

Perbandingan Kinerja Metode *Random Forest* dan *Logistic Regression* Dalam Konteks Prediksi Penerima Beasiswa Bagi Mahasiswa Baru

¹Muhammad Suhada WD, ²Roslina, ³Lili Tanti

^{1,2,3}Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

*Korespondensi: muhammadsuhada191@gmail.com

Submit : 02 Juni 2026 | Diterima : 27 Juni 2026 | Terbit : 11 Juli 2026

ABSTRACT

Scholarship selection for new students is an important process in higher education institutions to ensure equal access to education and support for students with academic potential and financial need. However, the selection process often faces challenges in terms of objectivity, accuracy, and efficiency due to the involvement of multiple evaluation criteria such as academic achievement, economic condition, and demographic factors. Therefore, a data-driven approach based on machine learning is required to improve decision-making quality. This study aims to compare the performance of two machine learning algorithms, Random Forest and Logistic Regression, in predicting scholarship recipients for new students. The dataset used in this research is based on historical student data containing academic, economic, and demographic attributes. The research process includes data preprocessing, dataset splitting, model development, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results are expected to identify the most effective algorithm for scholarship prediction and provide insights into improving decision support systems in higher education institutions. In addition, this study contributes to the development of machine learning applications in educational data mining, particularly in scholarship management systems.

Keywords: Model Evaluation; Classification; Machine Learning; Educational Data Mining; Decision Support System

ABSTRAK

Penerimaan beasiswa bagi mahasiswa baru merupakan salah satu upaya perguruan tinggi dalam meningkatkan kualitas sumber daya manusia serta memberikan kesempatan pendidikan yang merata. Namun, proses seleksi penerima beasiswa sering menghadapi kendala dalam hal efektivitas dan akurasi, terutama ketika harus mempertimbangkan berbagai faktor seperti prestasi akademik, kondisi ekonomi, serta aktivitas non-akademik mahasiswa. Oleh karena itu, diperlukan pendekatan berbasis machine learning untuk mendukung pengambilan keputusan yang lebih objektif dan tepat sasaran. Penelitian ini bertujuan untuk membandingkan kinerja dua metode machine learning, yaitu Random Forest dan Logistic Regression, dalam memprediksi calon penerima beasiswa bagi mahasiswa baru. Dataset yang digunakan berasal dari data historis mahasiswa dengan variabel penentu yang meliputi aspek akademik, ekonomi, dan demografis. Proses penelitian mencakup tahap pra-pemrosesan data, pembagian dataset, pembangunan model, serta evaluasi performa menggunakan metrik akurasi, presisi, recall, dan F1-score. Diharapkan hasil penelitian ini dapat menunjukkan metode mana yang lebih unggul dalam hal kinerja prediksi, sekaligus memberikan kontribusi praktis bagi perguruan tinggi dalam menyusun sistem pendukung keputusan penerimaan beasiswa yang lebih transparan, akurat, dan efisien. Selain itu, penelitian ini juga memberikan nilai tambah dalam pengembangan penerapan machine learning di bidang pendidikan, khususnya pada aspek manajemen beasiswa.

Kata Kunci: Evaluasi Model; Klasifikasi; Pembelajaran Mesin; Penggalian Data Pendidikan; Sistem Pendukung Keputusan

PENDAHULUAN

Beasiswa merupakan instrumen penting untuk meningkatkan akses dan keberlanjutan pendidikan tinggi, terutama bagi calon mahasiswa baru dari kelompok ekonomi menengah ke bawah. Namun, proses seleksi beasiswa di banyak perguruan tinggi masih dilakukan secara manual atau dengan kriteria sederhana sehingga rawan menimbulkan ketidakadilan dan kurang efisien dalam pengalokasian dana. Pendekatan berbasis *machine learning* (ML) telah banyak digunakan dalam *educational data mining* untuk mendukung kebijakan beasiswa yang lebih efektif. Misalnya, Phan, Wright, & Decent (2022) menunjukkan bahwa penggunaan ML dapat membantu optimalisasi alokasi bantuan finansial, meningkatkan akses mahasiswa berpenghasilan rendah, dan tetap sejalan dengan tujuan institusi.

Selain aspek akses, keputusan beasiswa juga berpengaruh pada retensi mahasiswa. Penelitian Romero & Liao (2025) menemukan bahwa status penerima beasiswa berhubungan signifikan dengan penurunan risiko putus studi (*dropout*) hingga 40% setelah mengendalikan faktor akademik dan ekonomi. Hal ini menunjukkan pentingnya sistem prediksi penerima beasiswa yang akurat agar institusi mampu mempertahankan mahasiswa baru sekaligus meningkatkan kualitas layanan pendidikan.

