

## Klasifikasi Mahasiswa Calon Penerima Beasiswa KIP Menggunakan Algoritma Naive Bayes di Universitas Tomakaka Mamuju

Hidayat<sup>1</sup>, Musliadi KH<sup>2</sup>, Indar Kusmanto<sup>3</sup>, Munawirah Kadir<sup>4</sup>, Kristian<sup>5</sup>

<sup>1,4,5</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Tomakaka Mamuju

<sup>2</sup>Teknik Informatika, Fakultas Komputer, Universitas Universal

<sup>3</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Tomakaka Mamuju

<sup>1</sup>hidayatfikom@gmail.com, <sup>2</sup>musliadikh@gmail.com\*, <sup>3</sup>indarkusmanto88@gmail.com, <sup>4</sup>munawirahkadir@gmail.com, <sup>5</sup>kristiansisteminformasi09@gmail.com

### Abstract

*Education is a fundamental aspect of national development that demands equal access to quality education for all. The Smart Indonesia Card (KIP) program is a government initiative aimed at supporting education for underprivileged communities. Tomakaka University, Mamuju, as one of the universities in West Sulawesi, plays an active role in distributing KIP scholarships to students who meet certain criteria. However, the selection process for prospective scholarship recipients has been carried out manually, which may lead to inefficiencies and inaccurate targeting. This study aims to apply the Naïve Bayes algorithm to classify prospective KIP scholarship recipients to make the selection process more objective, fast, and accurate. The research method uses a data mining approach with stages of data preprocessing, dividing training and test data, model training, and testing using the Python programming language on the Google Colab platform. The dataset used is 171 student data, with a division of 75% training data and 25% test data. The test results showed that the Naïve Bayes model achieved an accuracy of 95.35%, with a precision of 97%, a recall of 97%, and a loss of 4.65%, indicating excellent classification performance. Thus, this research contributes to improving administrative efficiency and targeting of KIP scholarship distribution at Tomakaka University, Mamuju.*

*Keywords: indonesia smart card, naïve bayes, data mining, student classification, KIP*

### Abstrak

Pendidikan merupakan aspek fundamental dalam pembangunan bangsa yang menuntut pemerataan akses terhadap pendidikan berkualitas bagi seluruh masyarakat. Program Kartu Indonesia Pintar (KIP) hadir sebagai inisiatif pemerintah untuk mendukung pendidikan bagi masyarakat kurang mampu. Universitas Tomakaka Mamuju sebagai salah satu perguruan tinggi di Sulawesi Barat berperan aktif dalam menyalurkan beasiswa KIP kepada mahasiswa yang memenuhi kriteria tertentu. Namun, proses seleksi calon penerima beasiswa selama ini masih dilakukan secara manual sehingga berpotensi menimbulkan ketidakefisienan dan ketidaktepatan sasaran. Penelitian ini bertujuan menerapkan algoritma Naïve Bayes dalam mengklasifikasikan mahasiswa calon penerima beasiswa KIP agar proses seleksi lebih objektif, cepat, dan akurat. Metode penelitian menggunakan pendekatan data mining dengan tahapan preprocessing data, pembagian data latih dan data uji, pelatihan model, serta pengujian menggunakan bahasa pemrograman Python pada platform *Google Colab*. Dataset yang digunakan berjumlah 171 data mahasiswa, dengan pembagian 75% data latih dan 25% data uji. Hasil pengujian menunjukkan bahwa model Naïve Bayes memperoleh akurasi sebesar 95,35%, dengan nilai presisi 97%, recall 97%, dan loss sebesar 4,65%, yang mengindikasikan performa klasifikasi yang sangat baik. Dengan demikian, penelitian ini berkontribusi dalam meningkatkan efisiensi administrasi serta ketepatan sasaran penyaluran beasiswa KIP di Universitas Tomakaka Mamuju.

Kata kunci: kartu indonesia pintar, naïve bayes, data mining, klasifikasi mahasiswa, KIP

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

### 1. Pendahuluan

Penyelenggaraan program beasiswa Kartu Indonesia Pintar (KIP) pada tingkat perguruan tinggi memegang peranan penting dalam menjamin akses pendidikan tinggi bagi mahasiswa dari keluarga kurang mampu [1]. Namun di banyak perguruan tinggi, proses seleksi calon penerima masih bergantung pada verifikasi manual dan aturan administrasi yang rentan terhadap keterlambatan, ketidakkonsistenan, dan beban kerja

biro kemahasiswaan yang tinggi [2]. Akibatnya, potensi kesalahan penentuan kelayakan dan waktu proses yang lama dapat menghambat penyaluran bantuan tepat waktu kepada mahasiswa yang benar-benar memenuhi kriteria. Selain itu, data calon penerima sering tersebar dalam berbagai format, seperti *hardcopy*, *spreadsheet* terpisah, sehingga menyulitkan analisis pola kelayakan secara sistematis [3]. Untuk Universitas Tomakaka Mamuju, fenomena tersebut menimbulkan tantangan operasional dan etis dalam

menyelenggarakan program KIP secara adil dan efisien. Kondisi tersebut menunjukkan perlunya pendekatan berbasis teknologi yang mampu membantu proses pengambilan keputusan secara objektif dan efisien dalam seleksi penerima beasiswa KIP. Salah satu pendekatan yang dapat diterapkan adalah pemanfaatan *data mining* dengan algoritma klasifikasi untuk menganalisis pola dan menentukan kelayakan mahasiswa secara otomatis berdasarkan data historis dan kriteria tertentu.

