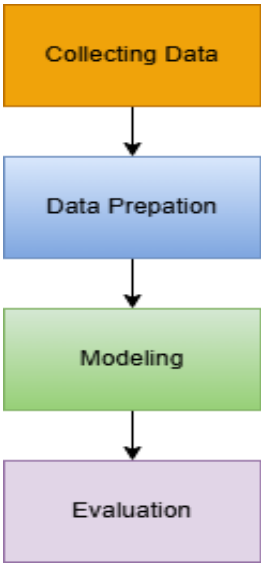
Random Forest adalah metode pembelajaran kelompok yang dibangun dari kumpulan pohon keputusan [10]. Algoritma ini bekerja dengan cara:

a. Sampel dataset diambil dengan *bootstraping* untuk setiap pohon.
b. Subset fitur acak dipilih secara acak untuk setiap simpul pohon.
c. Pemilihan: Prediksi akhir dibuat melalui proses mayoritas suara, atau mayoritas suara, dari semua pohon keputusan yang terbentuk.

Algoritma ini dipilih karena dapat menangani dataset yang kompleks, mengurangi masalah *overfitting*, dan biasanya memberikan prediksi yang sangat akurat [5].

2.4 Tahapan Penelitian

Tahapan Data Mining dalam penelitian ini didasarkan pada metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM), yang mencakup pemahaman data, pra-pemrosesan, pemodelan, evaluasi, dan implementasi. Seluruh proses penelitian dari pengumpulan data hingga pengimplementasian Sistem Peringatan Dini disajikan secara sistematis dalam diagram di bawah ini.



Gambar 1. Diagram Alir

2.4.1 Pra-pemrosesan Data (Data Preprocessing)

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Pra-pemrosesan data merupakan tahap pertama setelah pengumpulan data mentah [12], bertujuan untuk memastikan kualitas dan kesiapan data (30 variabel akademik dan demografi) sebelum digunakan dalam pemodelan. Data mentah diolah sesuai dengan alur yang ditunjukkan pada Gambar 1, meliputi:

1. Pembersihan Data (Data Cleaning): Meliputi pemeriksaan inkonsistensi data dan penanganan missing value. Untuk data numerik, missing value ditangani dengan metode imputasi rata-rata, sementara untuk data kategorikal digunakan modus.
2. Transformasi Data (Data Transformation):
   a) Encoding: Variabel kategorikal (misalnya jenis kelamin, status pendaftaran) diubah menjadi format numerik menggunakan teknik One-Hot Encoding atau Label Encoding agar dapat diproses oleh algoritma.
   b) Normalisasi: Data numerik diskalakan ke rentang nilai yang seragam (misalnya 0 hingga 1) menggunakan Min-Max Scaler untuk mencegah variabel dengan nilai terbesar mendominasi model.
3. Pembagian Data (Data Splitting): Data hasil pra-pemrosesan kemudian dibagi menjadi Data Pelatihan (80%) dan Data Pengujian (20%) untuk melatih dan menguji performa model.

2.4.2 Pemodelan dan Optimasi

Algoritma Random Forest (RF) dipilih sebagai metode klasifikasi utama dalam penelitian ini berdasarkan pertimbangan sebagai berikut:

1. Mengatasi Kelemahan Decision Tree: Sebagai metode Ensemble Learning, RF menggabungkan hasil prediksi dari banyak Decision Tree individu. Proses agregasi ini secara signifikan mengurangi risiko overfitting yang sering terjadi pada Decision Tree tunggal, menjamin model memiliki kemampuan generalisasi yang baik.
2. Stabilitas dan Akurasi Tinggi: RF dikenal menghasilkan prediksi yang lebih stabil dan akurat dalam konteks data kompleks, seperti data akademik yang melibatkan variabel campuran (numerik dan kategorikal).
3. Penanganan Variabel Beragam: RF dapat menangani set data dengan dimensi tinggi dan mampu bekerja secara efektif dengan variabel prediktor yang bervariasi antara akademik dan demografi.
4. Bukti Empiris Terdahulu: Penelitian serupa tentang prediksi kelulusan, termasuk studi yang berfokus pada optimasi teknis model [6], telah membuktikan bahwa Random Forest secara konsisten memberikan metrik kinerja yang unggul.