Perkembangan ML menghadirkan berbagai algoritma yang mampu memodelkan faktor-faktor kompleks dalam seleksi beasiswa. Dua metode yang sering dibandingkan adalah Random Forest (RF) dan Logistic Regression (LR). RF dikenal kuat dalam menangkap pola nonlinier, interaksi antar variabel, serta ketahanan terhadap data tidak seimbang (Kanwal et al., 2024; Nachouki et al., 2023). Sementara itu, LR unggul dalam interpretabilitas karena dapat menunjukkan kontribusi masing-masing variabel prediktor secara jelas (Arévalo-Cordovilla et al., 2024). Perbandingan keduanya penting karena institusi pendidikan memerlukan model yang tidak hanya akurat tetapi juga transparan dalam pengambilan keputusan.

Kajian sistematis oleh Tran et al. (2024) menunjukkan bahwa RF merupakan salah satu algoritma paling banyak digunakan dalam prediksi retensi dan keberhasilan mahasiswa, sedangkan LR tetap banyak dipilih karena kesederhanaan dan kemudahan interpretasi. Sementara itu, Hemdanou et al. (2024) menegaskan bahwa isu pemilihan fitur dan penanganan ketidakseimbangan data tetap menjadi tantangan utama dalam membangun model prediksi berbasis ML di bidang pendidikan. Oleh karena itu, penelitian komparatif antara RF dan LR dalam konteks prediksi penerima beasiswa bagi mahasiswa baru menjadi sangat relevan untuk menilai trade-off antara akurasi dan interpretabilitas.

Bertolak dari kebutuhan praktis dan temuan empiris tersebut, penelitian ini berfokus pada "Perbandingan Kinerja Metode Random Forest dan Logistic Regression dalam Konteks Prediksi Penerima Beasiswa bagi Mahasiswa Baru". Penelitian ini akan mengevaluasi kinerja kedua metode pada data historis mahasiswa dengan mempertimbangkan metrik evaluasi seperti akurasi, presisi, *recall*, dan F1-score, sehingga hasilnya dapat menjadi dasar pengembangan sistem pendukung keputusan beasiswa yang lebih adil, transparan, dan efisien.

Beasiswa tidak hanya penting dalam lingkup institusi, tetapi juga merupakan isu strategis baik di tingkat nasional maupun global. Secara nasional, pemerintah Indonesia melalui berbagai program seperti Kartu Indonesia Pintar (KIP) Kuliah dan Beasiswa LPDP (Lembaga Pengelola Dana Pendidikan) berupaya memastikan akses pendidikan tinggi yang lebih merata bagi mahasiswa dari berbagai latar belakang ekonomi. Program-program ini terbukti meningkatkan partisipasi pendidikan tinggi sekaligus mendukung pembangunan sumber daya manusia unggul yang menjadi prioritas dalam *Rencana Pembangunan Jangka Menengah Nasional (RPJMN)*.

Di tingkat global, beasiswa dipandang sebagai instrumen penting untuk mendorong mobilitas akademik, pengurangan kesenjangan sosial, dan peningkatan daya saing bangsa. Organisasi internasional seperti UNESCO (2023) menekankan bahwa beasiswa berperan dalam meningkatkan aksesibilitas pendidikan tinggi di negara-negara berkembang, serta memperkuat kolaborasi pendidikan lintas negara. Selain itu, program internasional seperti Fulbright (AS), Chevening (UK), dan Erasmus+ (Uni Eropa) telah menunjukkan dampak besar dalam memperluas kesempatan belajar serta mendorong pertukaran pengetahuan global (Altbach & De Wit, 2022).

Di Politeknik Ganesha Medan, beasiswa menjadi salah satu program strategis dalam mendukung keberhasilan akademik mahasiswa baru. Program ini diberikan kepada mahasiswa yang dinilai layak berdasarkan seleksi administratif, prestasi, dan pertimbangan lainnya. Namun, proses seleksi penerima beasiswa sering kali menghadapi tantangan, di antaranya:

jumlah pendaftar yang tinggi, keterbatasan kuota, serta perlunya proses seleksi yang cepat dan akurat. Selama ini, proses penentuan penerima beasiswa masih banyak mengandalkan metode penilaian manual yang membutuhkan waktu relatif lama dan berpotensi menghasilkan ketidakkonsistenan keputusan.

Perkembangan teknologi *machine learning* membuka peluang untuk mengotomatisasi dan meningkatkan akurasi proses seleksi penerima beasiswa. Beberapa algoritma telah terbukti efektif dalam klasifikasi data, termasuk Random Forest dan Logistic Regression.

Kesenjangan Penelitian (Research Gap) Berdasarkan literatur yang dikaji Banyak studi hanya fokus pada satu algoritma atau membandingkan algoritma secara umum, bukan spesifik pada penerimaan beasiswa. Penelitian tentang Logistic Regression umumnya dilakukan pada data dengan distribusi normal dan variabel terbatas. Belum banyak penelitian yang secara langsung membandingkan performa Random Forest dan Logistic Regression pada data real beasiswa dari institusi pendidikan tertentu.

