

Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi

¹Marwa Sulehu, ²Wisda, ³First Wanita, ⁴Markani,
^{1,2,3,4}Universitas Teknologi Akba Makassar

¹marwasulehu@akba.ac.id, ²wisda@akba.ac.id, ³firstwanita@akba.ac.id, ⁴markani@akba.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model prediksi berbasis algoritma Random Forest guna mengidentifikasi mahasiswa yang berisiko putus kuliah di Universitas Teknologi AKBA Makassar, di mana tingginya angka putus kuliah menjadi tantangan signifikan yang disebabkan oleh berbagai faktor akademik, sosial-ekonomi, dan psikologis. Dengan pendekatan kuantitatif melalui metode data mining, data dari sistem informasi akademik dan survei primer, menggunakan 1.425 data mahasiswa yang diolah melalui tahap preprocessing dan dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Hasil penelitian menunjukkan bahwa IPK, motivasi belajar, dan kehadiran merupakan variabel paling signifikan dalam memprediksi kelulusan, dengan model prediksi yang mencapai akurasi 70% serta performa precision dan recall yang memadai. Penelitian ini memberikan kontribusi penting dalam mendukung pengambilan keputusan strategis di institusi pendidikan tinggi untuk meningkatkan tingkat retensi mahasiswa dan mengurangi angka putus kuliah, meskipun masih terdapat ruang untuk peningkatan kinerja model, khususnya dalam menangani ketidakseimbangan data.

Kata Kunci: Retensi Mahasiswa, Random Forest, Prediksi Kelulusan, Data Mining, Pendidikan Tinggi.

PENDAHULUAN

Untuk menjaga retensi mahasiswa dalam menyelesaikan studi dengan menghindari putus kuliah, dimana putus kuliah mahasiswa dapat disebabkan oleh berbagai faktor yang saling terkait, penyebab utama kegagalan mahasiswa dalam akademik yaitu 1) gagalnya membangun integrasi akademik dan sosial, hal ini sering kali menimbulkan perasaan terisolasi, 2) kondisi psikologis, seperti pola pikir fixed mindset yang memandang kemampuan sebagai hal yang tidak dapat berubah, 3) kurangnya motivasi diri dan stabilitas emosional, turut meningkatkan risiko mahasiswa untuk berhenti kuliah, 4) Atribusi dan persepsi negatif terhadap keberhasilan, seperti menganggap tantangan berada di luar kendali mereka, juga menjadi pemicu utama. Faktor lain yang berkontribusi adalah 5) karakteristik individual, seperti rendahnya tingkat agreeableness, conscientiousness dan kestabilan emosional, yang memengaruhi daya tahan mereka terhadap tekanan, 6) masalah kinerja akademik yang berdampak pada kepercayaan diri, serta 7) kurangnya dukungan sosial dari keluarga, teman, dan institusi, memperburuk situasi ini. Lebih lanjut, 8) kebijakan perguruan tinggi yang kurang mendukung, seperti minimnya bimbingan akademik atau konseling, menjadi hambatan signifikan dalam mempertahankan mahasiswa hingga lulus (Moesarofah 2021). Dengan memahami faktor-faktor ini, perguruan tinggi dapat mengambil langkah strategis untuk meningkatkan retensi dan mengurangi angka putus kuliah.

Masalah keterbatasan layanan konseling di Perguruan Tinggi diperparah oleh rendahnya kesadaran mahasiswa terhadap kebutuhan bimbingan. Banyak mahasiswa yang tidak menyadari bahwa mereka sedang menghadapi masalah yang memerlukan pendampingan, sehingga cenderung

menganggapnya sebagai hal yang biasa dan tidak memerlukan intervensi. Selain itu, tidak semua perguruan tinggi memiliki unit layanan bimbingan dan konseling yang memadai, baik dari segi infrastruktur maupun sumber daya manusia. Jumlah konselor profesional yang tersedia sering kali tidak mencukupi untuk memenuhi kebutuhan mahasiswa yang beragam, mengakibatkan banyak permasalahan psikososial yang tidak tertangani secara optimal. Kondisi ini semakin memperbesar risiko kegagalan akademik dan meningkatkan angka putus kuliah di kalangan mahasiswa (Azhari dkk. t.t.)

Berbagai referensi menawarkan solusi untuk mengatasi masalah putus kuliah dan menjaga retensi mahasiswa dalam menyelesaikan studi melalui pendekatan yang komprehensif dan inovatif diantaranya 1) penyediaan layanan bimbingan akademik dan konseling yang komprehensif dapat membantu mahasiswa mengatasi kesulitan belajar dan permasalahan pribadi, sehingga berkontribusi pada peningkatan retensi. Selain itu, 2) penciptaan lingkungan kelas yang inklusif serta kebijakan dukungan akademik yang proaktif mampu meningkatkan keterlibatan mahasiswa dan mencegah mereka dari risiko putus kuliah, 3) pelaksanaan pelatihan yang mendorong pola pikir berkembang juga penting, karena dapat membantu mahasiswa memandang tantangan sebagai peluang untuk belajar dan berkembang. Di sisi lain, 4) keterlibatan orang tua dan komunitas dalam proses pendidikan memberikan dukungan tambahan yang dapat memotivasi mahasiswa untuk menyelesaikan studi mereka (Al Farabi dkk. 2024, Anon 2024; Astuti 2024; Moesarofah 2021).

