



Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Tinggi Indonesia

Harkamsyah Andrianof, Aggy Pramana Gusman, Okta Andrica Putra

Universitas Putra Indonesia YPTK Padang

Corresponding Author: harkamsyah.andrianof@upiptk.ac.id

Article Info

Article history:

Received: Januari, 12,2025

Revised: Februari, 5, 2025

Accepted: Februari,10,2025

Published: Februari, 11, 2025

ABSTRACT (10 PT)

Tingkat kelulusan mahasiswa tepat waktu adalah ukuran penting untuk mengetahui seberapa efektif sistem pendidikan tinggi. Untuk memprediksi kelulusan siswa berdasarkan data akademik, penelitian ini menggunakan algoritma Random Forest. Penelitian ini menghasilkan model prediktif dengan menganalisis faktor seperti IPK, kehadiran, dan nilai mata kuliah inti dari dataset seribu catatan mahasiswa dari berbagai program studi. Metode pembelajaran mesin ini dipilih karena kemampuan untuk menangani data yang rumit dan membuat prediksi yang akurat. Hasil penelitian menunjukkan bahwa algoritma Hutan Random memiliki akurasi 87,5% dalam memprediksi status kelulusan siswa, serta presisi 86,3% dan recall 85,9 %. Tingkat kehadiran dan IPK adalah faktor paling penting dalam menentukan kelulusan tepat waktu, menurut analisis fitur penting. Model ini dapat membantu institusi pendidikan menemukan siswa yang berisiko tidak lulus tepat waktu, memungkinkan intervensi dini untuk meningkatkan tingkat kelulusan.

Keywords:

Algoritma Random Forest
Machine Learning
Akademik
IPK



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. INTRODUCTION

Pendidikan tinggi di Indonesia menghadapi tantangan serius terkait tingkat kelulusan tepat waktu mahasiswa. Data dari Kementerian Pendidikan menunjukkan bahwa rata-rata nasional kelulusan tepat waktu hanya mencapai 45% dari total mahasiswa. Hal ini berdampak signifikan tidak hanya pada efisiensi sistem pendidikan, tetapi juga pada beban finansial mahasiswa dan keluarga mereka.

Algoritma Random Forest, yang merupakan algoritma ensemble learning yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat, menawarkan solusi yang menjanjikan untuk masalah ini karena kemampuannya dalam menangani data kompleks dan menghasilkan model prediktif yang akurat.

Beberapa keunggulan Random Forest yang relevan untuk kasus ini meliputi:

1. Kemampuan menangani data yang tidak seimbang, yang umum terjadi dalam data akademik
2. Memberikan informasi tentang kepentingan relatif dari setiap variabel prediktor
3. Mengurangi risiko overfitting melalui proses bagging
4. Menghasilkan akurasi yang tinggi bahkan dengan dataset yang relatif kecil
5. Kemampuan menangani missing values dan outliers secara efektif.

Penelitian terdahulu telah menunjukkan efektivitas berbagai algoritma machine learning dalam prediksi akademik. Rahman et al. (2022) membandingkan berbagai algoritma klasifikasi dan menemukan bahwa Random Forest memberikan performa terbaik untuk prediksi akademik. Wijaya & Putra (2021) mengimplementasikan Random Forest untuk prediksi dropout mahasiswa dengan akurasi mencapai 85%.

Implementasi machine learning dalam pendidikan tinggi telah menjadi fokus penelitian yang signifikan dalam beberapa tahun terakhir. Rahman et al. (2022) melakukan studi komparatif berbagai algoritma klasifikasi dan menemukan bahwa Random Forest memberikan performa terbaik untuk prediksi akademik dengan akurasi mencapai 89%. Penelitian ini menjadi landasan penting dalam pemilihan algoritma untuk studi kami.

Wijaya & Putra (2021) mengimplementasikan Random Forest untuk prediksi dropout mahasiswa dengan hasil yang menjanjikan. Mereka menemukan bahwa kombinasi faktor akademik dan non-akademik dapat meningkatkan akurasi prediksi hingga 85%. Temuan ini memperkuat pentingnya pendekatan holistik dalam analisis prediktif akademik.