Sejumlah penelitian sebelumnya telah menunjukkan bahwa metode klasifikasi berbasis algoritma *machine learning*, khususnya *Naïve Bayes* sering digunakan dalam proses pengelompokan data kategorikal [4]. Seperti hasil penelitian sebelumnya menunjukkan bahwa metode *Naïve Bayes* mampu mengklasifikasikan penerima beasiswa dengan tingkat akurasi yang baik [5]. Penerapan algoritma ini juga telah digunakan untuk memprediksi penerima beasiswa PPA dan KIP-Kuliah di Institut Teknologi Garut [6]. Namun, sebagian besar penelitian tersebut masih terbatas pada konteks sekolah menengah atau universitas di wilayah Pulau Jawa, tanpa mempertimbangkan kondisi sosial ekonomi serta karakteristik mahasiswa di wilayah Indonesia Timur.

Studi-studi terdahulu umumnya berfokus pada peningkatan akurasi model tanpa menganalisis kontribusi masing-masing fitur terhadap hasil klasifikasi. Pada penelitian ini, evaluasi *feature importance* dilakukan dengan menganalisis pengaruh setiap atribut, seperti IPK, pendapatan orang tua, jurusan, dan jenis kelamin, terhadap probabilitas kelas pada algoritma *Naïve Bayes*. Analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang paling dominan dalam menentukan kelayakan penerimaan beasiswa, sehingga hasil prediksi tidak hanya akurat, tetapi juga dapat dijelaskan dan relevan sebagai dasar pengambilan keputusan [7]. Padahal, interpretasi terhadap fitur tersebut penting untuk mendukung pengambilan keputusan oleh pihak universitas [8]. Oleh karena itu, diperlukan penelitian yang tidak hanya membangun model klasifikasi *Naïve Bayes* untuk penerima beasiswa KIP, tetapi juga mengintegrasikan hasil analisis ke dalam kebijakan kampus. Model dengan tingkat akurasi yang kurang optimal berpotensi menimbulkan kesalahan rekomendasi, seperti mahasiswa yang layak tidak terpilih atau sebaliknya, sehingga berdampak pada ketidaktepatan sasaran dan menurunnya transparansi seleksi beasiswa. Integrasi model yang teruji ke dalam sistem pendukung keputusan diharapkan dapat meminimalkan risiko tersebut dan meningkatkan objektivitas dalam proses seleksi.

Berdasarkan tinjauan literatur lima tahun terakhir, penelitian ini mengidentifikasi keterbatasan utama pada studi sebelumnya, seperti penggunaan dataset berskala kecil, kurangnya konteks lokal, serta minimnya integrasi hasil klasifikasi ke dalam sistem pendukung keputusan. Penelitian ini mengatasi

keterbatasan tersebut dengan memanfaatkan *dataset* kampus-spesifik Universitas Tomakaka, menerapkan tahapan pra-pemrosesan dan evaluasi model yang sistematis, serta mengimplementasikan hasil pemodelan ke dalam sistem pendukung keputusan untuk mendukung proses seleksi beasiswa secara nyata dan berbasis data. Sebagian besar studi sebelumnya menggunakan *dataset* berukuran kecil dan bersifat umum sehingga membatasi kemampuan model dalam merepresentasikan kondisi riil suatu institusi. Penelitian ini menggunakan *dataset* kampus-spesifik yang berasal dari data mahasiswa Universitas Tomakaka Mamuju, mencakup karakteristik lokal seperti jurusan, capaian IPK, dan kondisi sosial ekonomi mahasiswa. Penggunaan data lokal ini memungkinkan model klasifikasi merefleksikan pola dan kebijakan internal universitas secara lebih akurat, sehingga hasil analisis menjadi lebih representatif, relevan, dan aplikatif dalam konteks seleksi penerima beasiswa KIP di Universitas Tomakaka [9]. Eksplorasi fitur berbasis konteks lokal, khususnya variabel pendapatan dan IPK, juga masih terbatas, padahal faktor tersebut berpotensi meningkatkan ketepatan model klasifikasi [10]. Di sisi lain, implementasi model berbasis kecerdasan buatan yang mampu memberikan penjelasan prediktif secara transparan (*explainable AI*) masih jarang ditemukan [11]. Pada penelitian sebelumnya, juga belum banyak menguji model secara langsung dalam lingkungan operasional biro kemahasiswaan, sehingga dampak praktis dan tingkat penerimaan pengguna dalam konteks implementasi nyata belum terukur secara optimal [12].

Oleh karena itu, penelitian ini diarahkan untuk mengisi kekosongan tersebut dengan menggunakan dataset representatif dari Universitas Tomakaka Mamuju, mengembangkan model *Naïve Bayes* dengan fitur lokal yang relevan, serta menyertakan validasi ketat dan antarmuka *decision support system* yang memberikan penjelasan probabilistik bagi pengambil keputusan. Tujuan utamanya adalah membangun dan mengevaluasi model klasifikasi *Naïve Bayes* untuk menentukan kelayakan mahasiswa calon penerima beasiswa KIP di Universitas Tomakaka Mamuju dengan dataset kampus-spesifik, serta mengembangkan prototipe sistem pendukung keputusan yang menampilkan hasil klasifikasi beserta alasan probabilistiknya. Dengan demikian, penelitian ini diharapkan dapat mempercepat proses seleksi beasiswa, meningkatkan objektivitas keputusan, serta mengurangi beban administratif biro kemahasiswaan, sekaligus memberikan kontribusi ilmiah pada bidang penerapan *machine learning* dalam sistem beasiswa perguruan tinggi di Indonesia.