Keterbatasan tenaga kependidikan dalam menyediakan layanan konseling di perguruan tinggi menjadi tantangan signifikan, terutama ketika institusi menghadapi jumlah mahasiswa yang besar serta ragam permasalahan yang kompleks, yang berdampak langsung pada keberhasilan akademik dan kesejahteraan emosional mahasiswa. Untuk mengatasi isu ini sekaligus menjaga retensi mahasiswa dalam menyelesaikan studi, diperlukan pendekatan yang komprehensif dan inovatif. Penyediaan layanan bimbingan akademik dan konseling yang komprehensif menjadi langkah strategis untuk membantu mahasiswa mengatasi kesulitan belajar dan permasalahan pribadi. Selain itu, penciptaan lingkungan kelas yang inklusif serta kebijakan dukungan akademik yang proaktif dapat meningkatkan keterlibatan mahasiswa dan mencegah risiko putus kuliah. Pelatihan yang mendorong pola pikir berkembang juga penting untuk membantu mahasiswa melihat tantangan sebagai peluang belajar, sementara keterlibatan orang tua dan komunitas memberikan dukungan tambahan yang dapat memotivasi mahasiswa menyelesaikan studi mereka. Pendekatan-pendekatan ini menawarkan solusi konkret untuk mengatasi hambatan dalam mendukung kesejahteraan dan keberhasilan akademik mahasiswa (Azhari dkk. t.t.; Putra, Neviyarni, dan Firman 2023; Suriyanti 2024).

Menyimpulkan permasalahan yang begitu kompleks tersebut yang akan berdampak langsung pada keberhasilan akademik dan kesejahteraan emosional mahasiswa dan memperbesar risiko putus kuliah bila tidak disikapi lebih bijak. Peneliti menggaris bawahi bahwa banyak mahasiswa tidak menyadari bahwa mereka menghadapi masalah yang memerlukan pendampingan, ditambah lagi, tidak semua perguruan tinggi memiliki unit layanan bimbingan dan konseling yang memadai, baik dari segi infrastruktur maupun jumlah konselor yang tersedia, beberapa penelitian juga telah memaparkan solusi dari masalah-masalah tersebut. 1) Permasalahan yang terjadi adalah kapan konseling tersebut dilaksanakan?, 2) kepada siapa saja konseling tersebut dilakukan dan siapa konselornya?. Dalam penelitian ini, algoritma Random Forest digunakan untuk memprediksi mahasiswa yang berisiko mengalami putus kuliah dan variabel apa saja yang menjadi penyebab putus kuliah, sehingga memberikan dasar untuk melakukan intervensi yang terarah. Keterlibatan orang tua dan komunitas juga menjadi faktor pendukung yang penting dalam memotivasi mahasiswa menyelesaikan studi mereka. Dengan pendekatan tersebut, perguruan tinggi dapat mengidentifikasi dan menangani masalah sejak dini, meningkatkan retensi mahasiswa, dan mendukung tercapainya kelulusan tepat waktu, sekaligus mengurangi angka putus kuliah.

Beberapa penelitian tentang retensi dan pemanfaatan algoritma random forest dan masalah retensi mahasiswa,

Penelitian berjudul Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika, penelitian ini mengimplementasikan algoritma Random Forest untuk membantu klasifikasi bidang MSIB mahasiswa Prodi Pendidikan Informatika berdasarkan nilai akademik. Dengan dataset sebanyak 100 data, model menunjukkan performa

yang baik dengan tingkat akurasi sebesar 80%, precision 80%, dan recall 82%. Hasil ini menunjukkan bahwa algoritma Random Forest efektif dalam memetakan bidang MSIB, seperti Multimedia, Rekayasa Perangkat Lunak, Kecerdasan Buatan, dan Teknik Komputer Jaringan, sesuai dengan kemampuan mahasiswa. Meskipun jumlah data yang digunakan terbatas, penelitian ini memberikan kontribusi penting dalam mempermudah mahasiswa memilih bidang yang sesuai dan menjadi dasar untuk pengembangan lebih lanjut dengan dataset yang lebih besar (Aini dkk. 2024).

Berdasarkan penelitian Optimasi Algoritma Random Forest dengan Hyperparameter Tuning Menggunakan GridSearchCV untuk Prediksi Nasabah Churn pada Industri Perbankan, menyatakan implementasi Random Forest yang dioptimalkan menggunakan GridSearchCV berhasil meningkatkan performa model dengan akurasi 89,61%, recall 90,15%, presisi 88,72%, dan F1-Score 89,43%. Proses optimasi dilakukan melalui eksplorasi kombinasi berbagai hyperparameter, seperti jumlah estimators, kedalaman maksimal, dan lainnya, yang membantu memaksimalkan konsistensi model dalam memprediksi nasabah churn. Hasil ini menunjukkan potensi penggunaan teknologi AI yang lebih efektif dalam strategi manajemen churn di sektor perbankan untuk meminimalisir risiko kehilangan nasabah (Amalia t.t.).

Hasil penelitian berjudul Analisis Karakteristik Retensi Mahasiswa di Perguruan Tinggi menunjukkan bahwa retensi mahasiswa tahun kedua di Universitas PGRI Adi Buana Surabaya didominasi oleh faktor akademik dengan nilai rata-rata tertinggi dibandingkan faktor study skill dan keterlibatan dalam kegiatan kampus. Responden umumnya memiliki efikasi diri yang baik, yang tercermin dalam kemampuan menyesuaikan diri dengan tugas akademik dan memenuhi tuntutan prodi. Selain itu, keterampilan mengatur waktu antara kegiatan akademik dan non-akademik serta keterlibatan sosial juga menjadi aspek penting dalam retensi. Temuan ini menegaskan bahwa retensi mahasiswa dipengaruhi oleh faktor akademik, keterampilan belajar, dan keterlibatan sosial, dengan efikasi diri sebagai prediktor utama, serta didukung oleh pola pikir berkembang (growth mindset) yang mendorong ketekunan akademik (Moesarofah 2021).