2.4.3 Evaluasi Kinerja Model

Untuk mengetahui seberapa baik model dapat diterapkan pada data baru, evaluasi dilakukan pada data pengujian[13].

a. Akurasi (*Accuracy*): Persentase perkiraan yang benar.
b. Presisi (*Precision)*: prediksi kasus positif yang benar dari semua prediksi positif.
c. Daya Ingat (*Recall*): prediksi kasus positif yang benar dari total kasus positif yang ada.
d. F1-Score: Standard Harmonik Presisi dan *Recall*.
e. *Area Under the Curve* (AUC): Mengevaluasi kapasitas model untuk membedakan kelas.

2.5 Perancangan dan Implementasi Aplikasi

Pengembangan sistem dalam penelitian ini dilakukan menggunakan lingkungan pemrograman Python 3.10.x sebagai dasar utama pengembangan sistem dan pengolahan data. Untuk implementasi algoritma Random Forest serta perhitungan metrik evaluasi kinerja model, digunakan library Scikit-learn (v1.2.x), sedangkan proses persiapan data yang meliputi pembersihan serta transformasi fitur kategorikal dan numerik didukung oleh pustaka Pandas (v1.5.x) dan NumPy (v1.23.x). Melalui framework Streamlit (v1.18.x), model Random Forest yang telah diuji dan ideal diimplementasikan ke dalam aplikasi berbasis web yang berfungsi sebagai alat untuk memprediksi status kelulusan mahasiswa secara real-time atau berdasarkan data masukan. Implementasi ini bertujuan untuk menyediakan Sistem Peringatan Dini (Early Warning System) guna membantu institusi pendidikan dalam memberikan intervensi dini dan dukungan yang tepat sasaran kepada siswa yang diprediksi berisiko tidak lulus, sehingga diharapkan dapat meningkatkan tingkat kelulusan secara keseluruhan.

3. HASIL DAN PEMBAHASAN

Dalam bagian ini, topik penelitian dibahas dan hasilnya disajikan. Hasil pengujian model Random Forest untuk memprediksi tingkat kelulusan siswa berdasarkan variabel akademik dan demografi juga disajikan di bagian ini. Untuk mendapatkan pemahaman yang lebih baik tentang efektivitas metode yang digunakan, hasil evaluasi kinerja model dinilai menggunakan metrik akurasi, *presisi*, *recall*, dan F1-Score. Hasil ini kemudian dibandingkan dengan penelitian yang relevan.

3.1 Analisis Hasil Model

Visualisasi data awal menunjukkan distribusi keseluruhan mahasiswa yang lulus dan tidak lulus dalam dataset. Grafik ini memberikan konteks untuk tugas klasifikasi yang dilakukan oleh model.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



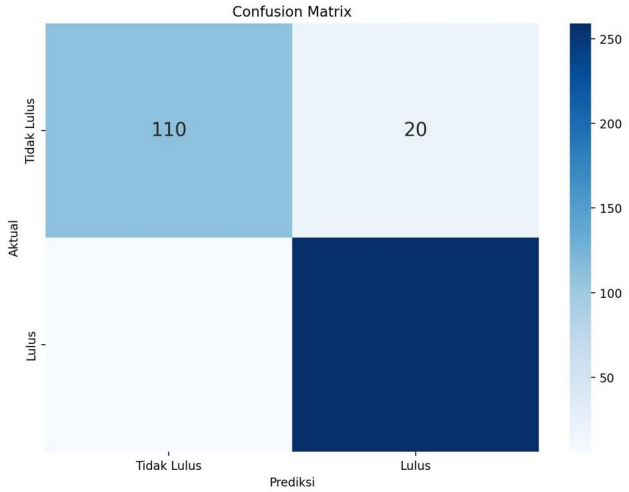
Gambar 2. Jumlah mahasiswa Lulus & Tidak Lulus

Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa prediksi yang sangat baik dalam membedakan siswa yang berpotensi lulus dan siswa yang tidak lulus. Hasil ini menunjukkan bahwa model tidak hanya mampu melakukan klasifikasi dengan baik, tetapi juga konsisten dalam memprediksi kelas positif dan negatif.

Studi sebelumnya menunjukkan bahwa Random Forest seringkali memiliki hasil yang lebih baik dalam memprediksi kelulusan mahasiswa. Misalnya, studi oleh Nguyen et al. menemukan algoritma Random Forest memiliki akurasi sebesar 89,5%.[14]. Studi lain juga menemukan bahwa algoritma boosting seperti XGBoost memiliki akurasi hingga 92% pada dataset mahasiswa. Selain itu, penelitian kontemporer menunjukkan bahwa algoritma Random Forest memiliki kemampuan untuk menangani variabel kategorikal dan numerik secara bersamaan, serta memiliki kekuatan yang kuat untuk menangani *overfitting*[14]. ini menjadi alasan kuat untuk memilih algoritma ini untuk penelitian ini.