## 2. Metode Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan eksperimen, karena berfokus pada penerapan dan pengujian algoritma klasifikasi dalam menentukan kelayakan calon

penerima beasiswa Kartu Indonesia Pintar (KIP) [13]. Pendekatan kuantitatif dipilih karena data yang digunakan berupa data numerik hasil pengumpulan data mahasiswa, seperti pendapatan orang tua, tanggungan keluarga, nilai akademik, dan status kepemilikan aset [14].

Metodologi penelitian meliputi beberapa tahapan utama, yakni pengumpulan data, pra-pemrosesan data, Penerapan *Naive Bayes*, Pelatihan dan Pengujian Model, evaluasi, hingga interpretasi hasil. Seluruh proses komputasi dilakukan menggunakan *Python* di lingkungan *Google Colab*, yang mendukung pemanfaatan *library machine learning* seperti *pandas*, *scikit-learn*, dan *numpy* secara optimal serta memungkinkan replikasi eksperimen berbasis *cloud computing* [15].



Gambar 1. Alur Penelitian

### 2.1. Pengumpulan Data

Tahap pengumpulan data merupakan langkah awal dalam proses penelitian yang berfungsi untuk memperoleh informasi yang relevan dan akurat mengenai calon penerima beasiswa Kartu Indonesia Pintar (KIP) [16]. Data diperoleh dari bagian kemahasiswaan Universitas Tomakaka Mamuju, yang menyimpan basis data mahasiswa penerima maupun calon penerima beasiswa KIP. Data tersebut mencakup variabel-variabel penting seperti identitas mahasiswa, program studi, indeks prestasi kumulatif (IPK), kondisi sosial ekonomi, dan tingkat pendapatan orang tua.

Selain itu, untuk memperkuat data primer, dilakukan penyebaran kuesioner secara daring kepada mahasiswa penerima dan non-penerima KIP. Validasi data dilakukan dengan mencocokkan hasil kuesioner dengan data akademik dan administrasi yang tercatat pada database resmi universitas. Untuk meminimalkan potensi bias, kuesioner disebarakan secara merata pada berbagai jurusan dan angkatan, serta menggunakan pertanyaan tertutup dengan skala yang konsisten. Data yang tidak lengkap atau tidak konsisten dieliminasi agar hasil penelitian tetap akurat dan representatif terhadap populasi mahasiswa Universitas Tomakaka.

### 2.2. Pra-pemrosesan Data

Tahap pra-pemrosesan data atau data *preprocessing* dilakukan setelah data terkumpul untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam proses klasifikasi [17]. Proses ini meliputi data *cleaning*, transformasi data, dan normalisasi data. Pada tahap data *cleaning*, dilakukan penghapusan data duplikat untuk mencegah pengulangan informasi yang dapat menyebabkan bias pada model, serta penanganan nilai kosong (*missing value*) dengan cara penghapusan atau pengisian nilai berdasarkan nilai rata-rata atribut terkait. Selain itu, data yang tidak valid atau tidak logis diperbaiki atau dihilangkan agar tidak memengaruhi proses perhitungan probabilitas.

Selanjutnya, atribut kategorikal seperti jurusan, jenis kelamin, dan status KIP ditransformasikan ke bentuk numerik menggunakan label *encoding* agar dapat diproses oleh algoritma *Naive Bayes*. Tahap normalisasi diterapkan pada atribut numerik seperti IPK dan pendapatan untuk menyamakan skala nilai antaratribut. Proses pra-pemrosesan ini berperan penting dalam meningkatkan akurasi model, karena data yang bersih dan konsisten dapat mengurangi kesalahan klasifikasi dan menghasilkan model yang lebih stabil dan reliabel.

### 2.3. Penerapan *Naive Bayes*

Tahap ini merupakan inti dari proses klasifikasi, di mana algoritma *Naive Bayes* diterapkan untuk memprediksi status penerimaan beasiswa KIP. Algoritma *Naive Bayes* dipilih karena keunggulannya dalam mengolah data dengan jumlah atribut yang relatif sedikit namun mampu menghasilkan performa klasifikasi yang tinggi [18].

Selain itu, algoritma ini memiliki kemampuan komputasi yang efisien serta bekerja dengan baik pada *dataset* berukuran kecil hingga menengah seperti kasus penelitian ini. Prinsip utama *Naive Bayes* adalah menghitung probabilitas posterior dari suatu kelas berdasarkan probabilitas kondisi dari atribut-atribut input yang diasumsikan saling independen. Melalui pendekatan probabilistik ini, model dapat menentukan kemungkinan seorang mahasiswa termasuk dalam kategori penerima atau bukan penerima KIP berdasarkan variabel-variabel seperti jurusan, jenis kelamin, IPK, dan pendapatan keluarga.

### 2.4. Pelatihan dan Pengujian Model

Setelah algoritma dipilih, dilakukan proses pelatihan dan pengujian model dengan membagi *dataset* menjadi data latih sebesar 75% dan data uji sebesar 25%. Pembagian ini dipilih sebagai kompromi antara kebutuhan model untuk mempelajari pola data secara optimal dan kebutuhan evaluasi yang objektif, mengingat ukuran *dataset* yang relatif terbatas. Proporsi 75:25 merupakan praktik umum dalam penelitian *machine learning* dan dinilai mampu memberikan keseimbangan antara stabilitas pembelajaran model dan reliabilitas pengujian, sehingga hasil evaluasi model dapat mencerminkan performa yang lebih representatif dibandingkan proporsi ekstrem seperti 90:10 atau 80:20 [19].