TINJAUAN PUSTAKA

Retensi Mahasiswa

Retensi dalam konteks pendidikan, khususnya di universitas, merujuk pada kemampuan institusi untuk mempertahankan mahasiswa selama masa studi mereka hingga lulus. Tingkat retensi mengukur persentase mahasiswa yang melanjutkan studi mereka dari tahun ke tahun, atau dari awal kuliah hingga kelulusan dengan tujuan untuk memastikan bahwa mahasiswa tidak keluar atau berhenti di tengah jalan, yang sering disebabkan oleh berbagai faktor seperti masalah akademik, keuangan, atau motivasi pribadi (Moesarofah 2021).

Retensi mahasiswa dipengaruhi oleh interaksi antara faktor akademik, sosial, dan individual. Mahasiswa yang memiliki performa akademik rendah, kurang keterlibatan sosial, atau menghadapi tekanan finansial cenderung memiliki risiko putus kuliah yang lebih tinggi. Faktor-faktor seperti IPK, kehadiran, dan jumlah SKS mencerminkan kinerja akademik mahasiswa, sedangkan motivasi belajar dan tingkat stres menunjukkan aspek psikologis yang memengaruhi keberhasilan studi. Faktor sosial-ekonomi, termasuk dukungan keluarga dan pekerjaan paruh waktu, juga menjadi aspek penting dalam menentukan keberlanjutan studi mahasiswa (Ghozali 2013).

Algoritma Random Forest dalam Prediksi

Algoritma Random Forest diperkenalkan oleh Breiman (2001) adalah metode pembelajaran ensemble berbasis pohon keputusan. Teknik ini terkenal karena keakuratannya dalam menangani dataset yang kompleks, termasuk yang memiliki fitur heterogen seperti dalam penelitian pendidikan tinggi. Keunggulan utama Random Forest meliputi 1) Kemampuan menangani data dengan banyak variabel, 2) Toleransi terhadap missing values dan outlier, dan 3) Menghasilkan evaluasi feature importance, yang berguna untuk menentukan variabel dominan dalam prediksi (Brenab 2001).

Random Forest adalah algoritma pembelajaran mesin berbasis pohon keputusan yang menggabungkan beberapa pohon keputusan untuk membuat prediksi yang lebih akurat. Algoritma ini bekerja dengan membangun banyak pohon keputusan dari subset data yang acak, dan hasilnya digabungkan untuk menghasilkan prediksi yang lebih stabil dan robust. Kelebihan dari Random Forest adalah kemampuannya dalam mengatasi masalah overfitting yang sering terjadi pada pohon keputusan tunggal dan kemampuannya dalam memproses data dengan banyak variabel (Rui dkk. 2022). Dalam penelitian mengenai retensi mahasiswa, Random Forest dapat digunakan untuk memprediksi mahasiswa yang berisiko putus kuliah berdasarkan variabel-variabel seperti nilai akademik, partisipasi dalam kegiatan kampus, dan interaksi sosial.

Dalam penelitian terkait pendidikan, Random Forest telah digunakan untuk memprediksi kelulusan mahasiswa, kinerja akademik, serta risiko putus kuliah. menunjukkan bahwa Random Forest menghasilkan akurasi tinggi dalam memprediksi hasil belajar dibandingkan algoritma lainnya seperti Decision Tree dan Logistic Regression (Ajiwerdhi, Kesiman, dan Wirawan 2012).

Evaluasi Model Prediksi

Evaluasi model prediksi dilakukan dengan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Metrik ini digunakan untuk memastikan bahwa model bekerja secara efektif dalam mendeteksi mahasiswa yang berisiko putus kuliah (Sugiono 2013). Selain itu, teknik validasi seperti cross-validation sering digunakan untuk memastikan generalisasi model pada data baru.

Implementasi Teknologi untuk Meningkatkan Retensi

Integrasi teknologi seperti prediksi berbasis data menggunakan Random Forest dapat membantu institusi pendidikan dalam mengambil tindakan preventif terhadap mahasiswa yang rentan putus kuliah. Dengan analisis yang mendalam terhadap variabel seperti IPK, motivasi belajar, dan status sosial-ekonomi, institusi dapat merancang intervensi yang lebih efektif, seperti pemberian bimbingan akademik, konseling psikologis, atau bantuan finansial.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk memprediksi mahasiswa yang berpotensi putus kuliah. Pendekatan kuantitatif dipilih karena memungkinkan analisis statistik terhadap data numerik yang besar, sehingga dapat mengidentifikasi pola dan hubungan antar variabel yang mempengaruhi risiko putus kuliah. Metode data mining, khususnya algoritma Random Forest, digunakan karena kemampuannya dalam melakukan klasifikasi dan prediksi dengan akurasi tinggi. Studi sebelumnya menunjukkan bahwa Random Forest efektif dalam mengklasifikasikan siswa yang berpotensi putus sekolah (Rofi, Setiawan, dan Riana 2024). Selain itu, algoritma ini mampu menangani data dengan banyak variabel dan menentukan atribut yang berpengaruh signifikan terhadap penurunan nilai mahasiswa (Thoriq, Pranoto, dan Faldi 2023). Dengan demikian, pendekatan ini diharapkan dapat menghasilkan model prediksi yang akurat untuk mendukung upaya pencegahan mahasiswa putus kuliah.