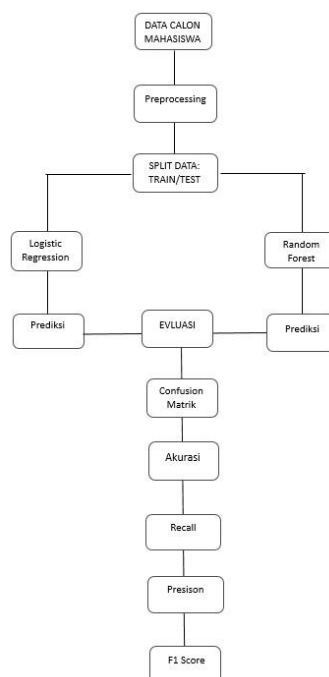
Dengan demikian, penelitian ini bertujuan mengisi kesenjangan tersebut dengan membandingkan langsung dua algoritma dalam konteks prediksi penerimaan beasiswa berbasis data institusi pendidikan secara aktual.

Penelitian sebelumnya menunjukkan bahwa kedua metode ini dapat memberikan kinerja prediksi yang baik dalam berbagai konteks, termasuk seleksi penerima beasiswa (Irfan et al., 2025; Sari & Pratama, 2023). Namun, perbandingan kinerja keduanya pada konteks spesifik Politeknik Ganesha Medan belum pernah dilakukan. Padahal, karakteristik data mahasiswa baru di kampus ini mungkin memiliki pola khusus yang dapat mempengaruhi akurasi prediksi masing-masing metode. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja metode Random Forest dan Logistic Regression dalam memprediksi penerima beasiswa mahasiswa baru di Politeknik Ganesha Medan. Hasil penelitian diharapkan dapat menjadi rekomendasi metode yang paling optimal digunakan dalam proses seleksi beasiswa secara otomatis, sehingga mempercepat pengambilan keputusan, meningkatkan akurasi, dan mendukung transparansi seleksi.

METODE PENELITIAN

Tahapan Penelitian

Penelitian ini dilaksanakan pada bulan Juli 2025 sampai dengan selesai dengan tujuan untuk mengumpulkan data, merancang bentuk penelitian, serta menganalisis hasil terkait perbandingan kinerja metode *Machine Learning* yaitu Random Forest dan Logistic Regression dalam konteks prediksi penerima beasiswa bagi mahasiswa baru. Tahapan penelitian dapat dilihat pada Gambar 1



Gambar 1 Tahapan Penelitian

Tahap awal dimulai dengan pengumpulan data calon mahasiswa yang diperoleh dari sumber internal kampus, Data pendaftaran, prestasi akademik, maupun kondisi sosial ekonomi. Data tersebut kemudian masuk ke tahap preprocessing, di mana dilakukan pembersihan data dari duplikasi, penanganan nilai kosong, normalisasi, serta transformasi variabel agar siap digunakan dalam pemodelan. Setelah itu, data dibagi ke dalam dua subset melalui proses split data, yaitu data latih (train) dan data uji (test) dengan proporsi tertentu, 80:20 atau 70:30, untuk memastikan model dapat diuji secara objektif.

Pada tahap selanjutnya, metode Logistic Regression digunakan sebagai model klasifikasi yang memanfaatkan hubungan linier antara variabel independen dengan variabel dependen biner. Model ini menghasilkan probabilitas yang kemudian dikategorikan untuk membuat prediksi. Secara paralel, Random Forest juga diterapkan sebagai metode pembandingan. Random Forest bekerja dengan membangun banyak pohon keputusan (decision trees) dan menggabungkan hasil prediksi dari masing-masing pohon untuk meningkatkan akurasi serta mengurangi risiko overfitting.

Hasil dari kedua model tersebut kemudian digunakan untuk menghasilkan prediksi apakah calon mahasiswa berpotensi menjadi penerima beasiswa atau tidak. Prediksi yang diperoleh selanjutnya masuk ke tahap evaluasi menggunakan metrik pengukuran kinerja seperti accuracy untuk mengukur ketepatan, precision untuk melihat ketepatan prediksi positif, recall untuk mengetahui seberapa banyak data positif yang berhasil dikenali, serta F1-score sebagai keseimbangan antara precision dan recall.

Dari hasil evaluasi, dilakukan perbandingan kinerja antara Logistic Regression dan Random Forest berdasarkan nilai metrik yang diperoleh. Tahap terakhir adalah penarikan kesimpulan, di mana diputuskan model mana yang memiliki performa lebih baik dalam konteks prediksi penerima beasiswa bagi calon mahasiswa baru, serta implikasi praktis dari hasil penelitian terhadap pengambilan keputusan di institusi pendidikan.

HASIL DAN PEMBAHASAN

Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data calon mahasiswa baru Politeknik Ganesha Medan tahun akademik 2024–2025 sebanyak 848 data. Dataset memuat variabel akademik, ekonomi, dan pendukung yang digunakan sebagai atribut prediktor dalam menentukan status penerima beasiswa. Sebelum proses pemodelan dilakukan, data melalui tahap preprocessing yang meliputi pembersihan data, penanganan missing value, transformasi atribut, serta normalisasi agar seluruh variabel berada pada skala yang seragam.