Data latih digunakan untuk membangun model klasifikasi dengan mengajarkan algoritma pola hubungan antar variabel terhadap status penerimaan beasiswa. Sementara itu, data uji digunakan untuk mengukur seberapa baik model dapat menggeneralisasi pola yang telah dipelajari terhadap data baru yang belum pernah dilihat sebelumnya. Proses ini dilakukan menggunakan *platform Google Colab*, yang memungkinkan pelatihan model secara efisien dengan dukungan komputasi berbasis *cloud*. Hasil dari tahap

ini berupa model *Naïve Bayes* yang terlatih dan siap dievaluasi berdasarkan performanya.

### 2.5. Evaluasi Model

Evaluasi model dilakukan dengan menerapkan model *Naïve Bayes* terlatih pada data uji untuk membandingkan hasil prediksi dengan data aktual. Hasil prediksi tersebut dianalisis menggunakan *confusion matrix* guna memperoleh jumlah *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Berdasarkan nilai-nilai tersebut, kinerja model dievaluasi menggunakan beberapa metrik, yaitu akurasi, presisi, *recall*, dan *F1-score*, sehingga penilaian kualitas model tidak hanya bergantung pada akurasi semata [20].

Akurasi dihitung untuk mengukur tingkat ketepatan keseluruhan prediksi, presisi untuk menilai ketepatan prediksi penerima beasiswa, *recall* untuk mengukur kemampuan model dalam mendeteksi seluruh mahasiswa yang benar-benar layak menerima KIP, sedangkan *F1-score* digunakan sebagai ukuran keseimbangan antara presisi dan *recall*. Metrik-metrik tersebut dihitung menggunakan persamaan berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (4)$$

### 2.6. Interpretasi Data dan Penarikan Pengetahuan

Tahap akhir penelitian ini adalah integrasi model klasifikasi *Naïve Bayes* ke dalam sistem pendukung keputusan. Model yang telah dilatih dan diuji disimpan dalam bentuk model terlatih dan diimplementasikan sebagai modul prediksi pada prototipe aplikasi [21]. Proses integrasi dilakukan dengan menghubungkan input data pengguna (jurusan, IPK, jenis kelamin, dan pendapatan orang tua) ke modul prediksi, sehingga sistem dapat secara otomatis melakukan perhitungan probabilitas dan menghasilkan rekomendasi status kelayakan penerima beasiswa KIP.

Integrasi model ke dalam sistem pendukung keputusan memungkinkan proses seleksi dilakukan secara otomatis, konsisten, dan berbasis data. Dampak dari integrasi ini adalah meningkatnya efisiensi waktu seleksi, berkurangnya subjektivitas dalam pengambilan keputusan, serta tersedianya hasil rekomendasi yang dapat ditelusuri dan dievaluasi kembali, sehingga memperkuat kontribusi praktis dan validitas hasil penelitian.

## 3. Hasil dan Pembahasan

Bagian ini menyajikan hasil analisis dan pembahasan penerapan algoritma *Naïve Bayes* dalam klasifikasi calon penerima beasiswa Kartu Indonesia Pintar (KIP) di Universitas Tomakaka Mamuju, yang mencakup

implementasi model, evaluasi performa, serta analisis variabel sosial ekonomi dan akademik yang memengaruhi akurasi prediksi penerimaan beasiswa.

Proses pengujian dan interpretasi hasil model dilakukan secara sistematis untuk memperoleh pemahaman menyeluruh terhadap pola data serta kontribusi setiap variabel dalam menentukan keputusan akhir. Selain itu, pembahasan dalam bagian ini juga diarahkan untuk mengaitkan hasil penelitian dengan teori dan temuan terdahulu, sehingga dapat memberikan gambaran komprehensif mengenai efektivitas penerapan *Naïve Bayes* sebagai pendekatan dalam sistem pendukung keputusan penerima beasiswa.

### 3.1. Pengumpulan Data

Pengumpulan data merupakan tahap awal yang sangat krusial karena kualitas model *machine learning* sangat dipengaruhi oleh kualitas data [22]. Penelitian ini menggunakan wawancara terstruktur dengan administrator kampus dan kuesioner daring sebagai instrumen utama. Wawancara dilakukan untuk memperoleh data resmi penerima beasiswa KIP serta memahami mekanisme dan kriteria penetapannya, sekaligus berfungsi sebagai data rujukan untuk memvalidasi hasil kuesioner.

Kuesioner daring menggunakan *Google Form* ditujukan kepada mahasiswa angkatan 2022/2023 karena angkatan ini telah memiliki IPK yang relatif stabil dan berada dalam sistem kurikulum yang homogen, sehingga meminimalkan bias lintas angkatan. Untuk menjamin keakuratan data, kuesioner dirancang dengan pengaturan wajib isi pada setiap variabel utama. Validasi data dilakukan melalui pemeriksaan kelengkapan data, konsistensi antarjawaban, pengecekan rentang nilai IPK sesuai standar akademik, serta identifikasi dan penghapusan data yang tidak rasional atau ekstrem. Respon yang tidak lengkap atau tidak konsisten tidak disertakan dalam analisis.