Dengan menggunakan desain penelitian deskriptif prediktif yang bertujuan untuk menjelaskan pola dan hubungan antar variabel yang memengaruhi kemungkinan mahasiswa putus kuliah serta memprediksi kelompok mahasiswa yang rentan. Desain ini cocok digunakan untuk membangun model berbasis data yang dapat menghasilkan prediksi akurat serta memberikan rekomendasi kebijakan berdasarkan analisis data (Sugiono 2013). Desain tersebut dirancang untuk menghasilkan wawasan mendalam mengenai faktor-faktor yang memengaruhi putus kuliah mahasiswa, serta menyediakan model prediksi yang dapat mendukung pengambilan keputusan di tingkat institusi pendidikan tinggi. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi nyata dalam meningkatkan retensi mahasiswa dan mencegah putus kuliah secara dini.

Data yang digunakan adalah data mahasiswa pada semester 4 Universitas Teknologi AKBA Makassar (UNITAMA) yang telah lulus dan putus kuliah dalam rentang lima tahun terakhir (2019 hingga 2024). Sampel penelitian dipilih menggunakan metode purposive sampling.

Pemilihan semester 4 didasarkan pada beberapa pertimbangan strategis. Semester 4 dianggap sebagai titik kritis dalam perjalanan akademik mahasiswa, di mana mereka telah melewati tahap adaptasi awal dan mulai menghadapi tantangan akademik yang lebih spesifik pada program studi mereka. Dengan menganalisis data dari lima tahun terakhir, penelitian ini tidak hanya mencerminkan pola dropout terkini, tetapi juga memungkinkan identifikasi tren dan faktor-faktor yang konsisten memengaruhi tingkat keberlanjutan studi mahasiswa. Pemanfaatan data historis ini memberikan landasan empiris yang kuat untuk validasi hasil penelitian dan meningkatkan relevansi solusi yang diusulkan dalam mengoptimalkan prediksi kelulusan mahasiswa.

Jumlah sampel yang diambil sebanyak 2.000 data mahasiswa untuk memastikan representasi yang memadai dan keakuratan analisis prediktif. Pengambilan sampel ini dilakukan untuk mencakup beragam variabel yang memengaruhi risiko putus kuliah, seperti kinerja akademik, status sosial-ekonomi, dan karakteristik psikologis, namun setelah tahap preprocessing dan pembersihan data, menghasilkan 1.425 data yang dapat dijadikan dataset.

Proses penelitian mencakup tiga tahap utama, yaitu pengumpulan data, pengolahan data dan pengembangan model, 1) data diperoleh dari dua sumber, yaitu pertama, data sekunder yaitu data yang berasal dari sistem informasi akademik Universitas Teknologi Akba Makassar, mencakup IPK, kehadiran, dan jumlah SKS yang diambil dan kedua, data primer, dikumpul menggunakan model kuesioner, wawancara dan observasi, data ini meliputi status sosial-ekonomi, tingkat stres, dan motivasi belajar. 2) Data yang terkumpul diproses melalui tahap preprocessing, termasuk pembersihan data, penanganan missing values, dan transformasi data kategorikal menjadi numerik. Proses ini dilakukan untuk memastikan kualitas data yang digunakan dalam analisis (Ghozali 2013). 3) Algoritma Random Forest digunakan untuk membangun model prediktif. Model ini melibatkan pembagian dataset menjadi data pelatihan (80%) dan data pengujian (20%). Parameter model seperti jumlah pohon ($n_{estimators}$) dan kedalaman pohon (max_depth) dioptimalkan untuk meningkatkan akurasi. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score untuk memastikan kinerjanya (Aini dkk. 2024)

Random Forest menggunakan kombinasi beberapa pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat. Berikut cara kerja Random Forest (Rui dkk. 2022)

1. Bootstrapping (Random Sampling)

Random Forest menggunakan teknik bagging (Bootstrap Aggregating), di mana dataset asli D dengan n sampel dibagi menjadi m subset dengan pengambilan sampel acak yang mengembalikan D_i . Dengan ketentuan

- Setiap subset D_i digunakan untuk melatih satu pohon keputusan.
- Beberapa sampel mungkin muncul lebih dari sekali dalam subset tertentu, sedangkan lainnya tidak muncul sama sekali.

Gunakan persamaan berikut untuk menentukan nilai D_i

$$D = \{X_i, Y_i\}, X_i \sim D$$

X_i adalah data acak yang diambil dari dataset D

Y_i adalah label dari X_i

2. Pembangunan Pohon Keputusan (Decision Tree)

Setiap pohon keputusan dibuat menggunakan subset D_i pada setiap node.

- Memilih fitur secara acak dari seluruh fitur yang tersedia
Jika ada k fitur, hanya fitur yang dipertimbangkan
- Memilih fitur terbaik berdasarkan metrik Gini Index

$$G = 1 - \sum_{i=1}^C P_i^2$$

P_i adalah probabilitas kelas i di node tersebut, dan C adalah jumlah kelas.

3. Pembuatan Hutan (Random Forest).

Setelah semua pohon keputusan dibuat, Random Forest menggabungkan hasil dari masing-masing pohon. Gunakan voting mayoritas untuk menentukan kelas akhir

$$\hat{Y} = mode(\{T_1(x), T_2(x), T_m(x)\})$$

$T_i(x)$ adalah prediksi dari pohon ke- i .

4. Evaluasi Model
5. Setelah prediksi dihasilkan, model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, guna mengetahui seberapa baik model memprediksi hasil sesuai dengan data aktual.

HASIL DAN PEMBAHASAN

Hasil penelitian yang diperoleh melalui analisis data menggunakan algoritma Random Forest, diikuti dengan pembahasan mendalam mengenai temuan utama. Hasil penelitian dijelaskan secara terperinci untuk menunjukkan pola, hubungan, dan faktor-faktor yang memengaruhi risiko mahasiswa putus kuliah, serta bagaimana model prediksi yang dikembangkan dapat digunakan sebagai alat bantu dalam pengambilan keputusan strategis di lingkungan Universitas Teknologi Akba Makassar.

