Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

Penerapan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Beasiswa Berdasarkan Data Sosial Ekonomi Mahasiswa

Kevin Rasi Dauly Pardede*, Jaya Tata Hardinata, Togi Lumbantobing, Rahul Sinurat

Program Studi Ilmu Komputer, Fakultas FMIPA, Universitas HKBP Nomensen, Pematangsiantar, Indonesia
Jl. Sangnawaluh No. 4, Siopat Suhu, Kec. Siantar Timur, Kota Pematang Siantar, Sumatera Utara, Indonesia
Email: [1,*]kevinpardede047@gmail.com, [2]jayatatahardinata@uhnp.ac.id,[3]togisamuel93@gmail.com, [4]rahulsinurat848@gmail.com
Email Penulis Korespondensi: kevinpardede047@gmail.com

Abstrak−Program beasiswa merupakan salah satu bentuk dukungan perguruan tinggi dalam meningkatkan kualitas pendidikan sekaligus membantu mahasiswa yang memiliki keterbatasan ekonomi. Proses seleksi penerima beasiswa umumnya mempertimbangkan berbagai aspek sosial ekonomi, seperti penghasilan orang tua, jumlah tanggungan keluarga, indeks prestasi kumulatif (IPK), kepemilikan Kartu Indonesia Pintar (KIP), dan kondisi tempat tinggal. Pada praktiknya, proses seleksi masih banyak dilakukan secara manual sehingga membutuhkan waktu yang relatif lama, rentan terhadap kesalahan pencatatan, serta berpotensi menghasilkan keputusan yang kurang objektif karena bergantung pada penilaian subjektif petugas. Permasalahan tersebut mendorong perlunya suatu metode komputasi yang dapat membantu proses pengambilan keputusan secara lebih cepat, konsisten, dan terukur. Penelitian ini bertujuan menerapkan algoritma Naive Bayes, yaitu metode klasifikasi berbasis probabilitas yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar atribut, untuk mengklasifikasikan kelayakan penerima beasiswa berdasarkan data sosial ekonomi mahasiswa. Metode penelitian meliputi lima tahapan utama, yaitu pengumpulan data, preprocessing data (penanganan missing value, duplikasi, dan transformasi atribut kategorikal), pembagian dataset menggunakan metode Hold-Out Validation dengan komposisi 80% data latih dan 20% data uji, pelatihan model Naive Bayes menggunakan pustaka Scikit-learn, serta evaluasi performa model menggunakan Confusion Matrix. Dataset penelitian terdiri atas 200 data mahasiswa dengan tujuh atribut, yaitu penghasilan orang tua, jumlah tanggungan, IPK, status tempat tinggal, kepemilikan KIP, semester, dan status kelayakan. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan kelayakan penerima beasiswa dengan nilai Accuracy sebesar 92,50%, Precision sebesar 95,65%, Recall sebesar 91,67%, dan F1-Score sebesar 93,62%. Nilai tersebut membuktikan bahwa algoritma Naive Bayes dapat dijadikan metode pendukung pengambilan keputusan dalam proses seleksi penerima beasiswa secara lebih cepat, objektif, dan efisien dibandingkan proses seleksi manual.

Kata Kunci: Naive Bayes; Machine Learning; Beasiswa; Klasifikasi; Data Mining

Abstract−*Scholarship programs are one of the efforts made by universities to improve educational quality while supporting students with limited economic conditions. The scholarship selection process generally considers various socio-economic factors, including parents' income, number of family dependents, Grade Point Average (GPA), ownership of the Indonesia Smart Card (KIP), and housing status. In practice, the selection process is still largely performed manually, which is time-consuming, prone to recording errors, and potentially less objective because it depends on the subjective judgment of the staff involved. This condition highlights the need for a computational approach that can support faster, more consistent, and measurable decision-making. This study aims to implement the Naive Bayes algorithm, a probability-based classification method that works based on Bayes' Theorem with the assumption of conditional independence among attributes, to classify scholarship eligibility based on students' socio-economic data. The research method consists of five main stages: data collection, data preprocessing (handling missing values, duplicate records, and categorical attribute transformation), dataset splitting using the Hold-Out Validation method with a composition of 80% training data and 20% testing data, model training using the Naive Bayes algorithm implemented with the Scikit-learn library, and performance evaluation using a Confusion Matrix. The dataset consists of 200 student records with seven attributes, namely parents' income, number of dependents, GPA, housing status, KIP ownership, semester level, and eligibility status. The results show that the Naive Bayes algorithm is able to classify scholarship eligibility with an Accuracy of 92.50%, Precision of 95.65%, Recall of 91.67%, and F1-Score of 93.62%. These results confirm that the Naive Bayes algorithm can be used as a decision support method to make the scholarship selection process faster, more objective, and more efficient than manual selection.*

Keywords: *Naive Bayes; Machine Learning; Scholarship; Classification; Data Mining*

1. PENDAHULUAN

Pendidikan merupakan salah satu faktor utama dalam meningkatkan kualitas sumber daya manusia serta mendukung pembangunan nasional. Perguruan tinggi sebagai lembaga pendidikan memiliki peran penting dalam memberikan kesempatan kepada seluruh mahasiswa untuk memperoleh pendidikan yang berkualitas tanpa terkendala kondisi ekonomi. Salah satu bentuk dukungan yang diberikan adalah melalui program beasiswa yang bertujuan membantu mahasiswa yang memiliki prestasi akademik maupun keterbatasan ekonomi. [1]. Dalam proses penentuan penerima beasiswa, pihak perguruan tinggi biasanya mempertimbangkan beberapa kriteria seperti pendapatan orang tua, jumlah tanggungan keluarga, Indeks Prestasi Kumulatif (IPK), kepemilikan Kartu Indonesia Pintar (KIP), kondisi tempat tinggal, serta semester mahasiswa. Banyaknya jumlah pendaftar menyebabkan proses seleksi menjadi semakin kompleks. Selain itu, proses seleksi yang masih dilakukan secara manual sering kali membutuhkan waktu yang lama, rentan terhadap kesalahan, dan berpotensi menimbulkan subjektivitas dalam pengambilan keputusan. [2].