Selain variabel akademik dan ekonomi, data demografis responden yang meliputi jenis kelamin, program studi, dan kategori pendapatan orang tua turut dikumpulkan untuk mengevaluasi sebaran karakteristik responden dan memastikan keterwakilan populasi penerima beasiswa di institusi penelitian.

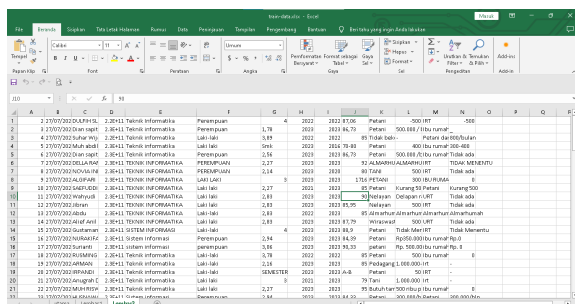
Data yang terkumpul berjumlah 171 record dan merepresentasikan seluruh populasi penerima beasiswa KIP pada periode pengamatan. Jumlah ini dinilai memadai untuk pembangunan model prediksi internal, memenuhi rasio data terhadap variabel input, serta didukung oleh performa model yang stabil pada data uji, sehingga menjamin validitas internal penelitian.

### 3.2. Pra-pemrosesan Data

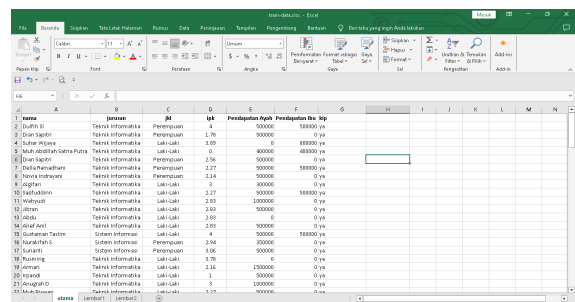
Pra-pemrosesan data merupakan tahap krusial untuk memastikan kualitas data sebelum digunakan dalam pembelajaran mesin. Pada tahap ini dilakukan

pembersihan data dengan menghapus nilai kosong, memperbaiki data tidak valid, serta menghilangkan atribut yang tidak relevan seperti NIM, timestamp, dan deskripsi pekerjaan. Proses data cleaning dilakukan menggunakan Microsoft Excel sehingga dihasilkan dataset yang bersih, konsisten, dan siap digunakan pada tahap pemodelan seperti pada Gambar 2.

Selanjutnya dilakukan transformasi data dengan mengonversi atribut kategorikal menjadi nilai numerik menggunakan teknik label encoding agar dapat diproses oleh algoritma Naïve Bayes seperti pada Gambar 3. Atribut jurusan, jenis kelamin, dan status penerima KIP masing-masing diberikan label numerik sesuai kategorinya. Proses ini bertujuan mengubah seluruh data kategorikal menjadi format numerik sehingga siap digunakan pada tahap pelatihan model.



Gambar 2. Data Awal



Gambar 3. Data Setelah Pembersihan

Tahap terakhir pada *preprocessing* adalah validasi hasil data setelah melalui proses *cleaning* dan *transformation*. Pada tahap ini, seluruh atribut telah memiliki format yang sesuai, baik numerik maupun kategorikal yang tersandi. Sehingga dataset sudah siap digunakan untuk pemodelan menggunakan algoritma *Naïve Bayes*. Hasil akhir *preprocessing* ditampilkan dalam tabel hasil data, yang menunjukkan representasi nilai numerik dari setiap atribut (data, jurusan, jenis kelamin, IPK, pendapatan, dan status KIP). *Dataset* inilah yang kemudian digunakan sebagai input pada proses pelatihan model klasifikasi.

### 3.3. Penerapan Naïve Bayes

Pada tahap ini dilakukan penerapan algoritma *Naïve Bayes* untuk membangun model klasifikasi penerima beasiswa KIP. Dataset hasil pra-pemrosesan dimuat ke dalam lingkungan komputasi berbasis awan dan selanjutnya dibagi menjadi data latih dan data uji. Pembagian data ini bertujuan untuk mengevaluasi

kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya.

Algoritma *Naïve Bayes* diterapkan dengan memanfaatkan asumsi independensi antar fitur, di mana probabilitas setiap kelas dihitung berdasarkan kombinasi probabilitas atribut akademik dan ekonomi mahasiswa. Model dilatih menggunakan data latih untuk mempelajari pola hubungan antara variabel input dan kelas target. Selanjutnya, model yang telah dilatih digunakan untuk melakukan prediksi pada data uji.

Hasil penerapan algoritma *Naïve Bayes* menunjukkan bahwa model mampu melakukan klasifikasi secara konsisten dengan tingkat performa yang stabil, yang mengindikasikan bahwa algoritma ini sesuai digunakan untuk memodelkan permasalahan prediksi penerima beasiswa KIP pada data penelitian.