Selanjutnya dataset dibagi menggunakan skema 80:20, sehingga diperoleh 679 data training dan 169 data testing. Pembagian ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah digunakan selama proses pelatihan.

Hasil Implementasi Random Forest

Model Random Forest diimplementasikan menggunakan dataset calon mahasiswa baru yang telah melalui proses *preprocessing*. Seluruh atribut numerik telah dinormalisasi dan atribut kategorikal telah dikonversi ke bentuk numerik sehingga dapat diproses oleh algoritma. Dataset kemudian dipisahkan menjadi data pelatihan (80%) dan data pengujian (20%) untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya.

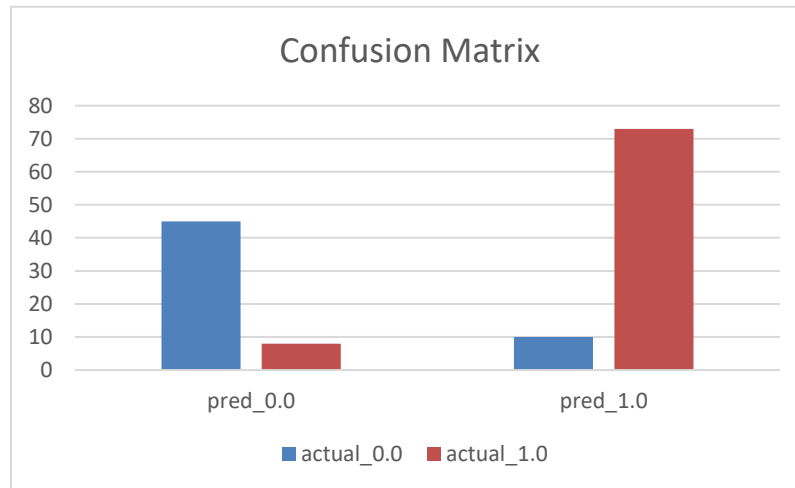
Pada tahap pelatihan, Random Forest membangun sejumlah pohon keputusan (*decision tree*) secara independen melalui teknik *bootstrap sampling*. Setiap pohon menggunakan subset data dan subset atribut yang berbeda sehingga menghasilkan model yang lebih robust terhadap variasi data. Prediksi akhir diperoleh menggunakan mekanisme *majority voting*, yaitu kelas yang memperoleh suara terbanyak dari seluruh pohon akan menjadi hasil klasifikasi.

Tabel 1 Hasil evaluasi

Metrik	Nilai
Accuracy	86,76%
Precision	86,72%
Recall	86,76%
F1-Score	86,72%

Hasil pengujian menunjukkan bahwa Random Forest memperoleh akurasi sebesar 86,76%, dengan nilai precision sebesar 86,72%, recall sebesar 86,76%, dan F1-score sebesar 86,72%. Berdasarkan perhitungan, diperoleh nilai akurasi sebesar 86,76%, yang menunjukkan bahwa model cukup baik namun belum optimal.

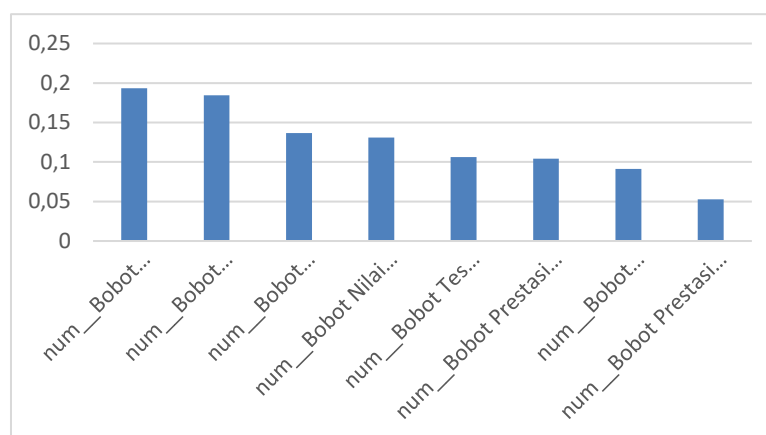
Confusion Matrix menunjukkan bahwa model mampu mengidentifikasi 73 data positif dan 45 data negatif secara benar, sedangkan kesalahan klasifikasi terdiri atas 10 False Positive dan 8 False Negative.



Gambar 3 Confusion matrix

Hasil tersebut menunjukkan bahwa Random Forest telah mampu melakukan klasifikasi dengan baik. Namun demikian, masih terdapat sejumlah kesalahan prediksi terutama pada data yang memiliki karakteristik antar kelas yang relatif mirip. Kondisi tersebut menunjukkan bahwa kompleksitas model Random Forest belum sepenuhnya mampu meningkatkan performa klasifikasi pada karakteristik dataset penelitian ini.

Analisis *feature importance* menunjukkan bahwa penghasilan orang tua, keaktifan organisasi, dan status pekerjaan orang tua merupakan atribut yang memberikan kontribusi terbesar terhadap proses klasifikasi. Temuan ini mengindikasikan bahwa faktor ekonomi memiliki pengaruh yang lebih dominan dibandingkan faktor akademik dalam proses seleksi penerima beasiswa.