Perkembangan teknologi informasi, khususnya pada bidang Machine Learning, memberikan solusi terhadap permasalahan tersebut. Machine Learning mampu mempelajari pola dari data historis sehingga dapat digunakan untuk membantu proses pengambilan keputusan secara otomatis. Salah satu algoritma klasifikasi yang banyak digunakan adalah Naive Bayes karena memiliki proses komputasi yang sederhana, cepat, dan mampu memberikan hasil klasifikasi yang cukup baik meskipun menggunakan jumlah data yang relatif terbatas. Naive Bayes merupakan algoritma klasifikasi

Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

berbasis probabilitas yang bekerja berdasarkan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Meskipun asumsi tersebut sederhana, berbagai penelitian menunjukkan bahwa algoritma Naive Bayes memiliki performa yang kompetitif pada berbagai kasus klasifikasi, seperti klasifikasi penyakit, prediksi kelulusan mahasiswa, analisis sentimen, hingga sistem pendukung keputusan penerima bantuan sosial.

Beberapa penelitian terdahulu telah menerapkan algoritma Naive Bayes pada permasalahan penentuan penerima beasiswa dengan hasil yang bervariasi. Nur dkk. menerapkan Naive Bayes untuk menentukan penerima Beasiswa Program Indonesia Pintar (PIP) dan memperoleh tingkat akurasi yang cukup baik menggunakan atribut nilai akademik dan kondisi ekonomi keluarga. Hidayat dkk. melakukan klasifikasi mahasiswa calon penerima beasiswa KIP di Universitas Tomakaka Mamuju dengan memanfaatkan atribut demografis dan akademik mahasiswa. Thoriq dkk. menerapkan Naive Bayes dalam memprediksi penerimaan mahasiswa penerima beasiswa KIP di Universitas Adzkia dan menunjukkan bahwa algoritma tersebut efektif digunakan pada dataset berukuran kecil hingga menengah. Siswanto dan Normawati membangun sistem klasifikasi monitoring dan evaluasi kelayakan penerima beasiswa UAD berbasis Naive Bayes, sedangkan Febri dan Sari menerapkan Naive Bayes Classifier untuk menentukan penerima Beasiswa Bank Indonesia. Putri dkk. mengoptimalkan pemilihan atribut menggunakan K-Means Clustering sebelum proses klasifikasi Naive Bayes untuk meningkatkan akurasi. Navizah dan Fatah serta Isnanto dan Sulaiman membandingkan performa Naive Bayes dengan algoritma lain seperti Decision Tree, C4.5, dan KNN dalam konteks seleksi beasiswa, dan menemukan bahwa Naive Bayes tetap kompetitif meskipun sederhana secara komputasi. Jelita dkk. juga menerapkan metode Machine Learning, termasuk Naive Bayes, untuk klasifikasi penerima Beasiswa Kartu Indonesia Pintar Kuliah.

Berdasarkan kajian terhadap penelitian-penelitian tersebut, sebagian besar penelitian sebelumnya menerapkan Naive Bayes pada satu jenis beasiswa tertentu, seperti PIP atau KIP, dengan atribut yang relatif terbatas, serta belum banyak yang secara khusus mengombinasikan atribut penghasilan orang tua, jumlah tanggungan keluarga, IPK, status tempat tinggal, kepemilikan KIP, dan semester aktif secara bersamaan dalam satu model klasifikasi yang dievaluasi secara menyeluruh menggunakan Confusion Matrix (Accuracy, Precision, Recall, dan F1-Score). Kondisi tersebut menjadi celah (gap) yang mendasari penelitian ini, yaitu perlunya model klasifikasi kelayakan penerima beasiswa yang mempertimbangkan atribut sosial ekonomi secara lebih komprehensif serta dievaluasi secara terukur agar dapat diandalkan sebagai alat bantu pengambilan keputusan. Kontribusi utama penelitian ini adalah membangun dan mengevaluasi model klasifikasi berbasis Naive Bayes dengan kombinasi atribut sosial ekonomi mahasiswa yang lebih lengkap, serta menyajikan proses perhitungan probabilitas secara rinci sehingga dapat menjadi acuan implementasi pada institusi pendidikan lain dengan karakteristik data yang serupa.

Pada penelitian ini, algoritma Naive Bayes diterapkan untuk mengklasifikasikan kelayakan penerima beasiswa berdasarkan data sosial ekonomi mahasiswa. Data yang digunakan meliputi penghasilan orang tua, jumlah tanggungan keluarga, IPK, status tempat tinggal, kepemilikan KIP, serta semester mahasiswa. Dengan memanfaatkan algoritma tersebut diharapkan proses seleksi penerima beasiswa menjadi lebih cepat, objektif, serta mampu meningkatkan akurasi dalam menentukan mahasiswa yang benar-benar layak menerima bantuan pendidikan. Penelitian ini bertujuan membangun model klasifikasi menggunakan algoritma Naive Bayes serta mengevaluasi performanya melalui pengukuran akurasi, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat menjadi rekomendasi bagi perguruan tinggi dalam mengembangkan sistem pendukung keputusan untuk proses seleksi penerima beasiswa yang lebih efektif dan transparan.

