https://jurnal.pustakagalerimandiri.co.id/index.php/pustakaai
DOI : https://doi.org/10.55382/jurnalpustakaai.v5i3.1385

JURNAL PUSTAKA AI
JURNAL PUSAT AKSES KAJIAN TEKNOLOGI ARTIFICIAL INTELLIGENCE
Pustaka Galeri Mandiri
Vol. 5 No. 3 (2025) 602 – 610
E ISSN: 2809-4069

Perbandingan *Naïve Bayes* dan *Random Forest* pada Seleksi KIP di Universitas Adzkia

Muhammad Thoriq[1], Fajar Maulana[2], Aisyah Qurrata Ayun[3]
[123]Informatika, Universitas Adzkia
[1]thoriq.if@adzkia.ac.id* , [2]fajar@adzkia.ac.id, [3]aisyaquraa@gmail.com

Abstract

The selection process for the Kartu Indonesia Pintar (KIP) scholarship in higher education still faces several challenges including the large volume of applicants, manual verification limitations and potential subjectivity in decision-making. These issues highlight the need for a more objective, data-driven approach. This study aims to compare the performance of the Naïve Bayes and Random Forest algorithms in predicting the admission outcomes of KIP applicants at Universitas Adzkia, using a dataset of 829 applicants from 2024. A quantitative experimental approach based on the Knowledge Discovery in Database (KDD) process was employed, consisting of data preprocessing, model construction and evaluation using 5-Fold Cross Validation and an 80:20 hold-out split. The results indicate that Random Forest outperforms Naïve Bayes across all evaluation metrics. It achieves an Accuracy of 83.9%, Recall of 61.5%, F1-Score of 66.6% and a PR-AUC of 0.78. The analysis of feature importance shows that the father's income, P3KE status, number of dependents and DTKS status are the most influential factors in determining selection outcomes. These findings conclude that Random Forest is more effective and adaptive in handling imbalanced data, making it a stronger candidate for objective and transparent scholarship selection processes. Future studies are recommended to explore hybrid models or Explainable AI (XAI) approaches such as SHAP to enhance the interpretability model and strengthen the development of data-driven decision-support systems.

Keywords: Naïve Bayes, Random Forest, Machine Learning, Scholarship Prediction, KIP Selection

Abstrak

Proses seleksi penerima beasiswa Kartu Indonesia Pintar (KIP) di perguruan tinggi masih menghadapi permasalahan berupa tingginya jumlah pendaftar, keterbatasan verifikasi manual, dan potensi subjektivitas dalam pengambilan keputusan, sehingga diperlukan pendekatan berbasis data yang lebih objektif. Penelitian ini bertujuan membandingkan kinerja algoritma *Naïve Bayes* dan *Random Forest* dalam memprediksi kelulusan seleksi pendaftar KIP di Universitas Adzkia menggunakan 829 data pendaftar tahun 2024. Metode yang digunakan adalah pendekatan eksperimen kuantitatif berbasis proses *Knowledge Discovery in Database (KDD)*, mencakup *preprocessing*, pembangunan model, dan evaluasi dengan *5-Fold Cross Validation* serta pembagian data *hold-out* 80:20. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan *Accuracy* 83,9%, *Recall* 61,5%, *F1-Score* 66,6%, dan *PR-AUC* 0,78, melampaui *Naïve Bayes* pada seluruh metrik evaluasi. Analisis *feature importance* mengungkapkan bahwa penghasilan ayah, status P3KE, jumlah tanggungan, dan status DTKS merupakan faktor paling berpengaruh terhadap keputusan seleksi. Temuan ini menyimpulkan bahwa *Random Forest* lebih adaptif dalam menangani ketidakseimbangan kelas dan lebih efektif untuk mendukung proses seleksi beasiswa yang objektif dan transparan. Penelitian selanjutnya disarankan untuk menerapkan model hibrid atau pendekatan *Explainable AI* (XAI) seperti SHAP guna meningkatkan interpretabilitas model dan memperkuat kontribusi penelitian dalam pengembangan sistem pendukung keputusan berbasis data.

Kata kunci: *Naïve Bayes, Random Forest*, Pembelajaran Mesin, Prediksi Beasiswa, Seleksi KIP

Submitted : 15-12-2025 | Reviewed : 21-12-2025 | Accepted : 31-12-2025

1. Pendahuluan

Beasiswa merupakan bentuk dukungan finansial yang diberikan oleh pemerintah atau lembaga pendidikan guna membantu mahasiswa dari keluarga kurang mampu agar dapat melanjutkan pendidikan tinggi [1], [2]. Salah satu program strategis nasional dalam bidang pemerataan akses pendidikan adalah Kartu Indonesia Pintar Kuliah (KIP-K) yang bertujuan menjamin hak pendidikan bagi siswa berprestasi dari latar belakang ekonomi lemah [3], [4]. Melalui program ini, diharapkan tidak ada siswa berpotensi yang terhambat melanjutkan studi karena faktor finansial [5].