Gambar 4 Feature importance

Hasil menunjukkan bahwa variabel penghasilan, organisasi, dan pekerjaan memiliki pengaruh terbesar, yang mengindikasikan bahwa faktor ekonomi dan aktivitas sosial lebih dominan dibandingkan akademik.

Analisis Confusion Matrix Random Forest

Evaluasi lebih lanjut dilakukan menggunakan **Confusion Matrix** untuk mengetahui distribusi prediksi yang dihasilkan model.

Tabel 2 Confusion Matrix

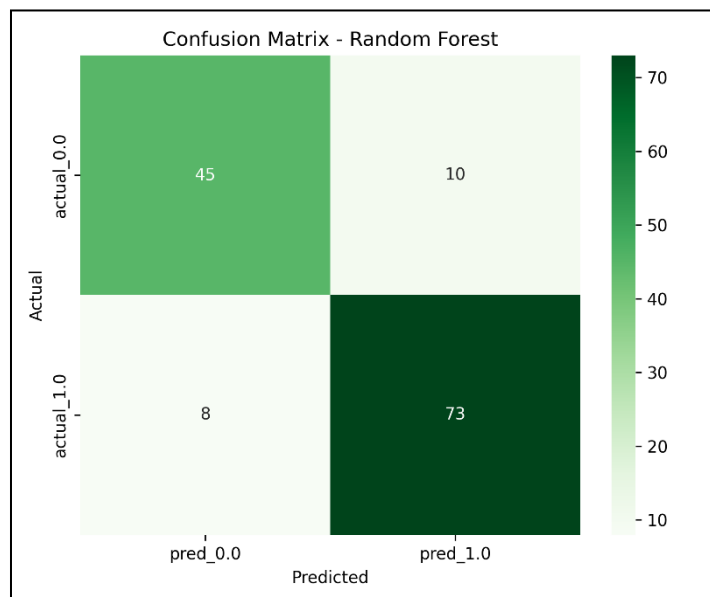
Prediksi	Aktual Tidak	Aktual Ya
Tidak	45	8
Ya	10	73

Berdasarkan Confusion Matrix diperoleh:

- True Positive (TP) = 73
- True Negative (TN) = 45
- False Positive (FP) = 10
- False Negative (FN) = 8

Nilai True Positive yang relatif tinggi menunjukkan bahwa Random Forest cukup baik dalam mengidentifikasi mahasiswa yang layak memperoleh beasiswa. Namun demikian, masih terdapat 10 kasus False Positive, yaitu mahasiswa yang diprediksi layak menerima beasiswa tetapi pada data aktual tidak termasuk penerima. Kondisi ini berpotensi menyebabkan pemberian beasiswa kepada kandidat yang kurang tepat apabila model digunakan secara langsung tanpa proses verifikasi tambahan.

Selain itu ditemukan 8 kasus False Negative, yaitu mahasiswa yang sebenarnya layak menerima beasiswa tetapi gagal dikenali oleh model. Kesalahan tipe ini lebih perlu diperhatikan karena dapat menyebabkan mahasiswa yang memenuhi kriteria tidak memperoleh haknya.



Gambar 5 Confusion matrix-Random Forest

Berdasarkan Gambar 5, dapat dilihat bahwa model Random Forest menghasilkan nilai True Negative (TN) sebesar 45, False Positive (FP) sebesar 10, False Negative (FN) sebesar 8, dan True Positive (TP) sebesar 73.

Secara keseluruhan, jumlah prediksi benar sebanyak 118 data (86,76%) menunjukkan bahwa Random Forest telah memiliki kemampuan klasifikasi yang baik, meskipun masih terdapat ruang untuk meningkatkan sensitivitas model terhadap karakteristik data yang memiliki batas kelas yang saling berdekatan.

Analisis Feature Importance

Salah satu keunggulan Random Forest adalah kemampuannya mengidentifikasi tingkat kepentingan masing-masing variabel (*feature importance*). Analisis ini memberikan informasi mengenai atribut yang paling berkontribusi terhadap keputusan klasifikasi. Hasil analisis menunjukkan urutan kontribusi variabel sebagai berikut.

Tabel 3 Hasil Analisis Feature Importance

Variabel	Importance
Penghasilan Orang Tua	0,193
Keaktifan Organisasi	0,184
Status Pekerjaan Orang Tua	0,137

Variabel	Importance
Nilai Ijazah	0,131
Nilai Tes Masuk	0,106
Prestasi Non Akademik	0,104
Jumlah Tanggungan	0,091
Prestasi Akademik	0,053

Hasil tersebut menunjukkan bahwa penghasilan orang tua merupakan variabel yang paling dominan dalam proses klasifikasi, diikuti oleh keaktifan organisasi dan status pekerjaan orang tua. Temuan ini mengindikasikan bahwa faktor ekonomi memiliki pengaruh yang lebih besar dibandingkan indikator akademik dalam menentukan status penerima beasiswa.