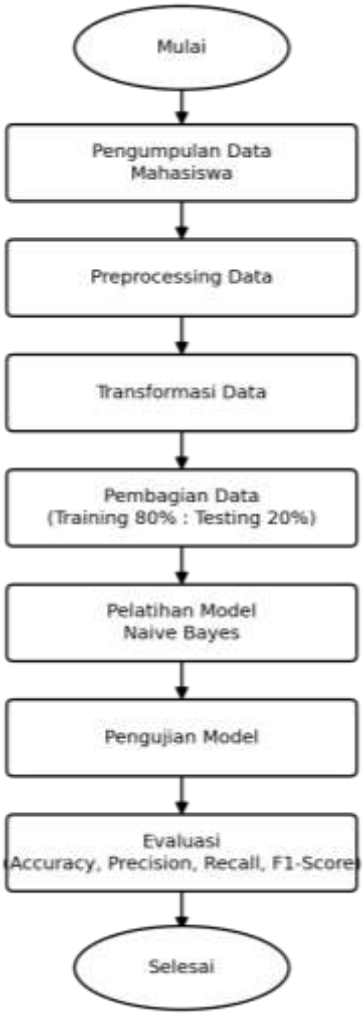
Selain unggul dari sisi kesederhanaan komputasi, algoritma Naive Bayes juga banyak dipilih pada penelitian bidang pendidikan karena kemampuannya bekerja secara stabil meskipun jumlah data latih yang tersedia relatif terbatas, suatu kondisi yang umum dijumpai pada institusi pendidikan dengan jumlah mahasiswa penerima beasiswa yang tidak terlalu besar. Kemampuan tersebut menjadikan Naive Bayes sebagai pilihan yang realistis untuk diterapkan pada skala perguruan tinggi tanpa memerlukan infrastruktur komputasi yang besar maupun proses tuning parameter yang rumit sebagaimana dibutuhkan oleh beberapa algoritma pembelajaran mesin lain seperti Support Vector Machine atau Artificial Neural Network. Dengan mempertimbangkan karakteristik permasalahan, ketersediaan data, serta kebutuhan akan hasil klasifikasi yang mudah diinterpretasikan oleh pengambil keputusan di lingkungan akademik, algoritma Naive Bayes dinilai sebagai pendekatan yang paling sesuai untuk diterapkan pada penelitian ini.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti



Gambar 1. Tahapan Penelitian

Tahapan penelitian yang dilakukan pada penelitian ini terdiri atas tujuh proses utama sebagaimana ditunjukkan pada Gambar 1. Tahap pertama adalah pengumpulan data, yaitu proses memperoleh data sosial ekonomi mahasiswa yang menjadi dasar dalam seleksi penerima beasiswa, dengan setiap data memiliki beberapa atribut yang menjadi variabel penelitian. Tahap kedua adalah preprocessing data, yang bertujuan membersihkan data dari nilai kosong (missing value), data ganda (duplicate), maupun data yang tidak konsisten agar kualitas dataset menjadi lebih baik sebelum digunakan pada proses klasifikasi. Tahap ketiga adalah transformasi data, yaitu proses mengonversi data kategorikal ke dalam bentuk numerik yang sesuai agar dapat diproses oleh algoritma Naive Bayes, termasuk melakukan normalisasi apabila diperlukan. Tahap keempat adalah pembagian dataset menjadi data latih (training data) sebesar 80% dan data uji (testing data) sebesar 20% menggunakan metode Hold-Out Validation. Tahap kelima adalah pelatihan model, yaitu proses membangun model klasifikasi menggunakan data latih dengan algoritma Naive Bayes yang bekerja berdasarkan Teorema Bayes. [3]

Tahap keenam adalah pengujian model, yaitu proses menguji model yang telah dibangun menggunakan data uji untuk mengetahui kemampuan model dalam mengklasifikasikan data baru yang belum pernah dipelajari sebelumnya. Tahap ketujuh atau tahap terakhir adalah evaluasi, yaitu proses pengukuran performa model menggunakan Confusion Matrix untuk memperoleh nilai Accuracy, Precision, Recall, dan F1-Score. [4] Penerapan solusi/metode penelitian secara konkret terjadi pada tahap kelima (Pelatihan Model) yang ditunjukkan pada Gambar 1, yaitu ketika algoritma Naive Bayes mempelajari pola dari data latih untuk membentuk model probabilitas yang selanjutnya diuji pada tahap keenam (Pengujian Model).

2.2 Kajian Pustaka Algoritma Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Meskipun asumsi tersebut sederhana, algoritma ini memiliki performa yang baik pada berbagai permasalahan klasifikasi. Probabilitas suatu kelas dihitung menggunakan persamaan berikut.