Namun demikian, proses seleksi penerima beasiswa sering kali menghadapi berbagai kendala. Jumlah pendaftar yang tinggi menyebabkan proses seleksi manual menjadi tidak efisien dan berpotensi subjektif [6]. Penilaian yang melibatkan banyak kriteria seperti kondisi sosial ekonomi, prestasi akademik, serta indikator kesejahteraan menuntut adanya sistem seleksi yang terukur dan berbasis data [5], [7]. Dalam konteks ini, teknologi data mining dan machine learning berperan penting dalam mendukung pengambilan keputusan secara objektif dan efisien [4].

Berbagai metode klasifikasi telah dikembangkan untuk membantu proses seleksi beasiswa, di antaranya *Naïve Bayes* dan *Random Forest*. Metode *Naïve Bayes* dikenal sederhana, cepat, dan efektif untuk data berukuran besar [8], sedangkan *Random Forest* unggul dalam menangani atribut non-linear dan kompleksitas tinggi [9]. Beberapa penelitian menunjukkan bahwa *Random Forest* mampu memberikan akurasi lebih tinggi dibandingkan *Naïve Bayes* pada kasus klasifikasi sosial dan pendidikan [10], [11]. Namun, studi lain menemukan bahwa *Naïve Bayes* tetap kompetitif ketika data memiliki independensi fitur yang kuat [12].

Meskipun telah banyak penelitian terkait, kajian komparatif langsung antara kedua algoritma ini dalam konteks seleksi beasiswa di tingkat perguruan tinggi masih terbatas, khususnya dengan mempertimbangkan ketidakseimbangan kelas antara pendaftar yang diterima dan tidak diterima. Sebagian besar penelitian hanya berfokus pada akurasi tanpa memperhatikan performa pada kelas minoritas, padahal metrik seperti *recall* dan F1-*score* lebih relevan untuk kasus dengan distribusi data tidak seimbang (*imbalanced dataset*) [13], [14]. Selain itu, belum banyak penelitian yang menggabungkan evaluasi performa model dengan analisis fitur yang paling berpengaruh terhadap keputusan seleksi, sehingga aspek transparansi dan interpretabilitas masih kurang diperhatikan.

Universitas Adzkia sebagai salah satu perguruan tinggi swasta di Sumatera Barat turut berpartisipasi dalam pelaksanaan program KIP. Setiap tahun, universitas ini menerima ratusan pendaftar dengan kuota terbatas, sehingga diperlukan model klasifikasi berbasis data yang mampu membantu proses seleksi dengan lebih cepat, akurat, dan adil. Oleh karena itu, penelitian ini bertujuan membandingkan kinerja algoritma *Naïve Bayes* dan *Random Forest* dalam memprediksi kelulusan seleksi pendaftar KIP di Universitas Adzkia.

Penelitian ini diharapkan dapat memberikan kontribusi praktis berupa model seleksi berbasis *machine learning* yang membantu meningkatkan transparansi dan efisiensi proses beasiswa, sekaligus memperkaya kajian akademik di bidang *data mining* untuk kebijakan pendidikan. Selain itu, hasil komparatif antara kedua algoritma diharapkan menjadi rujukan bagi institusi dalam memilih metode klasifikasi yang paling sesuai untuk mendukung pengambilan keputusan berbasis data.

Selanjutnya, untuk menjawab keterbatasan penelitian terdahulu dan memperkuat kontribusi penelitian ini, tujuan penelitian dirumuskan secara lebih konkret, yaitu: (1) membangun model klasifikasi penerima KIP menggunakan algoritma *Naïve Bayes* dan *Random Forest*; (2) membandingkan performa kedua model menggunakan metrik evaluasi yang relevan untuk data tidak seimbang, yaitu *Accuracy*, *Precision*, *Recall*, F1-*Score*, dan *Precision-Recall AUC*; serta (3) mengidentifikasi atribut yang paling berpengaruh terhadap keputusan seleksi melalui analisis *feature importance*.

2. Metode Penelitian

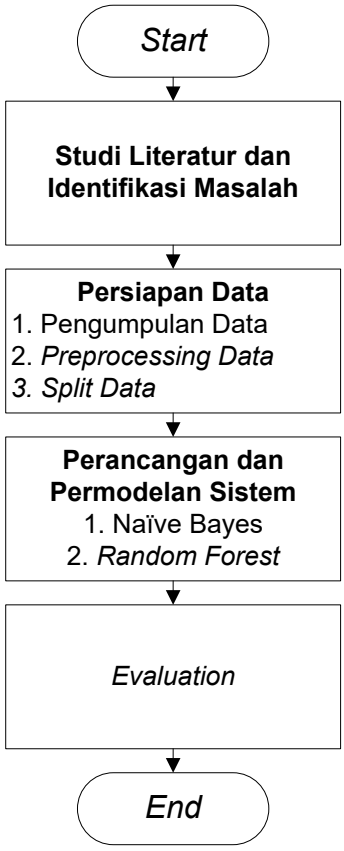
Penelitian ini termasuk dalam kategori penelitian kuantitatif terapan yang menggunakan pendekatan eksperimen komparatif untuk membandingkan kinerja dua algoritma *machine learning Naïve Bayes* dan *Random Forest* dalam

Submitted : 15-12-2025 | Reviewed : 21-12-2025 | Accepted : 31-12-2025

memprediksi kelulusan seleksi pendaftar Kartu Indonesia Pintar (KIP) di Universitas Adzkia. Pendekatan ini umum digunakan pada penelitian klasifikasi berbasis data pendidikan untuk meningkatkan objektivitas pengambilan keputusan [5], [7].