### 3.4. Pelatihan Model

Pada tahap penerapan model, dilakukan pemisahan data fitur (*X*) dan variabel target (*y*) seperti pada Gambar 4, di mana variabel yang diprediksi adalah status penerima KIP. Empat fitur yang digunakan, yaitu jenis kelamin, jurusan, IPK, dan pendapatan orang tua, dipilih karena memiliki keterkaitan langsung dengan kriteria penetapan penerima beasiswa KIP. Selain relevan secara substantif, fitur-fitur tersebut juga sesuai dengan karakteristik algoritma *Naïve Bayes* yang mampu menangani kombinasi data kategorikal dan numerik melalui perhitungan probabilitas bersyarat. Asumsi independensi antar fitur pada *Naïve Bayes* tetap dapat diterapkan secara efektif karena setiap variabel merepresentasikan aspek yang berbeda, yaitu demografis, akademik, dan ekonomi. Dengan demikian, keempat fitur tersebut dinilai relevan dan memadai untuk mendukung proses klasifikasi status penerima KIP secara akurat.

```
# Cek missing values
print(data.isnull().sum())

# Set x data, y data
x = data.iloc[:, [1,2,3,4]].values
y = data.iloc[:, -1].values
```

Gambar 4. Penentuan Alokasi Data

Pada tahap selanjutnya dilakukan proses data *splitting*, yaitu pembagian *dataset* menjadi data latih dan data uji. Data latih digunakan untuk membangun dan melatih model, sedangkan data uji digunakan untuk mengevaluasi kinerja model secara objektif. Pada penelitian ini, dataset dibagi dengan proporsi 75% sebagai data latih dan 25% sebagai data uji. Dengan total 171 data, pembagian ini menghasilkan 128 data latih dan 43 data uji, sebagaimana ditunjukkan pada Tabel X. Pembagian ini bertujuan memastikan ketersediaan data yang cukup untuk proses pembelajaran model sekaligus evaluasi yang reliabel.

Tabel 1. Pembagian Data Latih dan Data Uji

Jenid Data	Presentase	Jumlah Data
Data Latih (Training)	75%	128
Data Uji (Testing)	25%	43



merekomendasikan penerima beasiswa dapat diminimalkan.

Tabel 2. Hasil Presisi

Klasifikasi	TP / TP+FP	Prediksi "Tidak"
Ya	32/33	97%
Tidak	9/10	90%

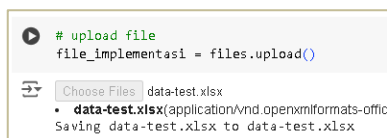
*Recall*, atau sensitivitas, menunjukkan kemampuan model dalam mendeteksi semua data yang termasuk sebagai kelas positif (penerima KIP). Hasil evaluasi menunjukkan bahwa *recall* untuk kelas "Ya" adalah 97%, dan untuk kelas "Tidak" sebesar 90%. Nilai *recall* yang tinggi mengindikasikan bahwa model mampu mendeteksi sebagian besar mahasiswa yang benar-benar memenuhi syarat sebagai penerima beasiswa KIP, dan hanya sedikit kasus positif yang terlewatkan.

Tabel 2. Hasil Recall

Klasifikasi	TP / TP+FN	Prediksi "Tidak"
Ya	32/33	97%
Tidak	9/10	90%

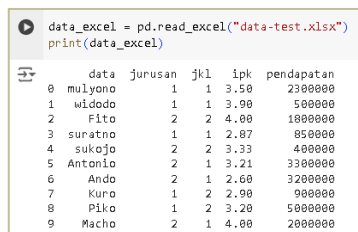
### 3.6. Implementasi Model

Implementasi model merupakan tahap penerapan model klasifikasi yang telah dibangun untuk memprediksi kemungkinan penerimaan beasiswa Kartu Indonesia Pintar (KIP) bagi mahasiswa baru. Tujuan dari tahap ini adalah untuk menguji sejauh mana model *Naive Bayes* mampu memberikan prediksi yang akurat terhadap data baru yang belum pernah digunakan dalam proses pelatihan.



Gambar 9. Upload File Implementasi

Pada tahap awal, dilakukan proses unggah file yang berisi data uji implementasi. Data ini berfungsi untuk menguji performa model terhadap kasus nyata. Dalam penelitian ini, digunakan lima data *dummy* sebagai representasi calon mahasiswa baru yang memiliki atribut sosial ekonomi dan akademik yang beragam. Proses ini digambarkan pada Gambar 9, sedangkan hasil dari data implementasi dapat dilihat pada Gambar 10.



Gambar 10. Hasil Data Implementasi

Langkah berikutnya adalah memuat model hasil pelatihan *Naive Bayes* ke dalam sistem. Proses ini dilakukan dengan mengeksekusi kode program yang memanggil model yang telah disimpan sebelumnya.

Model yang dimuat ini berfungsi untuk melakukan klasifikasi terhadap data baru berdasarkan pola dan parameter yang telah dipelajari dari dataset pelatihan.

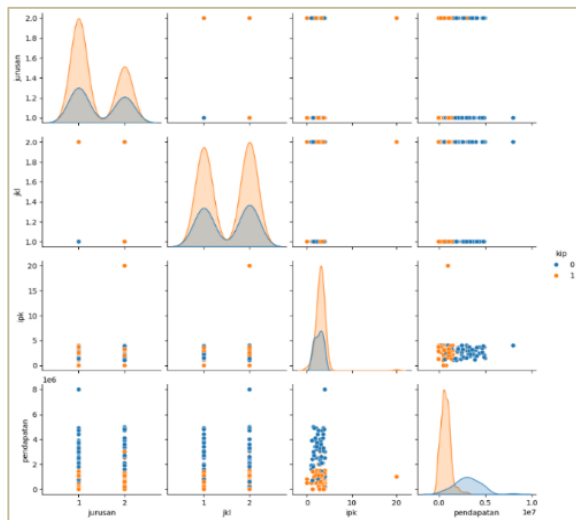
Setelah model *Naive Bayes* dilatih menggunakan data latih, dilakukan proses pengujian dengan menerapkan model pada data uji yang diperoleh dari hasil pembagian *dataset* sebelumnya. Data uji ini tidak digunakan pada tahap pelatihan model dan berfungsi untuk mengevaluasi kemampuan model dalam mengklasifikasikan data baru. Proses pengujian dilakukan dengan memasukkan data fitur mahasiswa ke dalam model terlatih untuk menghasilkan prediksi kelas, di mana nilai keluaran "1" menunjukkan potensi sebagai penerima beasiswa KIP dan nilai "0" menunjukkan tidak berpotensi. Hasil prediksi ini (Gambar 11) menunjukkan bahwa model mampu melakukan klasifikasi secara konsisten terhadap data yang belum pernah dilihat sebelumnya, sehingga layak digunakan sebagai dasar pendukung pengambilan keputusan.