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

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

Pada persamaan tersebut, P(C|X) menyatakan probabilitas suatu data termasuk ke dalam kelas C dengan diketahui atribut X, P(X|C) menyatakan probabilitas atribut X terhadap kelas C (likelihood), P(C) menyatakan probabilitas awal (prior) dari kelas C, dan P(X) menyatakan probabilitas atribut X secara umum. Keempat komponen tersebut dihitung dari data training untuk kemudian digunakan dalam menentukan kelas dari data baru pada tahap pengujian.
Pada penelitian ini, kelas yang digunakan terdiri atas:

a. Layak menerima beasiswa
b. Tidak layak menerima beasiswa

Model akan menghitung probabilitas masing-masing kelas berdasarkan data training, kemudian memilih kelas dengan nilai probabilitas terbesar sebagai hasil klasifikasi. Pemilihan algoritma Naive Bayes pada penelitian ini juga didasarkan pada beberapa kajian terdahulu yang menunjukkan keunggulan algoritma tersebut dibandingkan metode klasifikasi lain pada permasalahan yang serupa. Ikhsan dkk. membandingkan algoritma K-Nearest Neighbor, Naive Bayes, dan CART dalam memprediksi penerima beasiswa dan menemukan bahwa Naive Bayes memberikan hasil yang kompetitif dengan waktu komputasi yang lebih singkat Sejalan dengan hal tersebut, penelitian oleh Thoriq dkk. yang membandingkan Naive Bayes dengan Random Forest pada seleksi penerima KIP juga menunjukkan bahwa Naive Bayes tetap layak digunakan meskipun struktur modelnya jauh lebih sederhana.

Kesederhanaan struktur model Naive Bayes memberikan keuntungan tambahan berupa kemudahan interpretasi, yaitu setiap nilai probabilitas yang dihasilkan model dapat ditelusuri kembali ke kontribusi masing-masing atribut, sehingga pihak pengambil keputusan di perguruan tinggi dapat memahami alasan di balik suatu hasil klasifikasi dan tidak memperlakukan model sebagai kotak hitam (black box). Karakteristik inilah yang menjadikan Naive Bayes relevan digunakan pada konteks pengambilan keputusan yang bersifat administratif dan membutuhkan akuntabilitas, seperti proses seleksi penerima beasiswa yang menjadi fokus penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Dataset

Dataset penelitian terdiri dari 200 data mahasiswa yang diperoleh berdasarkan kriteria seleksi penerima beasiswa. Setiap data memiliki atribut yang merepresentasikan kondisi sosial ekonomi mahasiswa. Variabel yang digunakan meliputi penghasilan orang tua, jumlah tanggungan keluarga, Indeks Prestasi Kumulatif (IPK), status tempat tinggal, kepemilikan Kartu Indonesia Pintar (KIP), semester aktif, dan status kelayakan penerima beasiswa. Ketujuh variabel tersebut disajikan secara rinci pada Tabel 1. Contoh data mahasiswa yang digunakan pada penelitian ini ditunjukkan pada Tabel 2.

Tabel 1. Variabel Penelitian

No	Variabel	Keterangan
1	Penghasilan Orang Tua	Pendapatan per bulan
2	Jumlah Tanggungan	Jumlah anggota keluarga yang menjadi tanggungan
3	IPK	Indeks Prestasi Kumulatif
4	Status Tempat Tinggal	Milik sendiri, kontrak, atau kos
5	Kepemilikan KIP	Ya / Tidak
6	Semester	Semester aktif mahasiswa
7	Status Beasiswa	Layak / Tidak Layak

Masing-masing variabel dipilih karena memiliki pengaruh terhadap kondisi sosial ekonomi mahasiswa serta sering digunakan sebagai persyaratan dalam proses seleksi penerima beasiswa. [5]

Tabel 2. Contoh Dataset Penelitian

No	Penghasilan	Tanggungan	IPK	Tempat Tinggal	KIP	Semester	Status
1	< Rp2.000.000	5	3.82	Kos	Ya	4	Layak
2	Rp4–6 Juta	2	3.25	Milik Sendiri	Tidak	6	Tidak Layak
3	< Rp2.000.000	4	3.74	Kontrak	Ya	2	Layak
4	> Rp6.000.000	1	3.15	Milik Sendiri	Tidak	8	Tidak Layak
5	Rp2–4 Juta	3	3.68	Kos	Ya	4	Layak

Berdasarkan hasil pengumpulan data, terdapat dua kelas klasifikasi yaitu Layak dan Tidak Layak. [6] Distribusi kelas ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Kelas

Kelas	Jumlah
Layak	118
Tidak Layak	82
Total	200

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: 2963-0169 (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

Distribusi data menunjukkan bahwa sebagian besar mahasiswa berada pada kategori Layak, sehingga model perlu diuji agar tidak mengalami bias terhadap kelas mayoritas. [7]

3.2 Preprocessing Data

Tahapan preprocessing dilakukan untuk meningkatkan kualitas dataset sebelum digunakan pada proses klasifikasi. Tahapan yang dilakukan meliputi pemeriksaan nilai kosong (missing value), penghapusan data duplikat, pemeriksaan inkonsistensi penulisan data, transformasi atribut kategorikal menjadi format yang dapat diproses oleh algoritma, serta pembagian dataset menjadi data latih (training) dan data uji (testing). Hasil preprocessing menunjukkan bahwa seluruh data telah memenuhi kriteria sehingga dapat digunakan dalam proses pembentukan model klasifikasi. [8]

3.3 Pembentukan Model Naive Bayes

Dataset dibagi menggunakan metode Hold-Out Validation dengan komposisi 80% data training dan 20% data testing sebagaimana ditunjukkan pada Tabel 4. Pada tahap pelatihan (training), algoritma Naive Bayes menghitung probabilitas awal (prior probability) dan probabilitas setiap atribut terhadap masing-masing kelas. [9]