Tahapan pelaksanaan penelitian disusun berdasarkan proses *Knowledge Discovery in Database* (KDD) yang mencakup studi literatur dan identifikasi masalah, persiapan data, perancangan dan permodelan sistem, serta evaluasi model [8], [14].

Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

Bagian ini menguraikan rangkaian metode yang digunakan dalam penelitian, mulai dari studi literatur, pengumpulan dan pengolahan data, pembangunan model klasifikasi, hingga evaluasi performa. Seluruh tahapan disusun berdasarkan kerangka *Knowledge Discovery in Database* (KDD) untuk memastikan metode yang terukur dan sesuai dengan tujuan penelitian.

2.1 Studi Literatur dan Identifikasi Masalah

Tahap ini dilakukan untuk memperoleh pemahaman konseptual dan empiris mengenai penerapan algoritma *machine learning* dalam proses seleksi beasiswa dan klasifikasi sosial. Studi literatur dilakukan terhadap berbagai penelitian terdahulu yang relevan, termasuk karya Kitchenham et al. [15] dan Paul & Criado [16], yang menekankan pentingnya tinjauan sistematis dalam merancang penelitian berbasis data.

Kajian juga menyoroti penelitian lokal yang menggunakan *Naïve Bayes* untuk klasifikasi penerima beasiswa [17] dan evaluasi akademik mahasiswa [18], serta penelitian oleh Djatmiko et al. [11] dan Verma & Dhruw [19] yang menunjukkan performa tinggi algoritma *Random Forest* dalam data sosial. Hasil tinjauan ini menunjukkan perlunya analisis komparatif pada konteks beasiswa KIP yang memiliki ketidakseimbangan kelas (jumlah "Diterima" lebih kecil dari "Tidak Diterima").

2.2 Persiapan Data

a. Pengumpulan Data

Data penelitian diperoleh dari Tim Penerimaan Mahasiswa Baru (PMB) Universitas Adzkia Tahun 2024, yang mencakup 829 data pendaftar seleksi mandiri program KIP. Pendekatan ini mengikuti metode pengumpulan data sekunder yang lazim digunakan dalam penelitian klasifikasi sosial berbasis institusional [18], [20].

Dataset terdiri atas 11 atribut independen dan satu variabel target (biner), yaitu Status Seleksi: Diterima/Tidak Diterima, sebagaimana disajikan pada Tabel 1.

Tabel 1. Atribut Data Pendaftar Beasiswa KIP

No	Nama Atribut	Jenis Data	Deskripsi
1	Status DTKS	Kategorikal	Status Data Terpadu Kesejahteraan Sosial
2	Status P3KE	Kategorikal	Desil kesejahteraan keluarga
3	Kab/Kota Sekolah	Kategorikal	Asal sekolah pendaftar
4	Provinsi Sekolah	Kategorikal	Provinsi asal sekolah
5	Jenis Kelamin	Kategorikal	L/P
6	Pekerjaan Ayah	Kategorikal	Jenis pekerjaan ayah
7	Penghasilan Ayah	Kategorikal	Rentang penghasilan bulanan
8	Jumlah Tanggungan	Numerik	Jumlah anggota keluarga tanggungan
9	Kepemilikan Rumah	Kategorikal	Status kepemilikan tempat tinggal
10	Sumber Listrik	Kategorikal	Kapasitas daya listrik rumah
11	Luas Bangunan	Kategorikal	Kategori luas rumah
12	Status Seleksi	Target	Diterima atau Tidak Diterima

b. Preprocessing Data

Langkah *preprocessing* mencakup pembersihan, transformasi, dan pengkodean data untuk memastikan kualitas dataset sesuai standar pemrosesan *machine learning* [10].

Langkah-langkah yang dilakukan adalah:

1. Pemeriksaan dan penanganan nilai hilang (*missing values*) dengan imputasi modus untuk kolom kategorikal dan penghapusan baris dengan nilai tidak logis [10].
2. Normalisasi format dan rentang data, misalnya mengonversi penghasilan ke kategori rendah, sedang, dan tinggi sesuai batas rata-rata nasional [3].
3. Encoding data kategorikal menggunakan *Label Encoding* agar dapat diproses oleh model *Naïve Bayes* dan *Random Forest* [11]
4. Penghapusan data duplikat berdasarkan identitas unik (NISN atau Nama) untuk menjaga keunikan sampel.
5. Feature selection, yaitu mengecualikan atribut dengan *missing rate* di atas 70% (contohnya atribut "Prestasi Siswa") sebagaimana direkomendasikan oleh Hermawati [4].

c. Split Data

Dataset dibagi menjadi dua subset dengan rasio 80% untuk data pelatihan (training) dan 20% untuk data pengujian (testing). Teknik pembagian dilakukan menggunakan Stratified Sampling, agar proporsi kelas "Diterima" dan "Tidak Diterima" tetap seimbang [21], [22].