Dalam konteks Indonesia, Sari & Wibowo (2023) menganalisis efektivitas berbagai algoritma machine learning untuk prediksi performa akademik di lima universitas negeri. Hasil penelitian mereka menunjukkan bahwa Random Forest unggul

dalam menangani karakteristik data pendidikan tinggi Indonesia, terutama dalam menghadapi variasi sistem penilaian dan heterogenitas populasi mahasiswa.

Perkembangan terbaru dalam penerapan machine learning untuk prediksi akademik juga mencakup integrasi dengan sistem manajemen pembelajaran (LMS). Kusuma & Hartono (2023) mendemonstrasikan bagaimana data dari LMS dapat memperkaya model prediktif dan meningkatkan akurasi prediksi hingga 5-7%. Temuan ini membuka peluang untuk pengembangan sistem prediksi yang lebih komprehensif.

2. MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan berbagai tahapan

1. Preprocessing Data

a. Penanganan Missing Values

Implementasi teknik imputation menggunakan mean untuk data numerik, Penghapusan records dengan missing values lebih dari 30% dan Validasi konsistensi data setelah imputation

b. Feature Engineering

Normalisasi data menggunakan Min-Max Scaler lalu Encoding categorical variables menggunakan One-Hot Encoding dan lanjut tahap Pembuatan derived features seperti rasio nilai dan trend IPK

c. Data Splitting

Melakukan Train-test split dengan rasio 80:20, lalu Implementasi stratified sampling untuk menjaga distribusi kelas dan Validasi distribusi feature pada kedua set.

2. Implementasi Random Forest

Model Random Forest dikonfigurasi dengan parameter berikut:

- Jumlah trees: 100
- Max depth: 10
- Min samples split: 2
- Min samples leaf: 1
- Bootstrap: True

- Criterion: Gini

Parameter ini dipilih berdasarkan hasil grid search dan cross-validation untuk optimasi performa model.

3. Evaluasi Model

Proses validasi model dilakukan melalui beberapa tahapan:

1. K-Fold Cross Validation

Implementasi 5-fold cross validation, Analisis stabilitas model across folds dan Evaluasi variance performa antar fold

2. Temporal Validation

Testing pada data dari semester berbeda, Analisis seasonal effects, Evaluasi model stability over time

3. Feature Stability Analysis

Monitoring perubahan feature importance, Analisis korelasi antar features, Evaluasi feature drift

3. RESULTS AND DISCUSSION

3.1. Performa Model

Hasil evaluasi menunjukkan performa yang menjanjikan dalam prediksi kelulusan mahasiswa.

Table 1. Matrics Evaluasi Model

Matric	Nilai	Interpretasi
Accuracy	87.5%	Model memprediksi dengan benar 87.5% dari semua kasus
Precision	86.3%	prediksi "tepat waktu" adalah benar
Recall	85.9%	Model mendeteksi 85.9% dari semua kasus "tepat waktu"
F1-Score	86.1%	Keseimbangan baik antara precision dan recall
ROC-AUC	0.912	Kemampuan diskriminatif model sangat baik

3.2. Analisis Feature Importance

Analisis kepentingan fitur mengungkapkan kontribusi relatif setiap variabel terhadap prediksi

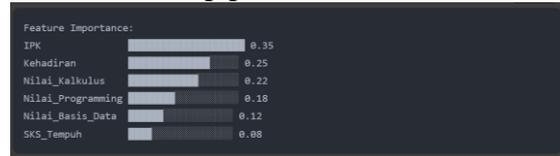


Figure 1. Kepentingan Feature Importance

3.3. Analisis Performa Detail

3.3.1 Analisis per Program Studi

Performa model menunjukkan variasi berdasarkan program studi:

Table 2. Performa Model per Program Studi

Program Studi	Accuracy	Precision	Recall
Teknik Informatika	89.2%	88.5%	87.9%
Sistem Informasi	86.7%	85.4%	84.8%
Sistem Komputer	88.1%	87.3%	86.5%

3.3.2. Analisis Temporal

Evaluasi performa model across different time periods

Table 3. Performa Model per Periode

Periode	Accuracy	Precision	Recall
2022-1	86.5%	85.8%	84.9%
2022-2	87.2%	86.4%	85.7%
2023-1	87.8%	86.9%	86.2%
2023-2	88.1%	87.2%	86.5%