Tabel 4. Pembagian Dataset

Jenis Data	Jumlah
Training Data	160
Testing Data	40
Total Dataset	200

Data training digunakan untuk membangun model klasifikasi, sedangkan data testing digunakan untuk mengevaluasi performa model terhadap data yang belum pernah dipelajari sebelumnya. [10]. Sebagai contoh, probabilitas awal kelas dihitung sebagai berikut: P(Layak) = 118/200 = 0,59 dan P(Tidak Layak) = 82/200 = 0,41. Selanjutnya dihitung probabilitas setiap atribut terhadap kelas menggunakan data training. Karena sebagian atribut penelitian ini bersifat numerik kontinu, seperti IPK dan penghasilan orang tua, algoritma Gaussian Naive Bayes menghitung nilai likelihood P(X|C) menggunakan fungsi kepadatan probabilitas (probability density function) distribusi normal berdasarkan nilai rata-rata (mean) dan simpangan baku (standard deviation) atribut tersebut pada setiap kelas. Sebagai ilustrasi, apabila pada kelas Layak atribut IPK memiliki rata-rata 3,55 dengan simpangan baku 0,22, maka data uji dengan IPK 3,60 akan memiliki nilai likelihood yang tinggi terhadap kelas Layak karena nilainya berada sangat dekat dengan rata-rata kelas tersebut. Nilai likelihood dari seluruh atribut kemudian dikalikan satu sama lain beserta probabilitas prior kelas untuk memperoleh nilai probabilitas posterior setiap kelas, dan kelas dengan nilai posterior terbesar dipilih sebagai hasil klasifikasi akhir data tersebut.[11].

Sebagai kelanjutan ilustrasi tersebut, nilai likelihood P(IPK=3,60|Layak) yang dihitung menggunakan fungsi kepadatan probabilitas Gaussian dengan rata-rata 3,55 dan simpangan baku 0,22 menghasilkan nilai sebesar 1,7671, sedangkan apabila diasumsikan kelas Tidak Layak memiliki rata-rata IPK sebesar 3,10 dengan simpangan baku 0,25, maka nilai likelihood P(IPK=3,60|Tidak Layak) yang diperoleh hanya sebesar 0,2160. Selisih yang cukup besar antara kedua nilai likelihood tersebut menunjukkan bagaimana atribut IPK berkontribusi kuat dalam membedakan kedua kelas.[11].

Nilai likelihood atribut numerik tersebut selanjutnya dikalikan dengan nilai likelihood atribut kategorikal serta probabilitas prior kelas untuk memperoleh nilai posterior akhir. Sebagai contoh, apabila data uji yang sama juga memiliki KIP dengan status kepemilikan “Ya” dan status tempat tinggal “Kos”, dengan asumsi P(KIP=Ya|Layak)=0,83, P(KIP=Ya|Tidak Layak)=0,22, P(Tempat Tinggal=Kos|Layak)=0,35, dan P(Tempat Tinggal=Kos|Tidak Layak)=0,18, maka nilai posterior tidak ternormalisasi untuk kelas Layak adalah $0,59 \times 1,7671 \times 0,83 \times 0,35 \approx 0,3029$, sedangkan untuk kelas Tidak Layak adalah $0,41 \times 0,2160 \times 0,22 \times 0,18 \approx 0,0035$. Karena nilai posterior kelas Layak jauh lebih besar dibandingkan kelas Tidak Layak, maka data uji tersebut diklasifikasikan ke dalam kelas Layak.

Contoh perhitungan ini memperlihatkan secara konkret bagaimana algoritma Naive Bayes menyelesaikan permasalahan klasifikasi kelayakan penerima beasiswa, yaitu dengan menggabungkan kontribusi seluruh atribut sosial ekonomi mahasiswa ke dalam satu nilai probabilitas akhir yang dapat dibandingkan secara langsung antar kelas, tanpa memerlukan proses optimisasi tambahan seperti pada algoritma pembelajaran mesin berbasis iteratif.[11]

3.4 Hasil Klasifikasi

Setelah proses pelatihan selesai, model diuji menggunakan 40 data testing. Hasil prediksi dibandingkan dengan data sebenarnya untuk mengetahui tingkat keberhasilan model, sebagaimana ditunjukkan pada contoh hasil prediksi pada Tabel 5.

Tabel 5. Contoh Hasil Prediksi

Data	Aktual	Prediksi
1	Layak	Layak
2	Tidak Layak	Tidak Layak
3	Layak	Layak
4	Layak	Tidak Layak

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

Data	Aktual	Prediksi
5	Tidak Layak	Tidak Layak

Dari hasil pengujian terlihat bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh algoritma Naive Bayes. [12]

3.5 Evaluasi Menggunakan Confusion Matrix

Evaluasi model dilakukan menggunakan Confusion Matrix yang hasilnya ditunjukkan pada Tabel 6.

Tabel 6. Confusion Matrix

	Prediksi Layak	Prediksi Tidak Layak
Aktual Layak	22	2
Aktual Tidak Layak	1	15

Berdasarkan Confusion Matrix tersebut diperoleh nilai True Positive (TP) sebesar 22, True Negative (TN) sebesar 15, False Positive (FP) sebesar 1, dan False Negative (FN) sebesar 2. Selanjutnya dilakukan perhitungan metrik evaluasi menggunakan persamaan (2) hingga (5) berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

Rekapitulasi hasil pengukuran performa model menggunakan keempat metrik evaluasi tersebut secara ringkas ditunjukkan pada Tabel 7.