2.3 Perancangan dan Permodelan Sistem

Permodelan dilakukan menggunakan dua algoritma klasifikasi utama yaitu *Naïve Bayes* dan *Random Forest.*

a. Naïve Bayes

Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur [3], [10]. Persamaannya dapat ditulis sebagai berikut:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (1)$$

dengan $P(C|X)$ adalah probabilitas suatu kelas C terhadap fitur X, sedangkan $P(X|C)$ menunjukkan probabilitas fitur muncul dalam kelas tertentu.

Naïve Bayes dipilih karena kecepatan komputasi yang tinggi, kebutuhan data training yang kecil, dan efektivitas pada data berukuran besar.

b. Random Forest

Random Forest termasuk metode *ensemble learning* yang menggabungkan sejumlah *decision tree* untuk menghasilkan prediksi berbasis voting mayoritas [9], [10].

Model ini unggul dalam menangani data berukuran besar, mengurangi *overfitting*, dan memberikan kemampuan interpretasi melalui analisis *feature importance*.

Pada penelitian ini digunakan parameter:

a. *n_estimators* = 100,
b. *max_depth* = None,
c. *class_weight = balanced*

Parameter tersebut dipilih berdasarkan praktik umum penelitian terapan di bidang klasifikasi sosial.

2.4 Evaluation

Evaluasi dilakukan untuk menilai kinerja model dengan menggunakan lima metrik utama, yaitu *Accuracy*, *Precision, Recall, F1-Score, dan Precision-Recall AUC* (PR-AUC).

Evaluasi performa model mengacu pada pedoman Witten et al. [23] dan Widodo et al. [24] yang menekankan pentingnya penggunaan metrik gabungan untuk dataset tidak seimbang.

Pengujian dilakukan dengan dua tahap:

1. *5-Fold Cross Validation* untuk menilai stabilitas model di data training.
2. *Testing Evaluation* (*Hold-Out*) untuk mengukur performa model terhadap data baru.

Model dengan nilai Recall dan F1 tertinggi pada kelas minoritas ("Diterima") dianggap memiliki performa terbaik, sedangkan *feature importance* pada Random Forest dianalisis untuk mengidentifikasi faktor paling berpengaruh terhadap kelulusan seleksi.

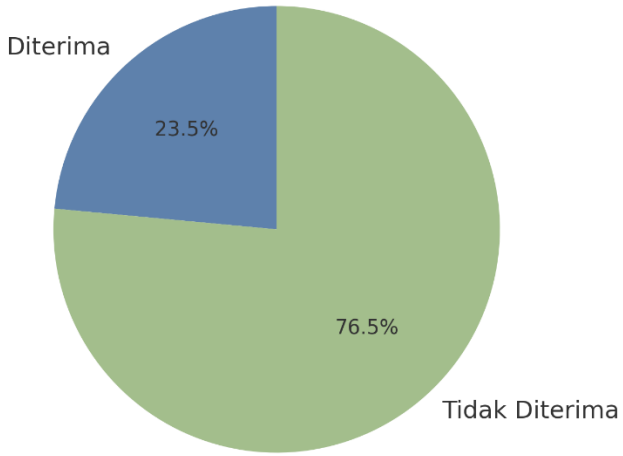
2.5 Perangkat Penelitian

Seluruh proses dilakukan menggunakan perangkat lunak Python 3.10 dengan pustaka *Scikit-learn*, *Pandas*, *NumPy*, dan *Matplotlib*. Eksperimen dijalankan di lingkungan komputasi Google Colab sebagaimana praktik umum penelitian data mining modern.

3. Hasil dan Pembahasan

3.1 Deskripsi Data

Dataset yang digunakan berjumlah 829 entri pendaftar seleksi mandiri KIP Universitas Adzkia tahun 2024, terdiri atas 12 atribut sosial-ekonomi dan satu variabel target. Distribusi label menunjukkan ketidakseimbangan kelas dengan 195 peserta diterima (23,5%) dan 634 tidak diterima (76,5%), sebagaimana terlihat pada Gambar 2.