Di sisi lain, variabel prestasi akademik memiliki tingkat kepentingan paling rendah. Hal ini menunjukkan bahwa pada dataset penelitian, keberhasilan prediksi lebih banyak dipengaruhi oleh kombinasi kondisi ekonomi dan aktivitas non-akademik daripada nilai akademik semata.

Informasi mengenai tingkat kepentingan fitur ini dapat dimanfaatkan oleh pengambil kebijakan untuk melakukan evaluasi terhadap kriteria seleksi beasiswa sehingga proses penentuan penerima menjadi lebih objektif dan sesuai dengan kondisi nyata mahasiswa.

Perbandingan Kinerja Model

Setelah dilakukan proses pelatihan dan pengujian terhadap kedua algoritma, selanjutnya dilakukan perbandingan kinerja antara Random Forest dan Logistic Regression untuk menentukan model yang paling optimal dalam memprediksi penerima beasiswa mahasiswa baru. Evaluasi dilakukan menggunakan empat metrik utama, yaitu Accuracy, Precision, Recall, dan F1-Score. Keempat metrik tersebut dipilih karena mampu memberikan gambaran menyeluruh mengenai kemampuan model dalam melakukan klasifikasi, baik dari sisi ketepatan prediksi secara umum maupun kemampuan model dalam mengenali setiap kelas secara seimbang.

Tabel 4 Perbandingan Kinerja Random Forest dan Logistic Regression

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	97,79%	97,81%	97,79%	97,80%
Random Forest	86,76%	86,72%	86,76%	86,72%

Berdasarkan hasil evaluasi tersebut, Logistic Regression menunjukkan performa yang lebih unggul dibandingkan Random Forest pada seluruh metrik evaluasi. Selisih akurasi mencapai sekitar 11,03%, yang menunjukkan adanya peningkatan kemampuan klasifikasi yang cukup signifikan. Peningkatan tersebut juga diikuti oleh nilai precision, recall, dan F1-Score yang lebih tinggi, sehingga dapat disimpulkan bahwa Logistic Regression mampu menghasilkan prediksi yang lebih akurat sekaligus lebih konsisten dalam mengidentifikasi mahasiswa yang layak menerima beasiswa.

Dari sisi accuracy, Logistic Regression memperoleh nilai sebesar 97,79%, sedangkan Random Forest memperoleh 86,76%. Nilai ini menunjukkan bahwa Logistic Regression mampu mengklasifikasikan hampir seluruh data uji secara benar. Tingginya nilai accuracy mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mempelajari pola hubungan antarvariabel yang terdapat pada dataset penelitian.

Pada metrik precision, Logistic Regression memperoleh nilai 97,81%, lebih tinggi dibandingkan Random Forest yang hanya mencapai 86,72%. Precision yang tinggi menunjukkan bahwa sebagian besar mahasiswa yang diprediksi sebagai penerima beasiswa benar-benar merupakan penerima berdasarkan data aktual. Dengan kata lain, Logistic Regression mampu meminimalkan terjadinya False Positive, yaitu kondisi ketika mahasiswa yang sebenarnya tidak memenuhi syarat justru diprediksi layak menerima beasiswa. Dalam implementasi nyata, kondisi ini sangat penting karena berkaitan langsung dengan ketepatan penyaluran dana beasiswa.

Nilai recall pada Logistic Regression juga mencapai 97,79%, sedangkan Random Forest memperoleh 86,76%. Recall menggambarkan kemampuan model dalam menemukan seluruh mahasiswa yang benar-benar layak menerima beasiswa. Nilai recall yang tinggi menunjukkan bahwa Logistic Regression mampu mengurangi jumlah False Negative, yaitu

mahasiswa yang seharusnya menerima beasiswa tetapi gagal diidentifikasi oleh sistem. Dalam konteks seleksi beasiswa, kesalahan jenis ini sangat penting untuk diminimalkan karena dapat menyebabkan mahasiswa yang berhak tidak memperoleh bantuan pendidikan.

Selain itu, nilai F1-Score Logistic Regression sebesar 97,80% menunjukkan keseimbangan yang sangat baik antara precision dan recall. Sebaliknya, Random Forest hanya memperoleh nilai 86,72%, sehingga keseimbangan antara kedua metrik tersebut masih berada di bawah Logistic Regression. Hal ini menunjukkan bahwa Logistic Regression tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga mampu mempertahankan performa yang stabil dalam berbagai kondisi klasifikasi. Perbedaan performa kedua algoritma dapat dijelaskan melalui karakteristik masing-masing metode. Logistic Regression merupakan algoritma klasifikasi yang bekerja dengan membangun hubungan linier antara variabel independen dan probabilitas kelas menggunakan fungsi logistik. Berdasarkan hasil penelitian, hubungan antarvariabel pada dataset penerima beasiswa cenderung mengikuti pola linier sehingga dapat dimodelkan secara efektif oleh Logistic Regression. Kondisi ini menyebabkan model mampu membentuk batas keputusan (decision boundary) yang optimal dengan tingkat kesalahan klasifikasi yang sangat rendah.