Tabel 7. Hasil Evaluasi Model

Metrik	Nilai
Accuracy	92,50%
Precision	95,65%
Recall	91,67%
F1-Score	93,62%

Tabel 7 menunjukkan bahwa model Naive Bayes yang dibangun memiliki performa klasifikasi yang tinggi pada seluruh metrik evaluasi, dengan nilai Accuracy sebesar 92,50%, Precision sebesar 95,65%, Recall sebesar 91,67%, dan F1-Score sebesar 93,62%. Nilai Precision yang lebih tinggi dibandingkan Recall mengindikasikan bahwa model lebih jarang salah memprediksi mahasiswa Tidak Layak sebagai Layak (False Positive) dibandingkan sebaliknya, sehingga model ini relatif aman digunakan sebagai alat bantu pengambilan keputusan pada proses seleksi penerima beasiswa.

3.6 Implementasi

Pada penelitian ini, implementasi dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka (library) Pandas, NumPy, dan Scikit-learn. Implementasi bertujuan untuk membangun model klasifikasi menggunakan algoritma Naive Bayes berdasarkan data sosial ekonomi mahasiswa serta mengevaluasi performa model dalam menentukan kelayakan penerima beasiswa. Tahapan implementasi dimulai dengan membaca dataset, melakukan preprocessing, membagi data menjadi data latih dan data uji, membangun model Naive Bayes, melakukan prediksi terhadap data uji, dan mengevaluasi hasil klasifikasi menggunakan Confusion Matrix. [13]

3.6.1 Lingkungan Implementasi

Spesifikasi perangkat yang digunakan pada penelitian ditunjukkan pada Tabel 8.

Tabel 8. Spesifikasi Perangkat Implementasi

Komponen	Spesifikasi
Sistem Operasi	Windows 11 64-bit
Bahasa Pemrograman	Python 3.12
IDE	Google Colaboratory / Visual Studio Code
Library	Pandas, NumPy, Scikit-learn, Matplotlib
Processor	Intel Core i5 atau setara
RAM	8 GB

Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

Lingkungan tersebut dipilih karena mampu mendukung proses pengolahan data, pelatihan model, serta evaluasi algoritma Naive Bayes secara efisien. [14]

3.6.2 Implementasi Algoritma Naive Bayes

Tahapan implementasi diawali dengan membaca dataset mahasiswa menggunakan library Pandas. Dataset kemudian dipisahkan menjadi atribut (feature) dan kelas (label). Selanjutnya dilakukan pembagian data menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan metode Hold-Out Validation. [15]. Model klasifikasi dibangun menggunakan algoritma Gaussian Naive Bayes dari pustaka Scikit-learn. Setelah model selesai dilatih, dilakukan proses prediksi terhadap data uji untuk memperoleh hasil klasifikasi kelayakan penerima beasiswa. Secara umum, tahapan implementasi terdiri atas proses impor dataset, preprocessing data, pembagian data training dan testing, pelatihan model Naive Bayes, prediksi data testing, dan evaluasi menggunakan Confusion Matrix.

3.6.3 Hasil Implementasi

Setelah model selesai dilatih menggunakan data training sebanyak 160 data, dilakukan pengujian terhadap 40 data testing. Hasil prediksi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan sesuai dengan kondisi sebenarnya, dengan performa evaluasi sebagaimana telah ditunjukkan pada Tabel 7. [16]

3.7 Pembahasan

Hasil pengujian menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan kelayakan penerima beasiswa dengan tingkat akurasi sebesar 92,50%. Nilai precision sebesar 95,65% menunjukkan bahwa sebagian besar mahasiswa yang diprediksi layak benar-benar termasuk dalam kategori layak menerima beasiswa, sedangkan nilai recall sebesar 91,67% menunjukkan bahwa model mampu mengidentifikasi mayoritas mahasiswa yang memang layak menerima beasiswa. Tingkat F1-Score sebesar 93,62% mengindikasikan keseimbangan yang baik antara precision dan recall sehingga model memiliki performa klasifikasi yang stabil. [17]

Secara teknis, proses klasifikasi pada algoritma Naive Bayes diawali dengan penghitungan probabilitas awal (prior) untuk setiap kelas berdasarkan proporsi data training, yaitu P(Layak) sebesar 0,59 dan P(Tidak Layak) sebesar 0,41 sebagaimana ditunjukkan pada bagian 3.3. Selanjutnya, untuk setiap data uji, model menghitung nilai likelihood setiap atribut terhadap masing-masing kelas menggunakan fungsi kepadatan probabilitas Gaussian, kemudian mengalikan seluruh nilai likelihood tersebut dengan probabilitas prior kelas yang bersesuaian sesuai persamaan (1). Kelas dengan nilai probabilitas posterior terbesar kemudian dipilih sebagai hasil prediksi akhir. Pendekatan inilah yang memungkinkan algoritma Naive Bayes menyelesaikan permasalahan klasifikasi kelayakan penerima beasiswa secara efisien tanpa memerlukan proses optimisasi parameter yang rumit seperti pada algoritma pembelajaran mesin lain.

Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi yang ditunjukkan oleh nilai False Positive dan False Negative pada Tabel 6. Kesalahan tersebut disebabkan oleh adanya karakteristik data mahasiswa yang memiliki kondisi sosial ekonomi hampir serupa, misalnya mahasiswa dengan penghasilan orang tua menengah namun jumlah tanggungan keluarga yang berbeda, sehingga probabilitas antar kelas menjadi sangat berdekatan. Selain itu, asumsi independensi antar atribut pada algoritma Naive Bayes juga dapat memengaruhi hasil klasifikasi ketika terdapat hubungan yang kuat antarvariabel, misalnya antara penghasilan orang tua dan status tempat tinggal.