	data	Prediksi Beasiswa KIP
0	mulyono	0
1	widodo	1
2	Fito	1
3	suratno	1
4	sukojo	1
5	Antonio	0
6	Ando	0
7	Kuro	1
8	Piko	0
9	Macho	0

Gambar 11. Hasil Prediksi

### 2.7. Interpretasi Data dan Penarikan Pengetahuan

Hasil interpretasi data divisualisasikan untuk memperoleh pemahaman yang lebih mendalam terhadap pola dan pengetahuan baru yang dihasilkan oleh model klasifikasi mahasiswa penerima beasiswa Kartu Indonesia Pintar (KIP) pada Universitas Tomakaka Mamuju. Visualisasi ini disajikan melalui diagram plot sebagaimana ditampilkan pada Gambar 12, yang menggambarkan hubungan antara beberapa variabel numerik dan kategorikal dengan warna berbeda untuk menunjukkan klasifikasi penerima beasiswa KIP, di mana nilai 0 merepresentasikan mahasiswa yang tidak menerima dan nilai 1 merepresentasikan mahasiswa yang menerima beasiswa. Setiap titik pada scatter plot mewakili data mahasiswa berdasarkan kombinasi variabel tertentu, sedangkan distribusi variabel numerik tampak sepanjang diagonal diagram.



Gambar 12. Diagram Plot Hasil

Berdasarkan hasil pengamatan pada diagram, variabel jurusan dan jenis kelamin menunjukkan distribusi yang relatif seimbang antara penerima dan non-penerima beasiswa, sehingga dapat disimpulkan bahwa kedua variabel tersebut tidak berpengaruh signifikan terhadap penentuan penerima KIP. Sebaliknya, variabel indeks prestasi kumulatif (IPK) memperlihatkan pola yang jelas, di mana sebagian besar mahasiswa penerima KIP cenderung memiliki IPK lebih rendah dibandingkan mahasiswa non-penerima yang tersebar pada rentang IPK lebih tinggi. Selain itu, variabel pendapatan orang tua juga menunjukkan tren yang konsisten, di mana penerima KIP umumnya berasal dari keluarga dengan tingkat pendapatan yang lebih rendah.

Berdasarkan hasil analisis visual, variabel IPK dan pendapatan orang tua menunjukkan pengaruh yang signifikan terhadap penerimaan beasiswa KIP, sedangkan variabel jurusan dan jenis kelamin tidak memiliki keterkaitan yang berarti. Temuan ini mengindikasikan bahwa mahasiswa dengan kondisi ekonomi kurang mampu dan IPK yang relatif lebih rendah cenderung memiliki peluang lebih besar untuk menerima beasiswa KIP, sejalan dengan tujuan utama program dalam mendukung pemerataan akses pendidikan bagi mahasiswa dari keluarga berpendapatan rendah.

#### 4. Kesimpulan

Berdasarkan hasil penelitian mengenai prediksi mahasiswa penerima beasiswa KIP di Universitas Tomakaka Mamuju menggunakan algoritma *Naïve Bayes*, dapat disimpulkan bahwa model yang dibangun melalui proses pelatihan di *Google Colab* menunjukkan performa sangat baik dengan tingkat akurasi mencapai 95,35%, serta mampu mengklasifikasikan data uji secara tepat. Temuan penelitian juga menunjukkan bahwa penerima beasiswa KIP umumnya berasal dari keluarga berpendapatan rendah dan memiliki IPK relatif lebih rendah. Untuk pengembangan lebih lanjut, disarankan agar penelitian berikutnya menggunakan data latih

yang lebih besar dan variatif, serta melakukan model tuning dan penambahan variabel yang lebih komprehensif agar dapat meningkatkan kinerja model dan menghasilkan prediksi yang lebih akurat.