3.2 Evaluasi Kinerja Klasifikasi

Pada bagian ini, kinerja model *Random Forest* dievaluasi secara mendalam. Berdasarkan pengujian, model menghasilkan *Confusion Matrix* sebagai berikut:



Gambar 3. Confusion Matrix Model Random Forest

Berdasarkan Gambar 3, model menunjukkan performa yang sangat kuat dengan berhasil mengklasifikasikan 110 data mahasiswa "Tidak Lulus" dan 259 data mahasiswa "Lulus" secara akurat. Kesalahan prediksi (*False Positive* dan *False Negative*) yang sangat rendah mengindikasikan bahwa model memiliki tingkat bias yang minim.

Performa ini divalidasi lebih lanjut melalui metrik evaluasi yang dirangkum dalam Tabel 4

Tabel 4. Hasil Evaluasi Performa Model.

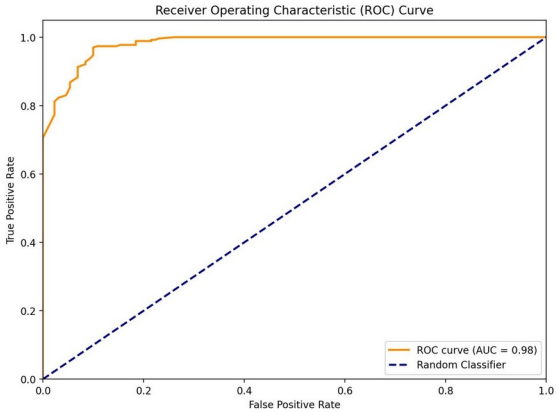
Model	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	94,44%	0,977	0,977	0,952	0,980

Selain metrik tabel, kemampuan model dalam membedakan antar kelas diuji melalui kurva AUC-ROC.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 4. Kurva AUC-ROC

Kurva ROC pada Gambar Y menunjukkan nilai AUC sebesar 0,98. Nilai yang mendekati 1.0 ini membuktikan bahwa model memiliki probabilitas yang sangat tinggi untuk memisahkan mahasiswa yang berisiko terlambat lulus dengan mereka yang akan lulus tepat waktu secara benar.

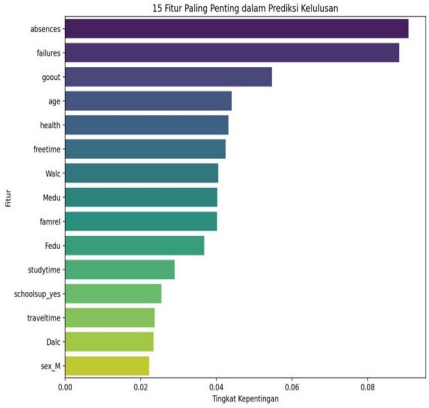
3.3 Perbandingan dengan Penelitian Terdahulu

Hasil model menunjukkan relevansi yang kuat dengan penelitian sebelumnya tentang prediksi kelulusan mahasiswa. Sebagai contoh, penelitian yang dilakukan di fakultas sains dan teknologi Universitas PGRI Sumatera Barat menemukan bahwa Random Forest memiliki akurasi kompetitif sebesar 94% dibandingkan dengan algoritma lain seperti SVM dan Naive Bayes[15].

Selain itu, model Random Forest telah terbukti berhasil dalam penelitian lain yang memprediksi prestasi akademik mahasiswa dengan sangat akurat dan stabil pada dataset besar. Secara umum, metode kelompok seperti Random Forest dan Gradient Boosting dianggap lebih baik daripada metode tunggal seperti Decision Tree atau KNN, terutama ketika jumlah fitur dan kompleksitas data meningkat [14].

3.4 Implementasi Temuan

Model Random Forest memungkinkan kita untuk mengekstrak tingkat kepentingan (*feature importance*) dari setiap variabel. Grafik berikut menunjukkan 15 fitur teratas yang paling berpengaruh dalam menentukan prediksi kelulusan mahasiswa.