Variabel yang mempengaruhi tingkat retensi mahasiswa untuk lulus yaitu, 1) akademik, mahasiswa dengan IPK tinggi cenderung termotivasi untuk menyelesaikan studinya, kehadiran konsisten pada perkuliahan meunjukkan keterlibatannya secara aktif, dan beban SKS yang dipenuhi mewakili keberhasilannya akademiknya; 2) Sosial dan Ekonomi, mahasiswa dari keluarga dengan penghasilan minim akan menghadapi tantangan finansial, dukungan moral dan finansial dari keluarga serta hubungan sosial yang baik akan meningkatkan motivasi, dan mahasiswa yang bekerja paruh waktu akan menghadapi konflik waktu; 3) psikologis, Mahasiswa dengan motivasi intrinsik yang tinggi cenderung akan menyelesaikan studi, tingkat stress yang tinggi akan menghambat retensi lulus dan rasa percaya diri dalam menghadapi tantangan akademik akan meningkatkan keberhasilan; 4) Keterlibatan dalam lingkungan kampus, mahasiswa yang aktif dalam organisasi kampus seringkali memiliki keterikatan emosional terhadap kampus, akses ke fasilitas seperti perpustakaan, laboratorium, dan bimbingan akademik akan mendukung retensi, dan interaksi yang baik antara mahasiswa dan dosen akan dapat memberikan dukungan akademik dan emosional. Dan variabel yang terahir adalah lingkungan belajar, metode pengajaran yang interaktif dan relevan memotivasi mahasiswa untuk tetap terlibat dalam pembelajaran, bimbingan yang tepat membantu mengatasi kesulitan akademik, dan Kebijakan terkait beasiswa, fleksibilitas kurikulum, atau cuti akademik memengaruhi retensi mahasiswa.

Untuk dapat diolah menggunakan algoritma Random Forest. Variabel tersebut diubah menjadi numerik agar dapat digunakan dalam proses pelatihan model, semua variabel dapat digunakan sebagai fitur (X) dengan kelulus sebagai target (y).beberapa variable dapat menggunakan skala liker untuk menentukan bobotnya seperti pada variabel Dukungan Keluarga dan Sosial, Motivasi Belajar, Tingkat Stres, Kepercayaan Diri Akademik, Fasilitas Kampus yang Digunakan, dan Hubungan dengan Dosen.

1. IPK, rentang nilai 0.0 hingga 4.0
2. Kehadiran, Persentase kehadiran, 0 hingga 100%.
3. Jumlah SKS yang Diambil, rentang: 0–24.
4. Status Sosial-Ekonomi, diberi nilai 1 untuk rendah, 2 menengah dan 3 tinggi
5. Pekerjaan Paruh Waktu, diberi nilai 0 jika Tidak bekerja dan 1 jika Bekerja
6. Dukungan Keluarga dan Sosial, digunakan skala 1–5
7. Motivasi Belajar, digunakan skala 1–5
8. Tingkat Stres, digunakan skala 1–5
9. Kepercayaan Diri Akademik, digunakan skala 1–5
10. Keterlibatan dalam Organisasi Kampus, 0 jika Tidak aktif dan 1 jika Aktif
11. Fasilitas Kampus yang Digunakan, digunakan skala 1–5
12. Hubungan dengan Dosen, digunakan skala 1–5
13. Kelulusan, 0 jika Tidak Lulus dan 1 jika Lulus

Untuk memastikan dataset bersih, peneliti melakukan langkah-langkah pembersihan dataset, dengan cara menghapus data yang bersifat Missing Values, Duplicate Values, Data Anomali dan Data Invalid

1. Missing Values

Hapus semua baris data yang tidak memiliki nilai atau biasa disebut missing value

2. Duplicate Values

Identifikasi dan hapus data duplikat

3. Data Anomali (Outliers)

Hitung data anomali menggunakan z-score dengan persamaan berikut

$$Z = \frac{(X - \mu)}{\sigma}$$

dimana

Z adalah Z-score.

X adalah nilai data.

μ adalah rata-rata populasi atau sampel.

σ adalah standar deviasi populasi atau sampel.

4. Data Invalid

Memastikan data berada dalam rentang yang wajar (misalnya, IPK antara 0.0 dan 4.0,

Kehadiran antara 0 dan 100%).

Dengan menggunakan dataset yang clear dan valid, hasil evaluasi model dengan metrik Akurasi, Presisi, Recall, dan F1-Score menunjukkan kinerja model yang sangat baik. Berikut adalah kesimpulan berdasarkan metrik yang diberikan

Tabel 1 Laporan Klasifikasi

	precision	recall	f1-score	support
0	0.67	0.83	0.74	149
1	0.75	0.56	0.64	136
accuracy			0.70	285
macro avg	0.71	0.69	0.69	285
weighted avg	0.71	0.70	0.69	285