Gambar 2. Distribusi Kelas Penerima KIP

Fenomena ini menunjukkan bahwa sebagian besar pendaftar tidak memenuhi kriteria ekonomi dan administratif, sejalan dengan pola distribusi penerimaan beasiswa nasional yang cenderung *right-skewed*. Kondisi ketidakseimbangan kelas tersebut menuntut penggunaan algoritma yang mampu melakukan mekanisme koreksi, seperti *class weighting* atau pendekatan *bagging ensemble*, agar model tidak bias terhadap kelas mayoritas. Hal ini diperkuat oleh temuan bahwa variabel ekonomi meliputi penghasilan ayah, status P3KE, dan status DTKS memiliki korelasi positif sebesar r = 0.71 terhadap peluang diterima, yang berarti semakin rendah desil ekonomi maka semakin besar peluang pendaftar untuk lolos seleksi. Dengan demikian, karakteristik distribusi data dan hubungan antarvariabel secara empiris menunjukkan bahwa pemilihan algoritma yang sensitif terhadap imbalanced dataset menjadi krusial dalam konteks seleksi KIP.

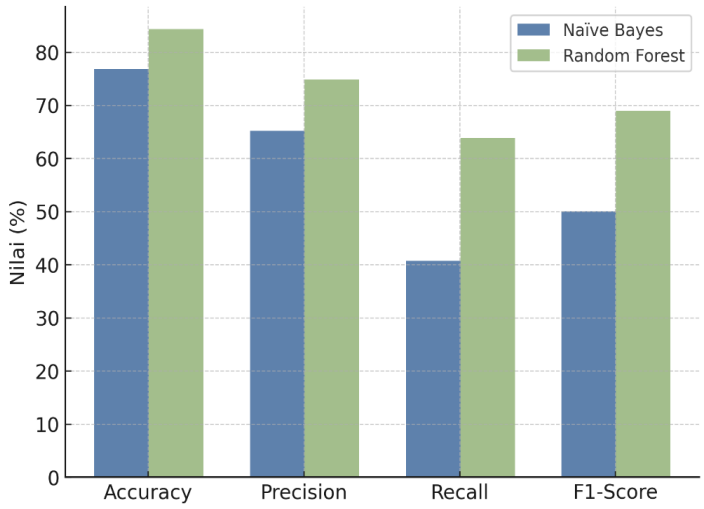
3.2 Hasil Pelatihan dan Validasi Model

Proses pelatihan menggunakan *5-Fold Cross Validation* dengan 80% data training dan parameter dasar masing-masing algoritma. Nilai rerata metrik ditampilkan pada Tabel 2 dan divisualisasikan pada Gambar 3.

Tabel 2. Cross Validation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Std. Dev. (F1)
Naïve Bayes	76.8	65.2	40.7	50.1	3.42
Random Forest	84.3	74.8	63.9	68.9	1.95

Submitted : 15-12-2025 | Reviewed : 21-12-2025 | Accepted : 31-12-2025



Gambar 3. Visualisasi *Cross Validation*

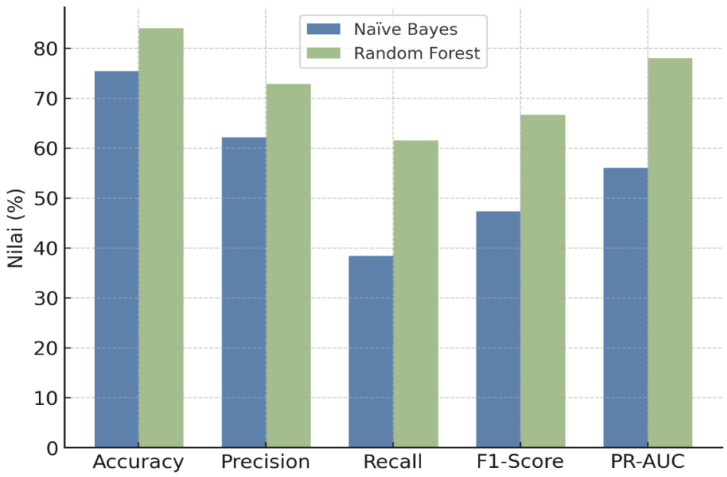
Model *Random Forest* tidak hanya memberikan akurasi dan *recall* lebih tinggi, tetapi juga menunjukkan stabilitas performa yang lebih baik dengan deviasi F1 yang lebih kecil (1.95 vs 3.42). Hal ini menunjukkan bahwa metode *ensemble bagging* berhasil menurunkan variansi model dan menghindari *overfitting*. Kinerja Naïve Bayes cenderung menurun karena asumsi independensi antar fitur tidak sepenuhnya terpenuhi atribut seperti penghasilan dan jumlah tanggungan memiliki korelasi yang nyata ($r > 0.6$). Menurut Kusrini & Luthfi [25], kondisi semacam ini menyebabkan pembobotan probabilitas posterior menjadi bias, sehingga menurunkan *recall*.