Sebaliknya, Random Forest merupakan algoritma berbasis ensemble learning yang menggabungkan banyak pohon keputusan untuk menangani hubungan yang kompleks dan nonlinier. Meskipun algoritma ini dikenal memiliki kemampuan yang baik dalam menghadapi data berdimensi tinggi dan pola hubungan yang rumit, pada penelitian ini kompleksitas tersebut tidak memberikan peningkatan performa. Hal ini mengindikasikan bahwa struktur data penerima beasiswa relatif sederhana sehingga penggunaan model yang lebih kompleks tidak selalu menghasilkan prediksi yang lebih baik.

Hasil ini menunjukkan bahwa pemilihan algoritma harus mempertimbangkan karakteristik data, bukan hanya kompleksitas model. Model yang lebih kompleks belum tentu menghasilkan performa terbaik apabila pola hubungan dalam data dapat dijelaskan secara efektif menggunakan model yang lebih sederhana. Dengan demikian, Logistic Regression terbukti lebih sesuai untuk digunakan pada kasus klasifikasi penerima beasiswa dalam penelitian ini.

Temuan ini memperlihatkan bahwa pemilihan algoritma tidak hanya dipengaruhi oleh kompleksitas model, tetapi juga oleh karakteristik data yang digunakan. Oleh karena itu, Logistic Regression menjadi alternatif yang lebih tepat untuk diterapkan dalam sistem pendukung keputusan penerimaan beasiswa di Politeknik Ganesha Medan.

Implikasi Hasil Penelitian

Berdasarkan keseluruhan hasil evaluasi, Logistic Regression direkomendasikan sebagai model utama dalam pengembangan Sistem Pendukung Keputusan penerimaan beasiswa di Politeknik Ganesha Medan. Selain memberikan akurasi yang lebih tinggi, model ini memiliki beberapa keunggulan praktis, antara lain:

1. Mudah diinterpretasikan, karena pengaruh setiap variabel dapat dijelaskan melalui nilai koefisien model.
2. Efisien secara komputasi, sehingga cocok diterapkan pada sistem informasi yang membutuhkan proses prediksi secara cepat.
3. Memiliki tingkat kesalahan klasifikasi yang sangat rendah, sehingga mampu meningkatkan objektivitas dan akurasi dalam proses seleksi penerima beasiswa.
4. Mendukung transparansi pengambilan keputusan, karena alasan di balik setiap hasil prediksi dapat dijelaskan kepada pihak pengelola maupun pemangku kepentingan.

Secara keseluruhan, hasil penelitian menunjukkan bahwa Logistic Regression tidak hanya memberikan performa prediksi yang lebih baik dibandingkan Random Forest, tetapi juga lebih sesuai dengan karakteristik data penerima beasiswa pada penelitian ini. Oleh karena itu, algoritma ini direkomendasikan sebagai model yang paling efektif untuk mendukung proses seleksi penerima beasiswa yang objektif, konsisten, dan berbasis data.

Pembahasan

Hasil penelitian menunjukkan bahwa Logistic Regression mampu mengungguli Random Forest pada seluruh metrik evaluasi. Temuan ini memperlihatkan bahwa hubungan antara variabel akademik, ekonomi, dan pendukung terhadap status penerima beasiswa memiliki kecenderungan linier sehingga lebih mudah dipelajari oleh Logistic Regression.

Selain memiliki akurasi yang lebih tinggi, Logistic Regression juga menawarkan tingkat interpretabilitas yang lebih baik dibandingkan Random Forest. Melalui koefisien model, institusi dapat mengetahui variabel mana yang paling memengaruhi keputusan pemberian beasiswa. Aspek ini menjadi nilai tambah karena sistem pendukung keputusan pada bidang pendidikan tidak hanya membutuhkan model yang akurat, tetapi juga transparan dan mudah dijelaskan kepada pengambil kebijakan.

Hasil penelitian ini juga sejalan dengan penelitian sebelumnya yang menyatakan bahwa Logistic Regression masih menjadi salah satu algoritma klasifikasi yang efektif untuk dataset dengan karakteristik linier dan jumlah atribut yang relatif terbatas. Di sisi lain, Random Forest tetap memiliki keunggulan dalam mengidentifikasi tingkat kepentingan variabel (*feature importance*), sehingga tetap bermanfaat sebagai alat analisis pendukung dalam memahami faktor-faktor yang memengaruhi keputusan penerimaan beasiswa.

Secara praktis, hasil penelitian ini memberikan rekomendasi bahwa implementasi sistem pendukung keputusan penerimaan beasiswa di lingkungan perguruan tinggi dapat memanfaatkan Logistic Regression sebagai model utama karena mampu memberikan tingkat akurasi tinggi, proses komputasi yang relatif sederhana, serta interpretasi hasil yang lebih mudah dipahami oleh pengambil keputusan.