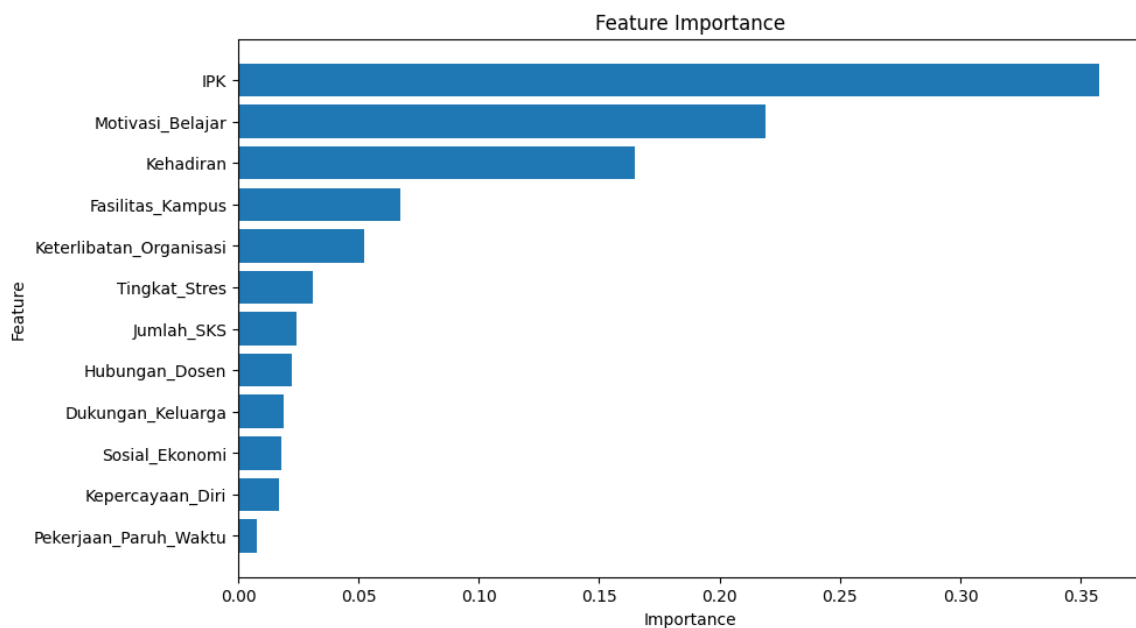
Berdasarkan hasil prediksi, data Tidak Lulus memiliki probabilitas benar sebesar 67%, sedangkan data Lulus memiliki probabilitas benar sebesar 75%. Precision cukup baik, terutama untuk kelas Lulus, namun precision pada kelas Tidak Lulus lebih rendah, menunjukkan adanya false positive atau prediksi salah pada kelas ini. Model berhasil menangkap 83% dari semua data aktual yang Tidak Lulus, tetapi hanya menangkap 56% dari data aktual yang Lulus, mengindikasikan recall rendah untuk kelas Lulus sehingga banyak mahasiswa yang sebenarnya Lulus tidak terdeteksi. F1-score pada kelas Tidak Lulus adalah 74%, menunjukkan keseimbangan yang baik antara precision dan recall, sedangkan pada kelas Lulus hanya mencapai 64%, mengindikasikan kinerja model yang lebih buruk untuk kelas ini. Secara keseluruhan, akurasi model sebesar 70% menunjukkan bahwa model Random Forest ini cukup baik, tetapi bukan yang terbaik, terutama jika terjadi ketidakseimbangan kelas. Nilai macro average dan weighted average stabil, namun macro average masih mengindikasikan adanya ruang untuk perbaikan pada performa model.

Sehingga model Random Forest yang dilatih menunjukkan performa yang cukup baik dalam memprediksi kelulusan mahasiswa. Dengan Akurasi, Presisi, Recall, dan F1-Score yang sangat tinggi, model ini dapat diandalkan untuk melakukan prediksi dengan tingkat kesalahan yang sangat rendah. Namun, tetap perlu berhati-hati untuk memastikan bahwa hasil ini dapat diandalkan pada data baru dan tidak mengalami overfitting. Penggunaan teknik seperti cross-validation dapat membantu memastikan bahwa model ini memiliki generalisasi yang baik.

Berdasarkan analisis variabel penting dari model Random Forest, kita dapat menyimpulkan bahwa beberapa variabel memiliki pengaruh yang lebih besar terhadap kelulusan mahasiswa dibandingkan yang lain. Pada tabel 1, adalah persentase variabel penting yang dapat memengaruhi retensi mahasiswa dalam menyelesaikan studi