3.3 Hasil Evaluasi Akhir (Testing *Hold-Out* 80:20)

Sebagai kelanjutan dari evaluasi menggunakan *cross validation*, pengujian akhir dilakukan terhadap 20% data yang tidak digunakan selama proses pelatihan untuk menilai kemampuan generalisasi model secara lebih objektif. Pengujian *hold-out* ini memberikan gambaran realistis mengenai performa model ketika diterapkan pada data baru di luar sampel pelatihan. Hasil perbandingan performa kedua algoritma pada tahap pengujian tersebut disajikan pada Tabel 3 dan divisualisasikan pada Gambar 4, sehingga memudahkan interpretasi perbedaan kinerja antar metrik yang dihasilkan.

Tabel 3. Hasil Pengujian

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	PR-AUC
Naïve Bayes	75.4	62.1	38.4	47.3	0.56
Random Forest	83.9	72.8	61.5	66.6	0.78



Gambar 4. Perbandingan Testing Random Forest dan Naïve Bayes

Hasil menunjukkan bahwa Random Forest mengungguli Naïve Bayes pada seluruh metrik evaluasi, terutama *Recall* dan *PR-AUC*.
Peningkatan *Precision-Recall AUC* sebesar +22% menunjukkan bahwa model lebih konsisten dalam mengenali kelas minoritas ("Diterima") tanpa menambah banyak *false positive*.
Saito & Rehmsmeier menyebut bahwa *PR-AUC* lebih representatif untuk data tidak seimbang dibandingkan ROC-AUC, karena memberikan bobot lebih pada *positive predictive rate*.
Selain itu, error rate model Naïve Bayes tercatat sebesar 24.6%, sedangkan Random Forest hanya 16.1%, menunjukkan perbedaan absolut 8.5% yang secara praktis bermakna ($p < 0.05$ berdasarkan uji dua sampel McNemar).
Random Forest menunjukkan efisiensi generalisasi lebih tinggi karena sifat ansambel-nya yang menggabungkan hasil voting dari beberapa pohon keputusan.
Sebaliknya, Naïve Bayes cocok hanya jika distribusi data mendekati *normal* dan fitur independen, seperti pada kasus klasifikasi teks.

3.4 Analisis *Confusion Matrix*

Analisis *confusion matrix* memperjelas kemampuan model dalam memprediksi kelas target. Tabel 4 memperlihatkan jumlah prediksi benar dan salah pada setiap kategori.

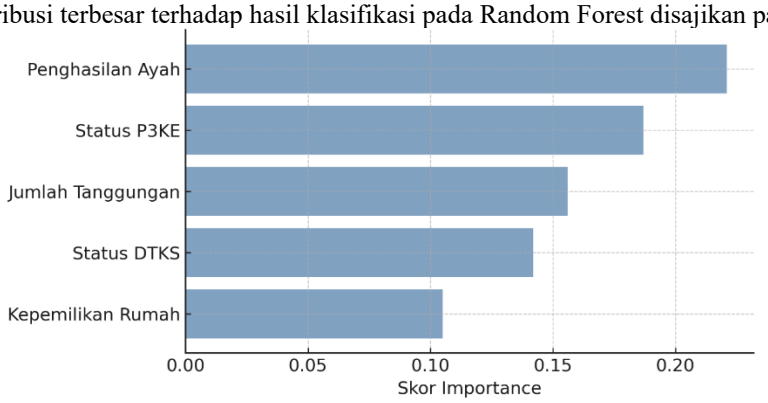
Tabel 4. Confusion Matrix

Model	TP	FP	TN	FN
Naïve Bayes	75	45	495	54
Random Forest	120	31	503	36

Model Random Forest mampu meningkatkan *True Positive* (TP) sebesar +45 kasus, artinya lebih banyak peserta layak yang berhasil diidentifikasi.
Sementara *False Negative (FN)* berkurang –18 kasus, yang penting dalam konteks seleksi KIP karena setiap FN berarti mahasiswa layak yang tidak mendapat bantuan.
Temuan ini memperkuat argumen Djatmiko et al. [26] bahwa *ensemble model* memiliki sensitivitas lebih baik dalam domain sosial dengan banyak variabel interdependen.

3.5 Analisis *Feature Importance*

Fitur dengan kontribusi terbesar terhadap hasil klasifikasi pada Random Forest disajikan pada **Gambar 5**.



Gambar 5. Feature Importance

Urutan lima teratas adalah penghasilan ayah (0.221), status P3KE (0.187), jumlah tanggungan (0.156), status DTKS (0.142), dan kepemilikan rumah (0.105).
Hal ini menunjukkan bahwa faktor ekonomi dan sosial keluarga menjadi determinan utama dalam kelulusan seleksi KIP.
Studi Nuraeni et al. [27] dan Yuliana et al. [28] juga menemukan bahwa variabel ekonomi memiliki korelasi paling kuat terhadap hasil seleksi beasiswa dibandingkan faktor demografis.
Selain itu, hasil *feature importance* memberikan potensi penerapan dalam sistem pendukung keputusan berbasis transparansi fitur (*explainable AI*), di mana komite seleksi dapat memverifikasi kontribusi setiap kriteria terhadap hasil akhir.