Dari sisi implementasi, hasil ini konsisten dengan pengamatan selama proses pelatihan model menggunakan pustaka Scikit-learn, di mana atribut seperti penghasilan orang tua, jumlah tanggungan keluarga, kepemilikan KIP, serta IPK memberikan kontribusi yang cukup besar terhadap pemisahan kedua kelas. Keunggulan utama algoritma Naive Bayes yang teramati pada implementasi ini adalah proses komputasi yang cepat dan sederhana, sehingga model dapat dilatih dan diuji dalam waktu yang singkat meskipun dijalankan pada spesifikasi perangkat kelas menengah sebagaimana ditunjukkan pada Tabel 8, dan tetap cocok digunakan pada dataset berukuran kecil hingga menengah tanpa memerlukan proses pelatihan yang kompleks. [18]

Temuan ini juga sejalan dengan penelitian-penelitian terdahulu yang menerapkan Naive Bayes pada konteks seleksi penerima beasiswa maupun bantuan pendidikan lainnya, yang secara konsisten melaporkan tingkat akurasi di atas 85% meskipun menggunakan jumlah atribut dan ukuran dataset yang berbeda-beda [1], [2], [4], [5]. Konsistensi hasil tersebut memperkuat argumen bahwa Naive Bayes merupakan algoritma yang robust untuk permasalahan klasifikasi berbasis data sosial ekonomi, sekaligus menunjukkan bahwa pendekatan probabilistik yang digunakan mampu menangkap pola kelayakan penerima beasiswa secara memadai tanpa memerlukan rekayasa fitur yang rumit. Dengan demikian, kontribusi utama penelitian ini terletak pada penyajian proses klasifikasi secara transparan mulai dari perhitungan probabilitas prior, likelihood, hingga posterior, sehingga pihak perguruan tinggi yang ingin mengadopsi pendekatan serupa dapat memahami dan mereplikasi seluruh tahapan pengambilan keputusan model secara utuh.[18]

Secara keseluruhan, baik dari sisi pengujian model maupun implementasinya, hasil penelitian menunjukkan bahwa algoritma Naive Bayes layak diterapkan sebagai metode klasifikasi dalam sistem pendukung keputusan seleksi penerima beasiswa. [19] Dengan proses komputasi yang sederhana, waktu pelatihan yang cepat, kemudahan implementasi, serta tingkat akurasi yang tinggi, algoritma ini dapat membantu pihak perguruan tinggi melakukan proses seleksi secara lebih objektif, efisien, dan transparan dibandingkan proses seleksi manual, sekaligus menjadi dasar pengembangan sistem pendukung keputusan berbasis data pada institusi pendidikan lain dengan karakteristik permasalahan yang serupa. [20]

Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Naive Bayes berhasil diterapkan untuk membangun model klasifikasi kelayakan penerima beasiswa dengan memanfaatkan tujuh atribut sosial ekonomi mahasiswa, yaitu penghasilan orang tua, jumlah tanggungan keluarga, Indeks Prestasi Kumulatif (IPK), status tempat tinggal, kepemilikan Kartu Indonesia Pintar (KIP), semester aktif, dan status kelayakan sebagai label kelas. Pengujian terhadap 40 data menggunakan skema Hold-Out Validation 80:20 menunjukkan bahwa model mampu mencapai nilai Accuracy sebesar 92,50%, Precision sebesar 95,65%, Recall sebesar 91,67%, dan F1-Score sebesar 93,62%, yang membuktikan bahwa model memiliki kemampuan generalisasi yang baik dalam membedakan mahasiswa yang layak dan tidak layak menerima beasiswa meskipun masih terdapat sejumlah kecil kesalahan klasifikasi akibat kemiripan karakteristik data antar kelas. Proses komputasi yang sederhana dan waktu pelatihan yang singkat menjadikan Naive Bayes sebagai pilihan yang praktis untuk diimplementasikan pada sistem pendukung keputusan seleksi beasiswa di lingkungan perguruan tinggi, khususnya pada institusi dengan jumlah pendaftar yang besar dan sumber daya evaluasi manual yang terbatas. Dengan demikian, pendekatan berbasis Machine Learning ini dapat menjadi alternatif yang objektif, konsisten, dan efisien dibandingkan proses seleksi manual, sekaligus membuka peluang pengembangan lebih lanjut melalui penambahan volume dan keberagaman data serta perbandingan dengan algoritma klasifikasi lain seperti Decision Tree, Random Forest, Support Vector Machine, atau K-Nearest Neighbor untuk memperoleh metode yang paling optimal bagi proses seleksi penerima beasiswa.

UCAPAN TERIMAKASIH

Penulis menyampaikan terima kasih kepada Program Studi Ilmu Komputer, Fakultas FMIPA, Universitas HKBP Nomensen Pematangsiantar, serta seluruh pihak yang telah mendukung terlaksananya penelitian ini.

REFERENCES

[1] M. Rayhans Adrian, F. Alfarizi Saragih, M. Rifqi Arrafi, and A. Taufik Al Afkari Siahaan, "Jurnal Restikom : Riset Teknik Informatika dan Komputer Sistem Monitoring dan Prediksi Kelayakan Kandidat Beasiswa Menggunakan Machine Learning Berbasis Web pada Biro KESRA Provinsi Sumatera Utara," vol. 7, no. 3, pp. 306–317, 2025, [Online]. Available: https://restikom.nusaputra.ac.id

[2] M. A. Juniawan and A. Sudrajat, "Prediksi Seleksi Penerima Beasiswa Kip Kuliah Menggunakan Algoritma Naïve Bayes (Studi Kasus Politeknik Tedc Bandung)," 2025.