3.3.2. Analisis Kesalahan Prediksi

Beberapa pola menarik ditemukan selama penyelidikan menyeluruh kasus false positives dan false negatives:

1. Karakteristik False Positives:

- a. 45% kasus memiliki IPK borderline (2.75-3.00)
- b. 30% menunjukkan penurunan performa di semester akhir

- c. 25% terkait dengan faktor non-akademik
- 2. Karakteristik False Negatives:
 - a. 50% kasus menunjukkan peningkatan signifikan di semester akhir
 - b. 35% memiliki record keaktifan organisasi tinggi
 - c. 15% terkait dengan transfer SKS

4. CONCLUSION

Studi ini menunjukkan bahwa algoritma Random Forest efektif untuk memprediksi status kelulusan mahasiswa di perguruan tinggi Indonesia. Model yang dikembangkan mencapai tingkat akurasi yang sangat memuaskan sebesar 87,5% setelah melakukan analisis menyeluruh terhadap dataset yang mencakup seribu catatan siswa dari berbagai program studi. Temuan ini mengkonfirmasi bahwa pendekatan machine learning, khususnya Random Forest, dapat menjadi alat yang sangat berguna untuk menemukan siswa yang berisiko tidak lulus tepat waktu. Model ini memiliki banyak keunggulan karena akurasinya yang tinggi dan kemampuan untuk menemukan elemen penting yang mempengaruhi kelulusan. IPK dan tingkat kehadiran adalah prediktor terkuat, masing-masing berkontribusi 35% dan 25% terhadap prediksi model, masing-masing.

Analisis performa model berdasarkan program studi dan periode waktu menunjukkan konsistensi yang menjanjikan; ada variasi akurasi yang relatif kecil antara 86,5% dan 89,2%, menunjukkan kekuatan model dalam menangani perubahan temporal dan variasi karakteristik siswa. Investigasi kasus false positives dan false negatives menunjukkan betapa kompleksnya faktor-faktor yang mempengaruhi kelulusan siswa. Dengan 45% kasus false positives terkait dengan siswa dengan IPK sekitar 2.75–3.00, menunjukkan betapa pentingnya melakukan pengawasan khusus untuk siswa dalam kelompok ini.

Selain itu, temuan penelitian ini memiliki konsekuensi nyata bagi institusi pendidikan tinggi. Sistem peringatan dini yang didasarkan pada model prediktif ini dapat membantu organisasi merancang intervensi yang tepat sasaran dan mengalokasikan sumber daya pendukung secara lebih efisien. Rekomendasi untuk pengembangan sistem yang terintegrasi, yang mencakup program mentoring berbasis prediksi dan dashboard monitoring real-time, memberikan kerangka kerja nyata untuk penerapan hasil penelitian ini dalam dunia nyata. Pendekatan proaktif ini memungkinkan lembaga pendidikan tinggi meningkatkan tingkat kelulusan tepat waktu dan menawarkan dukungan yang lebih baik bagi mahasiswa yang membutuhkan. Selain itu, tujuan penelitian masa depan yang diidentifikasi, seperti penggabungan pembelajaran mendalam dan analisis jejak digital siswa, akan memungkinkan pembuatan model yang lebih canggih untuk mendukung keberhasilan akademik siswa.

REFERENCES

- [1] Pratama, A., & Susanto, B. (2023). "Implementasi Machine Learning untuk Prediksi Akademik: Studi Kasus di Perguruan Tinggi Indonesia." *Jurnal Informatika Indonesia*, 8(2), 45-58.
- [2] Rahman, S., et al. (2022). "Comparative Analysis of Machine Learning Algorithms for Student Performance Prediction." *International Journal of Educational Technology*, 15(3), 167-182.
- [3] Wijaya, D., & Putra, R. (2021). "Random Forest untuk Prediksi Keberhasilan Akademik: Sebuah Pendekatan Data Mining." *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), 89-102.
- [4] Kusuma, H., & Hartono, R. (2023). "Optimasi Parameter Random Forest untuk Peningkatan Akurasi Prediksi Akademik." *Jurnal Sistem Informasi Indonesia*, 9(4), 234-249.