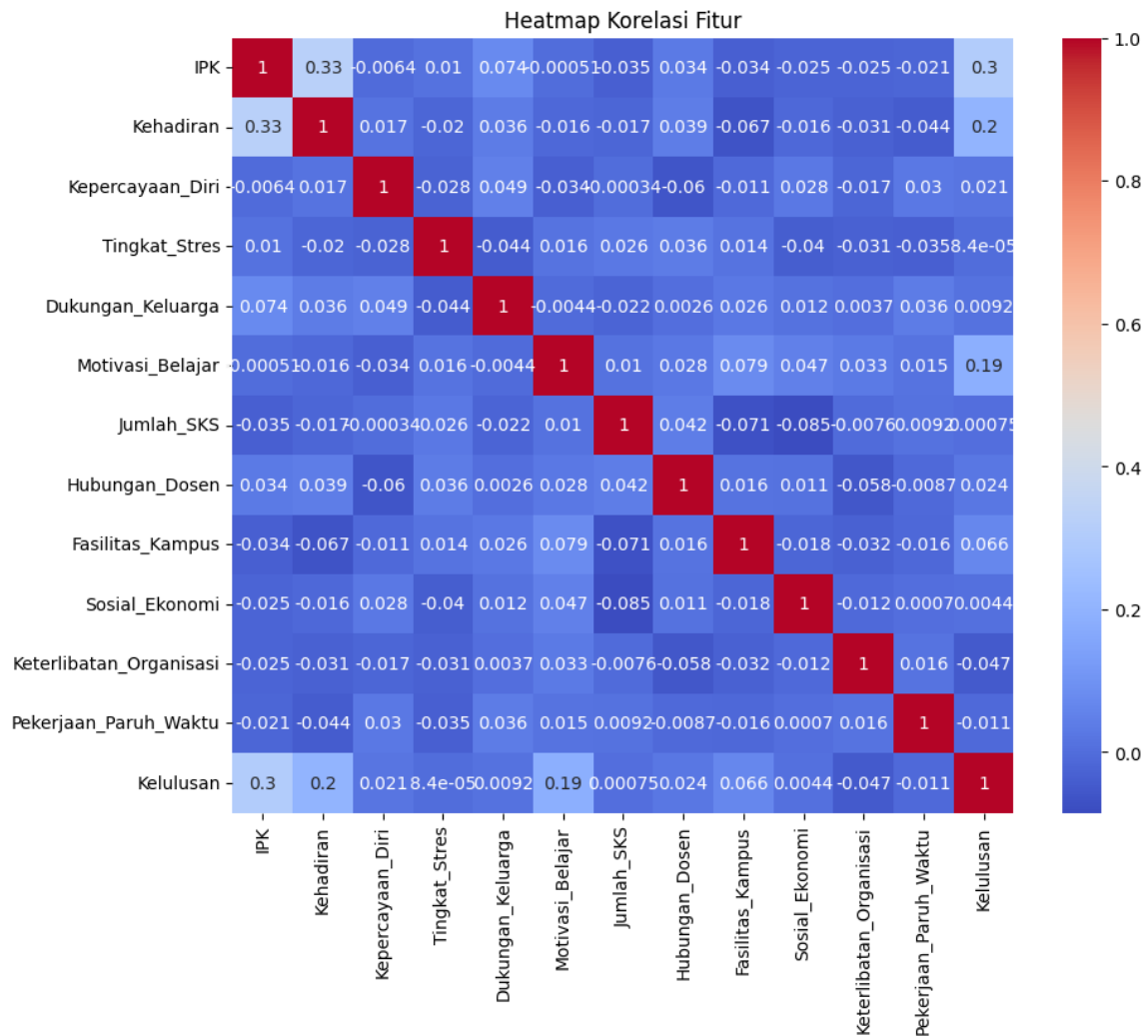
Tabel 2 Persentasi Kepentingan

Feature	Importance
IPK	0.357770
Motivasi_Belajar	0.218841
Kehadiran	0.164814
Fasilitas_Kampus	0.067453
Keterlibatan_Organisasi	0.052285
Tingkat_Stres	0.030997
Jumlah_SKS	0.024320
Hubungan_Dosen	0.022098
Dukungan_Keluarga	0.018844
Sosial_Ekonomi	0.018010
Kepercayaan_Diri	0.016784
Pekerjaan_Paruh_Waktu	0.007785



Gambar 1 Feature Importan

Dalam analisis faktor-faktor yang mempengaruhi kelulusan mahasiswa, ditemukan bahwa IPK memiliki pengaruh terbesar dengan nilai penting sebesar 36%, diikuti oleh Motivasi Belajar dengan nilai penting 22%, dan Kehadiran dengan nilai penting 16%. Fasilitas Kampus juga berperan cukup signifikan dengan nilai penting 7%, disusul oleh Keterlibatan dalam Organisasi dengan nilai penting 5%. Faktor-faktor lainnya seperti Tingkat Stres, Jumlah SKS, Hubungan dengan Dosen, Dukungan Keluarga, Sosial Ekonomi, dan Kepercayaan Diri memiliki pengaruh yang lebih kecil, masing-masing dengan nilai penting di bawah 3%. Pekerjaan Paruh Waktu memiliki pengaruh terkecil dengan nilai penting 0.8%. Temuan ini menunjukkan pentingnya IPK, motivasi belajar, dan kehadiran dalam menentukan kelulusan mahasiswa, sementara faktor-faktor lain juga turut berkontribusi meskipun dalam proporsi yang lebih kecil.



Gambar 2 Headmap Korelasi Fitur

Dalam analisis ini, ditemukan bahwa IPK memiliki korelasi positif dengan Kehadiran (0.33) dan Kelulusan (0.3), yang menunjukkan bahwa kehadiran yang baik berkorelasi dengan IPK yang lebih tinggi dan peluang kelulusan yang lebih besar. Selain itu, Kehadiran juga memiliki korelasi positif dengan IPK (0.33), memperkuat pentingnya kehadiran dalam pencapaian akademis. Motivasi Belajar memiliki korelasi positif dengan Kelulusan (0.19), menekankan peran motivasi internal dalam kelulusan mahasiswa. Di sisi lain, Sosial Ekonomi memiliki korelasi negatif dengan Jumlah SKS (-0.085), menunjukkan bahwa kondisi sosial ekonomi yang lebih rendah mungkin berkaitan dengan pengambilan jumlah SKS yang lebih sedikit.

KESIMPULAN

Penelitian ini bertujuan untuk memprediksi retensi mahasiswa, khususnya dalam mengidentifikasi mahasiswa yang rentan putus kuliah, menggunakan algoritma Random Forest. Algoritma Random Forest terbukti mampu menangani dataset yang kompleks dan heterogen dengan akurasi prediksi yang tinggi. Kemampuannya untuk mengevaluasi feature importance memberikan wawasan tentang variabel-variabel yang paling signifikan dalam menentukan risiko putus kuliah, sehingga memungkinkan pengambilan keputusan yang lebih tepat sasaran. Variabel seperti IPK, kehadiran, jumlah SKS, motivasi belajar, tingkat stres, dan status sosial-ekonomi memiliki pengaruh signifikan terhadap risiko mahasiswa untuk tidak menyelesaikan studi. Variabel-variabel ini tidak hanya mencerminkan kinerja akademik, tetapi juga menunjukkan aspek sosial dan psikologis yang berperan penting dalam keberhasilan studi.