#### Daftar Rujukan

- [1] N. Sufni, "Analisis Keberhasilan Program Kartu Indonesia Pintar (KIP) dalam Meningkatkan Akses Pendidikan di Indonesia," *Benefit J. Bussiness, Econ. Financ.*, vol. 2, no. 2, pp. 38–45, 2024, doi: 10.70437/benefit.v2i2.393.
- [2] A. S. Gutama, M. Fedryansyah, and E. Nuriyah, "Implementasi Program Kartu Indonesia Pintar (KIP) Berdasarkan Basis Nilai Keadilan Dalam Kebijakan Sosial," *J. Penelit. dan Pengabd. Kpd. Masv.*, vol. 2, no. 3, p. 389, 2022, doi: 10.24198/jppm.v2i3.35872.
- [3] S. Rosyadi, N. Rino A, Y. Muhammad, and Y. Tri, *Administrasi Publik di Era Disrupsi dan Big Data*, no. December. 2021.
- [4] O. B. Akanbi, "Application of Naive Bayes to Students' Performance Classification," *Asian J. Probab. Stat.*, vol. 25, no. 1, pp. 35–47, 2023, doi: 10.9734/ajpas/2023/v25i1536.
- [5] A. R. Siregar and M. Furqan, "Classification of Scholarships for Students in Schools Using the Naïve Bayes Method," vol. 7, no. 1, pp. 278–289, 2025.
- [6] A. Mulyani, D. Kurniadi, M. R. Nashrulloh, I. T. Julianto, and M. Regita, "The Prediction of PPA and KIP-Kuliah Scholarship Recipients Using Naive Bayes Algorithm," *J. Tek. Inform.*, vol. 3, no. 4, pp. 821–827, 2022, doi: 10.20884/1.jutif.2022.3.4.297.
- [7] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, no. December 2023, p. 102342, 2024, doi: 10.1016/j.is.2023.102342.
- [8] 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 Teknol.*, vol. 12, no. 2, pp. 309–316, 2023, doi: 10.23887/jstundiksha.v12i2.51745.
- [9] M. Arsyali, Nasruddin, and Syamsuriadi, "Sistem Akses Pintu Otomatis Berbasis Pengenalan Wajah Menggunakan Sensor Kamera Arduino," *Sinov. - Sci. Lit. Innov. Technol. J.*, vol. 01, no. 01, pp. 21–27, 2024.
- [10] Z. Rohman, A. Homaidi, and Z. Fatah, "Penerapan Metode Naïve Bayes Untuk Menentukan Penerima Kartu Indonesia Pintar (KIP)," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1806–1815, 2024, doi: 10.33379/gtech.v8i3.4573.
- [11] C. Fitri, N. Halizah, and M. Kartikasari, "Implementasi Algoritma Naïve Bayes Dalam Penentuan Pemberian Beasiswa Kip Kuliah (Studi Kasus Stiki Malang) Implementation of Naïve Bayes Algorithm in Determining Kip-Kuliah Scholarship (Case Study Stiki Malang)," vol. 12, no. 2, pp. 1–6, 2024.
- [12] M. Penerima and B. Kip, "Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan," vol. 15, no. 1, pp. 108–114, 2025.
- [13] K. Kaharuddin, K. Musliadi, H. Kurniawan, and S. Ilwan, "Penerapan Augmented Reality Dalam Pengenalan Peralatan Manufaktur Pada Prodi Teknik Industri Universitas Universal," *J. Tekinkom (Teknik Inf. dan Komputer)*, vol. 7, no. 1, pp. 402–409, 2024.
- [14] A. Ismail, M. Kh, and A. Hasnining, "Optimalisasi Pembelajaran Melalui Aplikasi Interaktif di Pondok Pesantren XYZ Polewali Mandar," *J. Digit. Ecosyst. Nat. Sustain.*, vol. 4, no. 1, pp. 14–20, 2024.
- [15] F. Masruroh, R. Wahyuningsih, and A. N. Halima, "Pengaruh Parenting Pada Orang Tua Terhadap Pola Asuh Anak Usia Dini," *INCARE, Int. J. Educ. Resour.*, vol. 5, no. 1, pp. 080–088, 2024, doi: 10.59689/incare.v5i1.907.
- [16] H. Jurnal, M. Andriana, Y. Sinta, and W. Ulfa, "Sistem Informasi Penggajian Karyawan Berbasis Web," *J. Tek. Inform. Dan Teknol. Inf.*, vol. 2, no. Agustus, pp. 2827–9379,

- 
- 2022.
- [17] J. Amalia, J. Pakpahan, M. Pakpahan, and Y. Panjaitan, "Model Klasifikasi Berita Palsu Menggunakan Bidirectional LSTM Dan Word2Vec Sebagai Vektorisasi," *J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 4, pp. 3319–3331, 2022.
- [18] A. Giachanou, B. Ghanem, E. A. Rissola, P. Rosso, F. Crestani, and D. Oberski, "The Impact Of Psycholinguistic Patterns In Discriminating Between Fake News Spreaders And Fact Checkers," *Data Knowl. Eng.*, vol. 138, no. December 2021, p. 101960, 2022, doi: 10.1016/j.datak.2021.101960.
- [19] V. R. Dănăilă, S. Avram, and C. Buiu, "The Applications Of *Machine learning* In HIV Neutralizing Antibodies Research - A Systematic Review," *Artif. Intell. Med.*, vol. 134, no. October 2021, p. 102429, 2022, doi: 10.1016/j.artmed.2022.102429.
- [20] M. Liu, J. Guo, and D. Bi, "Comparison Of Administrative And Regulatory Green Technologies Development Between China And The U . S . Based On Patent Analysis," *Data Sci. Manag.*, vol. 6, no. 1, pp. 34–45, 2023, doi: 10.1016/j.dsm.2023.01.001.
- [21] M. C. Neff *et al.*, "Development and expert inspections of the user interface for a primary care decision support system," *Int. J. Med. Inform.*, vol. 192, no. July, 2024, doi: 10.1016/j.ijmedinf.2024.105651.
- [22] Maspiati, Nasruddin, and K. Musliadi, "Sistem Informasi Pemesanan Baju Olahraga Berbasis Web Pada Konveksi ' Adher ' Tammerodo Sendana Majene," *Sinov. - Sci. Lit. Innov. Technol. J.*, vol. 01, no. 01, pp. 7–13, 2024.