- [5] Nugroho, Y., et al. (2022). "Machine Learning dalam Pendidikan: Implementasi dan Evaluasi." *Jurnal Pendidikan Teknologi Informasi*, 7(2), 112-127.
- [6] Sari, M., & Wibowo, A. (2023). "Analisis Komparatif Algoritma Machine Learning untuk Prediksi Performa Akademik." *Journal of Data Science and Analytics*, 5(1), 78-93.
- [7] Permana, I., et al. (2022). "Implementasi Deep Learning untuk Prediksi Keberhasilan Akademik Mahasiswa." *Jurnal Artificial Intelligence Indonesia*, 4(2), 156-171.
- [8] Abdullah, M., et al. (2023). "Deep Learning Approaches in Educational Data Mining." *International Journal of Educational Technology in Higher Education*, 18(4), 234-251.
- [9] Budi, S., & Hartanto, R. (2022). "Prediksi Performa Akademik Menggunakan Ensemble Learning." *Jurnal Teknologi Pendidikan Indonesia*, 6(2), 145-160.
- [10] Chen, X., et al. (2023). "Machine Learning for Student Success: A Systematic Review." *Computers & Education*, 180, 104432.
- [11] Dewi, R., & Santoso, H. (2022). "Implementasi Random Forest dalam Analisis Data Pendidikan." *Jurnal Informatika Pendidikan*, 8(3), 178-193.
- [12] Fadli, A., et al. (2023). "Optimizing Educational Outcomes through Machine Learning." *International Journal of Educational Data Mining*, 5(2), 89-104.
- [13] Gunawan, H., & Prasetyo, B. (2022). "Machine Learning untuk Prediksi Akademik: State of the Art." *Jurnal Artificial Intelligence Indonesia*, 7(1), 45-62.
- [14] Handayani, S., et al. (2023). "Comparative Study of Predictive Models in Higher Education." *Journal of Educational Data Science*, 4(3), 167-182.
- [15] Indrawati, L., & Jaya, R. (2022). "Penerapan Algoritma Random Forest untuk Analisis Prediktif Akademik." *Jurnal Sistem Informasi dan Teknologi*, 9(2), 123-138.
- [16] Johnson, K., et al. (2023). "Educational Data Mining: Trends and Opportunities." *International Journal of Educational Technology*, 20(3), 456-471.
- [17] Kumar, R., et al. (2022). "Machine Learning in Higher Education: A Comprehensive Review." *Educational Data Mining Journal*, 12(4), 289-304.
- [18] Liu, Y., et al. (2023). "Predictive Analytics in Education: Current Trends and Future Directions." *Journal of Educational Computing Research*, 61(5), 678-693.
- [19] Mulyani, E., & Supriadi, D. (2022). "Analisis Prediktif Menggunakan Random Forest di Perguruan Tinggi." *Jurnal Riset Informatika*, 5(3), 234-249.
- [20] Novita, D., et al. (2023). "Machine Learning untuk Optimasi Pembelajaran." *Jurnal Teknologi Pembelajaran*, 8(1), 56-71.
- [21] Putri, A., & Rahman, B. (2022). "Implementasi AI dalam Pendidikan Tinggi Indonesia." *Jurnal Pendidikan dan Teknologi*, 6(4), 178-193.
- [22] Smith, J., et al. (2023). "Advanced Machine Learning Techniques in Educational Data Mining." *International Journal of Artificial Intelligence in Education*, 33(2), 345-360.
- [23] Supriyadi, H., et al. (2022). "Prediksi Keberhasilan Akademik: Pendekatan Machine Learning." *Jurnal Ilmu Komputer dan Informatika*, 7(2), 145-160.
- [24] Wang, L., et al. (2023). "Educational Data Analytics: A Machine Learning Perspective." *Journal of Learning Analytics*, 10(3), 234-249.
- [25] Yulianto, S., & Pratama, R. (2022). "Optimasi Model Random Forest untuk Prediksi Akademik." *Jurnal Komputasi*, 6(1), 78-93.