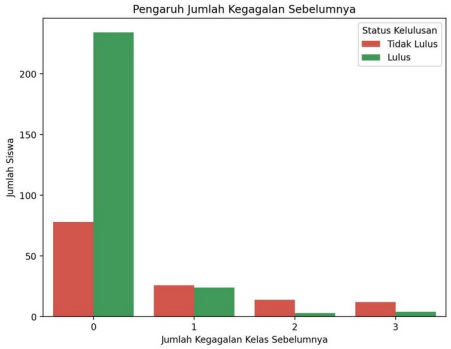
Gambar 5. Fitur Paling penting dalam prediksi kelulusan

Hasil implementasi menunjukkan bahwa elemen seperti nilai akademik, jumlah kegagalan mata kuliah, dan dukungan pendidikan keluarga sangat penting untuk menentukan kelulusan mahasiswa. Temuan ini selaras dengan temuan dari dataset pendidikan Portugal yang digunakan sebagai acuan. Seperti yang ditunjukkan pada Gambar 3.2, Fitur *failures* (kegagalan sebelumnya) dan *absences* (ketidakhadiran) adalah prediktor terkuat. Grafik di bawah ini secara jelas mengilustrasikan bagaimana jumlah kegagalan sebelumnya berdampak drastis pada status kelulusan.s

This is an open access article under the CC–BY-SA license

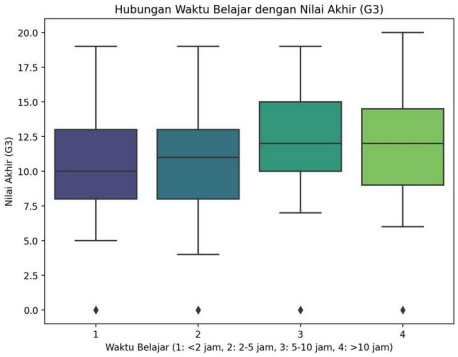
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



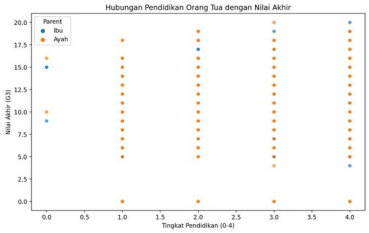
Gambar 6. Pengaruh Jumlah Kegagalan Sebelumnya

Hal ini menguatkan keyakinan bahwa faktor akademik masih menjadi penentu utama keberhasilan mahasiswa dalam pendidikan mereka. Selain itu, *studytime* (waktu belajar) juga muncul sebagai fitur penting. Hubungannya dengan nilai akhir (G3) dapat dilihat di bawah ini, di mana median nilai cenderung meningkat seiring dengan bertambahnya waktu belajar.



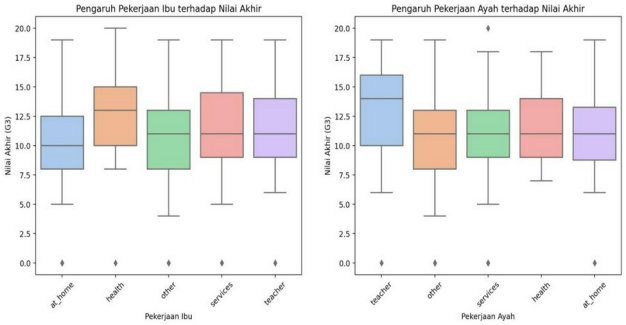
Gambar 7. Hubungan Waktu Belajar dengan Nilai Akhir (G3)

Faktor Keluarga dan Sosial, Namun demikian, faktor non-akademik seperti kondisi sosial dan demografi juga mempengaruhi kinerja model. dukungan pendidikan keluarga, yang dapat dianalisis melalui tingkat pendidikan dan pekerjaan orang tua.