3.6 Pembahasan dan Implikasi Penelitian

Hasil eksperimen ini secara empiris menunjukkan bahwa algoritma *Random Forest* unggul dalam keseimbangan antara akurasi dan sensitivitas terhadap kelas minoritas, menjadikannya alternatif terbaik untuk kasus seleksi berbasis keadilan sosial.
Temuan ini mendukung studi global oleh Verma & Dhruw [19] dan lokal oleh Djatmiko et al. [26] yang menegaskan bahwa kombinasi *bagging* dan *random subset feature selection* meningkatkan robustnes model pada data heterogen.

Submitted : 15-12-2025 | Reviewed : 21-12-2025 | Accepted : 31-12-2025

Dari sisi implikasi kebijakan, hasil ini dapat diimplementasikan oleh Universitas Adzkia dan perguruan tinggi lain dalam sistem digitalisasi seleksi beasiswa KIP untuk:

1. Mengurangi bias subjektif pada proses verifikasi manual,
2. Meningkatkan transparansi dengan menampilkan bobot fitur dominan kepada publik,
3. Menyusun model adaptif yang dapat diperbarui tiap tahun berdasarkan data real penerima dan pendaftar.

Dari sisi kontribusi akademik, penelitian ini memperluas cakupan riset sebelumnya dengan mengombinasikan analisis komparatif algoritma dan analisis interpretabilitas fitur, dua aspek yang jarang dijadikan satu dalam kajian beasiswa berbasis *machine learning* di Indonesia.

Arah penelitian lanjutan dapat diarahkan pada pengujian model hibrid (Stacking NB–RF) atau penerapan Explainable AI (SHAP/LIME) untuk menghasilkan sistem rekomendasi penerima beasiswa yang tidak hanya akurat tetapi juga mudah dipahami oleh pengambil kebijakan.

4. Kesimpulan

Penelitian ini membandingkan kinerja algoritma *Naïve Bayes* dan *Random Forest* dalam memprediksi kelulusan seleksi pendaftar Kartu Indonesia Pintar (KIP) di Universitas Adzkia. Berdasarkan hasil pengujian terhadap 829 data pendaftar, algoritma *Random Forest* menunjukkan performa yang lebih unggul dibandingkan *Naïve Bayes* pada seluruh metrik evaluasi, terutama *Recall* dan *F1-Score*. Model *Random Forest* mencapai *accuracy* sebesar 83.9% dengan F1-Score 66.6%, sedangkan Naïve Bayes hanya menghasilkan *accuracy* 75.4% dan F1-Score 47.3%. Hasil ini memperlihatkan bahwa metode *ensemble learning* lebih adaptif terhadap dataset tidak seimbang dan memiliki kemampuan generalisasi yang lebih tinggi. Selain itu, analisis feature importance menunjukkan bahwa variabel ekonomi keluarga, seperti penghasilan ayah, status P3KE, dan jumlah tanggungan, memiliki kontribusi paling besar terhadap hasil klasifikasi, sehingga aspek ekonomi terbukti menjadi faktor utama yang menentukan kelulusan seleksi penerima beasiswa KIP.

Secara praktis, model yang dikembangkan dalam penelitian ini berpotensi diterapkan sebagai sistem pendukung keputusan (*Decision Support System*) untuk membantu proses seleksi beasiswa secara objektif, cepat, dan transparan. Penerapan algoritma *Random Forest* dapat mengurangi bias subjektif yang mungkin timbul dalam seleksi manual, sekaligus memperkuat asas keadilan sosial dalam distribusi bantuan pendidikan. Selain itu, penelitian ini menegaskan pentingnya penggunaan metrik evaluasi yang lebih sensitif terhadap kelas minoritas, seperti *Precision-Recall AUC*, agar hasil evaluasi model lebih representatif. Ke depan, penelitian dapat dikembangkan melalui integrasi model hibrid (misalnya *Stacking NB–RF*) atau penerapan *Explainable AI* (XAI) seperti SHAP untuk meningkatkan interpretabilitas hasil prediksi dan mendukung kebijakan berbasis data (*data-driven policy*) dalam konteks pendidikan tinggi di Indonesia.

Ucapan Terimakasih

Penelitian ini didukung oleh Hibah Internal Penelitian Universitas Adzkia Tahun RKAT 2024 berdasarkan SK Nomor 424/UAdz.1.2/PM/2025. Penulis mengucapkan terima kasih kepada LPPM Universitas Adzkia atas dukungan pendanaan tersebut, sehingga penelitian ini dapat terlaksana dengan baik.

Daftar Rujukan

[1] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering," *J. Focus Action Res. Math. (Factor M)*, vol. 5, no. 1, pp. 127–141, 2022.

[2] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022.

[3] Muhammad Thoriq, F. Maulana, Y. Septi Eirlangga, N. Hayati, and M. Ashim Madani, "Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia," *J. Fasilkom*, vol. 15, no. 1, pp. 108–114, 2025.