[3] M. Thoriq, F. Maulana, and A. Qurrata Ayun, "Perbandingan Naïve Bayes dan Random Forest pada Seleksi KIP di Universitas Adzkia," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 5, no. 3, pp. 602–610, Dec. 2025, doi: 10.55382/jurnalpustakaai.v5i3.1385.

[4] D. Marpaung, W. Ramdhan, and M. Ishan, "Penerapan Naive Bayes Dalam Menentukan Kuota Program Indonesia Pintar Pada SMP Negeri 5 Kota Tanjungbalai," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1350–1358, Mar. 2025, doi: 10.29100/jipi.v10i2.6034.

[5] A. Rizki Siregar and Mhd. Furqan, "Classification of Scholarships for Students in Schools Using the Naïve Bayes Method," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 1, pp. 278–289, Feb. 2025, doi: 10.47709/cnahpc.v7i1.5417.

[6] M. Ariandi Sitorus, A. Anita, and E. Agustian, "Analisis Metode Naïve Byes Classifer pada Penentuan Penerima Beasiswa Bidikmisi di Universitas Prima Indonesia," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 5, no. 2, pp. 132–141, Aug. 2025, doi: 10.55382/jurnalpustakaai.v5i2.999.

[7] A. Rachmat Raharja, A. Pramudianto, and Y. Muchsam, "Application of Naïve Bayes Classifier Algorithm for Classification of Scholarship Recipients at SMA PGRI 2 Bandung," 2024. [Online]. Available: www.trigin.pelnus.ac.id

[8] D. Sebagai and S. S. Syarat, "Penerapan Metode Naïve Bayes Classifier Pada Klasifikasi Penerimaan Beasiswa Kartu Indonesia Pintar Tugas Akhir."

[9] J. Elektronika and D. Komputer, "Sistem Pendukung Keputusan Seleksi Penerima Beasiswa Dengan Metode Naïve Bayes Classifier," vol. 16, no. 1, pp. 156–162, 2023, [Online]. Available: https://journal.stekom.ac.id/index.php/elkom■page156

[10] I. Agustina, G. Dwilestari, and A. R. Rinaldi, "Implementasi Data Mining Pada Proses Seleksi Beasiswa Menggunakan Naive Bayes Dan Backward Elimination".

[11] A. Ceselli, "Università degli Studi di Milano Master Degree in Computer Science Information Management course."

[12] A. N. Ikhsan, P. Subarkah, and R. S. Alifian, "Komparasi Algoritme K-NN, Naïve Bayes, dan Cart untuk Memprediksi Penerima Beasiswa," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 2, Oct. 2023, doi: 10.23887/jstundiksha.v12i2.51745.

[13] M. Hidayat *et al.*, "INFO ARTIKEL ABSTRAK," vol. 2, no. 4, pp. 172–180, 2023, doi: 10.55123.

[14] A. nugraha, F. Al-ghifari, and A. Panigoran, "Implementasi Algoritma Naïve Bayes Classifier Untuk Klasifikasi Penerima Beasiswa BUMN Sekota Dumai," *Jurnal Karya Ilmiah Multidisiplin (JURKIM)*, vol. 14, no. 2, pp. 77–85, 2024.

[15] U. Islam Negeri Sultan Syarif Kasim Riau Simpang Baru *et al.*, "Klasifikasi Penerima Beasiswa Kartu Indonesia Pintar Kuliah dengan Metode Machine Learning," 2026. [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id/index.php/joincos

[16] B. Isnanto and R. Sulaiman, "An Empirical Comparison of C4.5, Naive Bayes, and KNN for Scholarship Selection," *Journal of Information Systems and Informatics*, vol. 8, no. 3, pp. 3252–3268, Jun. 2026, doi: 10.63158/journalisi.v8i3.1617.

[17] Z. Fatah and P. Navizah, "Perbandingan Klasfikasi Status Beasiswa Mahasiswa Menggunakan Algoritma Decision Tree dan Algoritma Naive Bayes," 2026.

Vol. 4, No. 2, Juli 2026, Hal 63 - 71
ISSN: **2963-0169** (Online - Elektronik)
DOI: 10.62866/jurikti.v4i2.323
https://journal.grahamitra.id/index.php/jurikti

[18] F. M. Febri and D. P. Sari, "Determination Of Bank Indonesia Scholarship Recipients Using Naïve Bayes Classifier," *Barekeng*, vol. 17, no. 3, pp. 1595–1604, Sep. 2023, doi: 10.30598/barekengvol17iss3pp1595-1604.

[19] D. B. Siswanto and D. Normawati, "Sistem Klasifikasi Monitoring dan Evaluasi Kelayakan Penerima Beasiswa UAD Menggunakan Algoritma Naïve Bayes," *Jurnal SAINTEKOM*, vol. 13, no. 2, pp. 161–172, Sep. 2023, doi: 10.33020/saintekom.v13i2.428.

[20] A. M. Nur, N. Nurhidayati, and I. Fathurrahman, "Penerapan Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP).," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 1, pp. 93–102, Jan. 2024, doi: 10.29408/jit.v7i1.23995.