Gambar 8. Hubungan Pendidikan Orang Tua dengan Nilai Akhir

Temuan yang paling menonjol dari visualisasi ini bukanlah adanya korelasi linear yang kuat, melainkan tingginya variabilitas pencapaian mahasiswa pada setiap tingkat pendidikan orang tua. Pada setiap level pendidikan orang tua (dari 1 hingga 4), kita menyaksikan spektrum pencapaian yang penuh: mulai dari mahasiswa yang menghadapi kesulitan besar (nilai mendekati 0) hingga mahasiswa yang meraih prestasi puncak (nilai mendekati 20). Pola ini secara visual membantah gagasan bahwa latar belakang pendidikan keluarga adalah sebuah takdir (deterministik)



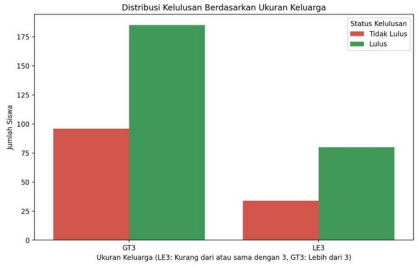
Gambar 9. Pengaruh Pekerjaan Ibu dan Ayah terhadap Nilai Akhir

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 143-151
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Faktor keluarga lainnya, seperti ukuran keluarga (famrel pada Gambar 3.2), juga teridentifikasi sebagai prediktor.



Gambar 10. Distribusi Kelulusan Berdasarkan Ukuran Keluarga

Temuan ini mengindikasikan adanya tren marginal namun signifikan di mana mahasiswa yang berasal dari keluarga kecil (LE3) memiliki proporsi kelulusan yang sedikit lebih baik daripada mahasiswa dari keluarga besar (GT3).

Secara humanis dan sosiologis, temuan ini dapat dikaitkan dengan teori dilusi sumber daya (*resource dilution theory*). Hipotesis ini menyatakan bahwa dalam keluarga yang lebih besar, sumber daya orang tua baik yang bersifat material (finansial, ruang belajar) maupun non-material (waktu, perhatian, bimbingan akademik) secara teoretis harus terbagi di antara lebih banyak anak. Pembagian sumber daya ini berpotensi mengurangi jumlah dukungan individual yang dapat diterima oleh setiap anak, yang pada gilirannya dapat berdampak tipis pada pencapaian akademik mereka. Ini sejalan dengan penelitian yang menyatakan bahwa kombinasi data akademik dan faktor eksternal mahasiswa diperlukan untuk prediksi kelulusan yang akurat[15].

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma *Random Forest* sangat efektif dalam memprediksi tingkat kelulusan mahasiswa dengan mencapai tingkat akurasi sebesar 94,44% pada data uji. Hasil tersebut menjawab pertanyaan penelitian mengenai keandalan model klasifikasi, di mana nilai AUC sebesar 0,98 serta nilai presisi dan *recall* masing-masing sebesar 0,97 membuktikan bahwa kombinasi variabel akademik dan demografi memiliki hubungan yang sangat kuat dan signifikan dengan kemungkinan kelulusan mahasiswa. Kontribusi ilmiah utama dari penelitian ini terletak pada integrasi model prediksi ke dalam aplikasi web interaktif berbasis *framework* Streamlit sebagai Sistem Peringatan Dini (*Early Warning System*) yang praktis. Kehadiran sistem ini memberikan nilai tambah dibandingkan studi terdahulu karena memungkinkan institusi pendidikan melakukan identifikasi dan intervensi secara *real-time* dan humanis. Meskipun demikian, penelitian ini memiliki keterbatasan karena menggunakan data sekunder yang bersifat statis dari satu sumber dataset. Oleh karena itu, pengembangan di masa mendatang disarankan untuk menggunakan teknik optimasi seperti *Particle Swarm Optimizer* (PSO) atau *Stacking Ensemble* serta melakukan pengujian pada skala dataset yang lebih luas guna meningkatkan ketahanan klasifikasi.

UCAPAN TERIMAKASIH

Penulis memanjatkan puji syukur kehadirat Allah Subhanahu Wa Ta'ala atas segala rahmat dan karunia-Nya, yang telah melimpahkan kekuatan, kesabaran, dan petunjuk, sehingga penelitian berjudul "Aplikasi Prediksi Tingkat Kelulusan Mahasiswa Berdasarkan Data Akademik dan Demografi Menggunakan Algoritma Klasifikasi Random Forest" ini dapat diselesaikan dengan baik.

Shalawat dan salam senantiasa tercurah kepada junjungan kita Nabi Muhammad Shallallahu 'Alaihi Wasallam, sebagai suri tauladan terbaik dan pembawa risalah kebenaran.