Model yang dikembangkan menghasilkan performa evaluasi yang baik, dengan metrik seperti akurasi, presisi, recall, dan F1-score menunjukkan hasil yang memadai. Hal ini menunjukkan bahwa algoritma Random Forest dapat digunakan sebagai alat prediksi yang andal untuk mendukung kebijakan pencegahan putus kuliah. Informasi yang dihasilkan dari model prediksi dapat dimanfaatkan oleh institusi pendidikan untuk merancang intervensi yang efektif, seperti program bimbingan akademik, konseling psikologis, dan bantuan finansial. Implementasi ini diharapkan dapat meningkatkan retensi mahasiswa dan mengurangi angka putus kuliah.

Sebagai saran penelitian ini memberikan kontribusi praktis dan teoritis dalam upaya meningkatkan retensi mahasiswa melalui pendekatan berbasis data. Ke depan, perlu dilakukan pengembangan lebih lanjut, seperti integrasi data real-time dan validasi model dengan dataset yang lebih besar dan bervariasi, untuk meningkatkan generalisasi hasil. .

REFERENSI

- Aini, Nuru, Muchamad Arif, Irka Tri Agustin, dan Zulfah Binti Toyibah. 2024. "Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika." *Jurnal Informatika* 11(1):11–16. doi: 10.31294/inf.v11i1.20637.
- Ajiwerdhi, Anak Agung Gde Putra, Made Windu Antara Kesiman, dan I. Made Agus Wirawan. 2012. "Pengembangan Sistem Pendukung Keputusan Berbasis Mobile Untuk Pengisian Kartu Rencana Studi Dengan Fuzzy Multi-Attribute Decision Making (FMADM) Metode Simple Additive Weighting (SAW) Di Jurusan Pendidikan Teknik Informatika Universitas Pendidikan Ganesha Singaraja." *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)* 1(2):155. doi: 10.23887/janapati.v1i2.9850.
- Al Farabi, Mohammad, Ayilzi Putri, Abdi Ardiansyah Sihotang, dan Bunga Alfi Syahria. 2024. "Strategi Inovatif Untuk Mengatasi Tantangan dalam Sistem Pendidikan di Amerika Serikat." *Pendas : Jurnal Ilmiah Pendidikan Dasar* 9(3).
- Amalia, Nisa. t.t. "Optimasi Algoritma Random Forest dengan Hyperparameter Tuning Menggunakan GridSearchCV untuk Prediksi Nasabah Churn pada Industri Perbankan."
- Anon. 2024. "Mengatasi Masalah Drop Out: Strategi untuk Meningkatkan Retensi Siswa." *Cikoneng*. Diambil (https://cikoneng-ciamis.desa.id/mengatasi-masalah-drop-out-strategi-untuk-meningkatkan-retensi-siswa?utm_source=chatgpt.com).
- Astuti, Peni. 2024. "Pengaruh Lingkungan Kelas Inklusif, Kebijakan Dukungan Akademik, dan Keaktifan Organisasi Mahasiswa terhadap Keterlibatan Akademik pada Universitas di Jakarta." *Jurnal Pendidikan West Science* 2(02):108–17. doi: 10.58812/jpdws.v2i02.1200.
- Azhari, Devi Syukri, Afnibar, Ulfatmi, dan Gusril Kenedi. t.t. "Konseling di Perguruan Tinggi." *Jurnal Review Pendidikan dan Pengajaran* 6(4):2023.
- Brenab, Leo. 2001. "Random Forests." dalam *Machine Learning*. Vol. 45. Kluwer Academic Publishers.
- Ghozali, Imam. 2013. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 26*. Badan Penerbit Universitas Diponegoro.
- Moesarofah. 2021. "Analisis Karakteristik Retensi Mahasiswa di Perguruan Tinggi."
- Putra, Ade Herdian, Neviyarni, dan Firman. 2023. "Bimbingan dan Konseling di Perguruan Tinggi pada Era Society 5.0." *Jurnal Ilmu Pendidikan dan Sosial* 1(1):128–36. doi: 10.58540/jipsi.v1i1.24.
- Rofi, Muhammad Maulana, Foni Agus Setiawan, dan Freza Riana. 2024. "Perbandingan Metode K-NN dan Random Forest Pada Klasifikasi Mahasiswa Berpotensi Dropout." *INFOTECH journal* 10(1):84–89. doi: 10.31949/infotech.v10i1.8856.

- Rui, Yong, Vicente Ivan Sanchez Carmona, Mohsen Pourvali, Yun Xing, Wei-Wen Yi, Hui-Bin Ruan, dan Yu Zhang. 2022. "Knowledge Mining: A Cross-Disciplinary Survey." *Machine Intelligence Research* 19(2):89–114. doi: 10.1007/s11633-022-1323-6.
- Sugiono. 2013. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suriyanti. 2024. "Analisis Kebutuhan Pelayanan Bimbingan Konseling pada Mahasiswa di Poltekkes Kemenkes RI Prodi Keperawatan Curup." Institut Agama Islam Negeri (IAIN) Curup, Curup.
- Thoriq, Muhammad Fath, Wawan Joko Pranoto, dan Faldi Faldi. 2023. "Penerapan Seleksi Fitur Analysis of Variance Pada Algoritma Random Forest Classifier Dalam Klasifikasi Nilai Mahasiswa." *Explore: Jurnal Sistem Informasi dan Telematika* 14(2):185. doi: 10.36448/jsit.v14i2.3187.