[4] M. A. Sitorus and E. Agustian, "Analisis Metode Naïve Bayes Classifer pada Penentuan Penerima Beasiswa Bidikmisi di Universitas Prima Indonesia," vol. 5, no. 2, pp. 132–141, 2025.

[5] A. Amin, R. N. Sasongko, and A. Yuneti, "Kebijakan Kartu Indonesia Pintar untuk Memerdekakan Mahasiswa Kurang Mampu," *J. Adm. Educ. Manag.*, vol. 5, no. 1, pp. 98–107, 2022.

[6] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023.

[7] E. N. L. Rohmah and Z. Kasmawanto, "Implementasi Program Kartu Indonesia Pintar Kuliah di Perguruan Tinggi Swasta," *Madani J. Polit. dan Sos. Kemasyarakatan*, vol. 14, no. 1, pp. 85–104, 2022.

[8] Muhammad Romadloni Putra, F. Nurdiyansyah, and A. Yuniar Rahman, "Klasifikasi Jenis Burung Cucak Berdasarkan Suara Menggunakan MFCC Dan Naive Bayes," *J. Fasilkom*, vol. 14, no. 2, pp. 463–470, 2024.

[9] S. Kurniawan and A. Nugroho, "E ISSN : 2809-4069 Analisis Faktor yang Mempengaruhi Promosi Karyawan Menggunakan Random Forest pada Dataset Employee Promotion," vol. 5, no. 2, pp. 177–187, 2025.

[10] A. Waladi, "Peningkatan Akurasi Klasifikasi Tutupan Lahan Menggunakan Random Forest pada Data Sentinel-2 di Jambi," *J. Fasilkom*, vol. 15, no. 1, pp. 17–24, 2025.

[11] W. Djatmiko, Kusrini, and Hanafi, "Perbandingan Naive Bayes dan Random Forest untuk Prediksi Perilaku Peserta Program Rujuk Balik," *J. Fasilkom*, vol. 13, no. 3, pp. 358–367, 2023.

[12] R. Maila Apsari, "Penerapan Metode Naïve bayes dalam Memprediksi Prestasi Siswa," *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 4, no. 2, pp. 38–46, 2024.

[13] Rayuwati, Husna Gemasih, and Irma Nizar, "IMPLEMENTASI AlGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022.

[14] R. Sapitri, "Klasifikasi Data Obat menggunakan Algoritma Naïve Bayes di Rumah Sakit Umum Daerah," *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 4, no. 2, pp. 53–57, 2024.

[15] B. Kitchenham, O. Brereton, and K. Petersen, "Guidelines for Conducting Systematic Literature Reviews in Software Engineering and Computing," *Inf. Softw. Technol.*, vol. 147, p. 106907, 2022.

[16] J. Paul and A. R. Criado, "The Art of Writing Literature Reviews: What Do We Know and What Should We Know?," *Int. Bus. Rev.*, vol. 29, no. 4, p. 101717, 2020.

[17] G. Suganda, M. Asfi, R. T. Subagio, and R. P. Kusuma, "Penentuan Penerima Beasiswa KIP Menggunakan Naïve Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022.

[18] H. Suyono and M. Ridwan, "Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes," *EECCIS*, vol. 7, no. 1, pp. 23–28, 2013.

[19] A. Verma and K. K. Dhruw, "Comparative Performance of Random Forest and Naïve Bayes for Classification Problems," *Int. J. Comput. Appl.*, vol. 184, no. 45, pp. 32–38, 2023.

[20] Junaidi, "Pemilihan Penerima Beasiswa Menggunakan Metode Profile Matching," *Paradigma*, vol. 19, no. 2, 2017.

[21] M. Asfi and N. Fitrianingsih, "Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi," *J. Nas. Inform. dan Teknol. Jar.*, vol. 5, pp. 45–50, 2020.

[22] Hidayatunnisa'i, Kusrini, and Kusnawi, "Perbandingan Kinerja Metode Naïve Bayes Dan Support Machine (Svm) Dalam Analisis Kualitas Butir Soal Pilihan Ganda," *J. Fasilkom*, vol. 13, no. 02, pp. 173–180, 2023.

[23] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Elsevier, 2011.

[24] P. P. Widodo, R. T. Handayanto, and Herlawati, *Penerapan Data Mining dengan MATLAB*. Bandung: Rekayasa Sains, 2020.

[25] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.

[26] W. Djatmiko, Kusrini, and Hanafi, "Perbandingan Naïve Bayes dan Random Forest untuk Prediksi Perilaku Peserta Program Rujuk Balik," *J. Fasilkom*, vol. 13, no. 3, pp. 358–367, 2023.

[27] F. Nuraeni, D. Kurniadi, and G. F. Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima KIP-K Menggunakan Algoritma K-Means++," *J. Sisfokom*, vol. 11, no. 3, pp. 437–443, 2023.

[28] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering," *J. Focus Action Res. Math.*, vol. 5, no. 1, pp. 127–141, 2022.

Submitted : 15-12-2025 | Reviewed : 21-12-2025 | Accepted : 31-12-2025