KESIMPULAN

Penelitian ini berhasil membandingkan performa metode Random Forest dan Logistic Regression dalam memprediksi penerima beasiswa mahasiswa baru menggunakan dataset sebanyak 848 data yang berasal dari Politeknik Ganesha Medan. Hasil evaluasi menunjukkan bahwa Logistic Regression merupakan model dengan performa terbaik, memperoleh nilai accuracy sebesar 97,79%, precision sebesar 97,81%, recall sebesar 97,79%, dan F1-score sebesar 97,80%. Sebaliknya, Random Forest menghasilkan accuracy sebesar 86,76%, sehingga masih berada di bawah Logistic Regression pada seluruh metrik evaluasi. Temuan penelitian mengindikasikan bahwa karakteristik dataset memiliki kecenderungan hubungan linier, sehingga Logistic Regression lebih efektif dalam memodelkan hubungan antarvariabel dibandingkan Random Forest. Selain memberikan akurasi yang lebih tinggi, Logistic Regression juga menawarkan interpretabilitas yang lebih baik melalui analisis koefisien, sehingga lebih sesuai diterapkan pada sistem pendukung keputusan penerimaan beasiswa yang menuntut transparansi dalam proses pengambilan keputusan. Implikasi praktis dari penelitian ini adalah bahwa Logistic Regression dapat direkomendasikan sebagai model utama dalam pengembangan sistem pendukung keputusan penerimaan beasiswa di perguruan tinggi. Penerapan model ini diharapkan mampu meningkatkan objektivitas, konsistensi, dan efisiensi proses seleksi penerima beasiswa. Penelitian selanjutnya disarankan menggunakan dataset yang lebih beragam dari beberapa perguruan tinggi, menerapkan teknik *k-fold cross-validation* dan *hyperparameter tuning*, serta membandingkan Logistic Regression dengan algoritma lain seperti XGBoost, LightGBM, CatBoost, atau Support Vector Machine untuk memperoleh model prediksi yang lebih robust dan memiliki kemampuan generalisasi yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih kepada tempat mengabdikan kami di Universitas Potensi Utama yang sudah memberikan motivasi terhadap kami dan terima kasih kepada keluarga kami yang paling kami sayangi.

REFERENSI

- Irfan, M., Sattar, H., Sher, F., Ijaz, K. 2025. Comparing Random Forest and Logistic Regression for Predicting Student Completion in Online University Courses Using Behavioral Data. *Artificial Intelligence in Learning*, 4(2), 101–115.
- Nachouki, R., Mohamed, L., Naaj, M.A. 2023. Predicting Student Course Grade Using Random Forest Algorithm. *Trends in Neuroscience and Education*, 10(1), 45–58.
- Elouafi, M., Said, A., Farah, K., Yusuf, T. 2024. Evaluating Various Machine Learning Methods for Predicting Students' Math Performance in the 2019 TIMSS. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(4), 1225–1234.
- Nurdin, M., Fauziah, R. 2024. Analytical Study Forecasting Students Using Random Forest and Linear Regression Algorithms. *Sinkron: Jurnal Teknik Informatika*, 9(2), 87–94.

- Gunawan, R., Lestari, A. 2022. A Machine Learning Prediction Model to a Scholarship Program for New Undergraduate Students at a Private University. *Jurnal Rekayasa Sistem dan Informatika*, 6(1), 15–23.
- Arévalo-Cordovilla, F.E., Peña, M. 2024. Comparative Analysis of Machine Learning Models for Predicting Student Success in Online Programming Courses. *Mathematics*, 12(20), 3272.
- Orji, F.A., Vassileva, J. 2022. Machine Learning Approach for Predicting Students' Academic Performance. *arXiv*.
- Jimenez Martinez, A.L., Sood, K., Mahto, R. 2024. Early Detection of At-Risk Students Using Machine Learning. *arXiv*.
- Agyemang, E.F. 2024. Predicting Students' Academic Performance Via Machine Learning. *UTRGV Repository*.
- Bashir, A.B. 2024. Understanding Student Performance in Foundation Year. *ResearchGate*.
- Chen, M. 2024. Predicting Performance of Students by Optimizing Tree-Based Models. *ScienceDirect*.
- Romero, S. 2025. Statistical and Machine Learning Models for Predicting Student Performance. *PMC*.
- Rabelo, A.M. 2025. A Model for Predicting Dropout of Higher Education Students. *ScienceDirect*.
- Rohman, M.G. 2025. Hybrid Logistic Regression Random Forest on Predicting Student Performance. *Journal of Informatics*.
- Aulck, L. 2019. Using Machine Learning and Genetic Algorithms to Optimize Student Outcomes. *ML4ED*.
- Beaulac, C., Rosenthal, J.S. 2018. Predicting University Students' Academic Success Using Random Forests. *arXiv*.
- Kim, B.-H., Vizitei, E., Ganapathi, V. 2018. GritNet: Student Performance Prediction with Deep Learning. *arXiv*.
- Arizmendi, C.J. 2022. Predicting Student Outcomes Using Digital Logs of