Penghargaan dan terima kasih yang tulus disampaikan kepada:

1. **Orang Tua Saya**, atas dukungan moral, doa tak terhingga, dan pengorbanan yang menjadi sumber motivasi utama penulis selama menempuh studi dan melaksanakan penelitian.
2. **Bapak Muhammad Ifan Rifani Ihsan, S.Kom., M.Kom.**, atas arahan, bimbingan, dan motivasi yang telah diberikan selama proses penelitian berlangsung.
3. **Ibu Riski Annisa, S.Kom., M.Kom.**, atas kontribusi ilmiah, masukan yang konstruktif, serta waktu yang diluangkan dalam membimbing dan mengarahkan penulisan artikel ini.
4. Pihak **Jurnal Media Informatika [JUMIN]**tempat artikel ini dipublikasikan serta tim **Editor dan *Reviewer*** yang telah bekerja keras memastikan kualitas ilmiah dan kelayakan publikasi naskah ini.
5. **Tim Pembantu Penelitian**, atas dedikasi, kerjasama, dan bantuan teknis yang sangat berharga dalam pengumpulan dan pemrosesan data, sehingga seluruh tahapan penelitian dapat berjalan lancar.

REFERENCES

[1] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, “Predicting University Student Graduation Using Academic

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**

Performance and Machine Learning: A Systematic Literature Review," *IEEE Access*, vol. 12, no. January, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

[2] D. Cahyani, D. Rachmatullah, D. Akbar, H. Malik, K. A. Akhmallahudin, and M. Fansyuri, "Data Mining Menggunakan Algoritma Decision Tree untuk Menentukan Kelulusan Mahasiswa dengan RapidMiner," vol. 1, no. 1, pp. 113–117, 2025, [Online]. Available: file:///C:/Users/User/Downloads/Artikel_DataMiningMenggunakanAlgoritmaDecisionTreeuntukMenentukanKelulusanMahasiswadenganRapidMiner_JITIE.pdf

[3] J. Kuswanto and L. Hakim, "Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 262–270, 2025, doi: 10.51454/decode.v5i1.1103.

[4] A. Asro, J. Chaidir, C. Cahairuddin, and J. Friadi, "Evaluasi Kinerja Algoritma Klasifikasi dalam Studi Kasus Prediksi Kelulusan di Universitas XYZ," *Zo. Tek. J. Ilm.*, vol. 19, no. 1, pp. 15–22, 2025, doi: 10.37776/zt.v19i1.1674.

[5] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG J.*, vol. 1, no. 3, pp. 237–253, 2023.

[6] K. B. Pso, I. Irawan, M. R. Qisthiano, M. Syahril, and P. M. Jakak, "Optimasi Prediksi Kelulusan Tepat Waktu : Studi Perbandingan Algoritma Random Forest dan Algoritma," vol. 4, no. 4, pp. 26–36, 2023.

[7] F. Budi Kurniawan and L. Farokhah, "Aplikasi Cerdas Prediksi Kelulusan Mahasiswa Berbasis Website Menggunakan Metode Support Vector Machine (SVM)," *J. Fasilkom*, vol. 15, no. 1, pp. 155–162, 2025, doi: 10.37859/jf.v15i1.8767.

[8] J. M. Polgan *et al.*, "Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi," vol. 13, pp. 2364–2374, 2025.

[9] R. Agustia, A. A. Supianto, and N. H. Wardani, "Aplikasi Data Mining menggunakan Algoritme Naive Bayes untuk Memprediksi Ketepatan Waktu Lulus Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 2548–964, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[10] H. Andrianof, A. P. Gusman, O. A. Putra, and A. Info, "Jurnal Sains Informatika Terapan (JSIT)," pp. 24–28, 2025.

[11] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.

[12] D. Septhya, K. Rahayu, S. Rabbani, and V. Fitria, "Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," vol. 3, no. April, pp. 15–19, 2023.

[13] Y. Yati, M. Ainol Yaqin, and A. Yusrotun Nadhiroh, "Application of The Support Vector Machine Algorithm for Timely Student Graduation Prediction Based on Streamlit Web at The Faculty of Informatics Engineering Nurul Jadid University," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 3, pp. 1066–1070, 2024, doi: 10.47709/cnahpc.v6i3.3918.

[14] J. Jurnal and T. Informasi, "Prediksi kelulusan mahasiswa menggunakan machine learning," vol. 16, no. 2, pp. 36–39, 2025.

[15] S. Junaidi, R. V. Anggela, and D. Kariman, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes , Random Forest , Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN)," vol. 5, no. 1, pp. 109–119, 2024.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**07/11/2025**; Accepted: **26/01/2026**; Published: **28/03/2026**