BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

Implementasi Naïve Bayes untuk Klasifikasi Peminatan Program Studi pada Penerimaan Mahasiswa Baru di Fakultas Ilmu Komputer Unika

Munawirah[1,*], Andriansyah Oktafiandi Arisha[2]

[1] Fakultas Ilmu Komputer, Sistem Informasi, Universitas Tomakaka, Mamuju, Indonesia
[2] Teknik Informatika, AMIK Tri Dharma Palu, Palu, Indonesia
Email: [1,*]munawirahkadir@gmail.com, [2]ecpand@gmail.com
(* : coressponding author; munawirahkadir@gmail.com)

Abstrak- Fakultas Ilmu Komputer Universitas Tomakaka di Mamuju, Sulawesi Barat, memiliki dua program studi, yaitu Sistem Informasi dan Teknik Informatika. Namun, dalam praktiknya, calon mahasiswa baru sering mengalami kebingungan dalam menentukan jurusan yang sesuai dengan kemampuan dan latar belakang akademiknya. Pemilihan program studi umumnya didasarkan pada tren jurusan favorit, dorongan eksternal, atau preferensi sosial tanpa mempertimbangkan jurusan asal di sekolah sebelumnya. Kondisi tersebut berpotensi menimbulkan ketidaksesuaian minat yang berdampak pada risiko penurunan motivasi belajar, pindah jurusan, berhenti kuliah, atau mengalami hambatan selama masa studi. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi program studi menggunakan metode klasifikasi Naïve Bayes guna memprediksi kecenderungan peminatan program studi berdasarkan atribut input seperti jenis kelamin, asal sekolah, dan jurusan asal sekolah. Dataset yang digunakan merupakan data historis penerimaan mahasiswa baru Fakultas Ilmu Komputer Universitas Tomakaka sejak tahun akademik 2015/2016 hingga 2024/2025, sebanyak 1.046 entri data. Proses analisis mencakup tahapan data *mining*, mulai dari seleksi dan pembersihan data, pembagian data latih dan data uji (80:20), hingga evaluasi performa menggunakan metode Confusion Matrix. Hasil evaluasi menunjukkan akurasi sebesar 87,14%, presisi 89,91%, *recall* 87,70%, dan F1-*score* 88,76%. Model ini diimplementasikan ke dalam aplikasi berbasis website menggunakan *framework* Flask, guna mempermudah pemberian rekomendasi jurusan secara *real-time*. Pendekatan ini memberikan kontribusi sistem rekomendasi berbasis data yang membantu institusi dalam memetakan minat mahasiswa, menyusun strategi promosi yang tepat sasaran, serta memberikan intervensi awal terhadap pilihan program studi mahasiswa baru yang kurang sesuai.

Kata Kunci: Naïve Bayes; Penambangan Data; Klasifikasi; Mahasiswa Baru; Peminatan Jurusan.

Abstract- *The Faculty of Computer Science at Universitas Tomakaka offers two study programs, namely Information Systems and Informatics Engineering. In practice, however, prospective students often encounter difficulties in selecting a program that aligns with their academic background and capabilities. Program selection is frequently influenced by prevailing trends, external encouragement, or social preferences, without due consideration of their previous fields of study. Such conditions may result in a mismatch between student interests and program requirements, potentially leading to decreased motivation, program transfer, academic delays, or even withdrawal from studies. This study aims to develop a study program recommendation system employing the Naïve Bayes classification method to predict students' program preferences based on input attributes such as gender, type of previous school, and prior field of study. The dataset utilized comprises historical admission data from the Faculty of Computer Science at Universitas Tomakaka, covering the academic years 2015/2016 to 2024/2025, with a total of 1,046 entries. The analytical process follows the stages of data mining, including data selection and cleaning, partitioning into training and testing datasets (80:20), and performance evaluation using the Confusion Matrix method. The evaluation results demonstrate an accuracy of 87.14%, a precision of 89.91%, a recall of 87.70%, and an F1-score of 88.76%, indicating robust model performance. The developed model has been implemented into a web-based application using the Flask framework, enabling the delivery of real-time program recommendations. This approach contributes to the development of a data-driven recommendation system that assists the institution in mapping student interests, formulating targeted promotional strategies, and providing early interventions to address potential mismatches in program selection among prospective students.*

Keywords: Naïve Bayes; Data Mining; Classification; New Students; Program Preferences.

1. PENDAHULUAN

Pemilihan program studi merupakan salah satu keputusan krusial dalam proses penerimaan mahasiswa baru di perguruan tinggi. Keputusan yang tidak tepat dalam pemilihan jurusan berpotensi menyebabkan berbagai permasalahan selama masa studi, seperti ketidakcocokan minat, kemampuan, lingkungan belajar, serta rendahnya motivasi belajar, hingga risiko berhenti kuliah atau pindah jurusan [1]. Di Fakultas Ilmu Komputer Universitas Tomakaka yang terletak di ibu kota Provinsi Sulawesi Barat, Mamuju, memiliki dua program studi yang ditawarkan, yaitu Sistem Informasi dan Teknik Informatika, menjadi pilihan utama yang seringkali membingungkan calon mahasiswa. Permasalahan utama yang dihadapi adalah banyaknya calon mahasiswa baru yang memilih program studi bukan berdasarkan kesesuaian dengan latar belakang pendidikan ataupun potensi akademik, melainkan karena pengaruh eksternal seperti tren jurusan favorit, faktor keluarga, teman sejawat, kepribadian calon mahasiswa, sekolah asal, citra kampus dan faktor sosial lainnya [2]. Hal ini menunjukkan bahwa proses pemilihan jurusan belum sepenuhnya berbasis data atau pendekatan yang sistematis. Ketidaktepatan dalam pemilihan jurusan akan berdampak pada keberlangsungan studi dan efektivitas proses pembelajaran. Sebagai alternatif solusi, beberapa pendekatan dapat digunakan, seperti tes minat bakat, wawancara akademik, hingga penilaian berbasis prestasi. Namun, metode-metode tersebut cenderung memerlukan sumber daya tambahan dan tidak mudah diotomatisasi [3]. Oleh karena itu, dibutuhkan pendekatan berbasis data (*data-driven*) yang

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

mampu mengolah informasi historis dan karakteristik mahasiswa secara efisien untuk menghasilkan rekomendasi peminatan yang lebih akurat dan objektif [4]. Salah satu metode yang relevan untuk mengatasi permasalahan tersebut adalah algoritma Naïve Bayes**,** sebuah teknik klasifikasi berbasis probabilistik yang dikenal efisien dalam menangani data kategorikal serta mampu memberikan hasil yang cukup akurat meskipun dengan data pelatihan yang terbatas [5]. Algoritma ini cocok digunakan dalam konteks peminatan program studi karena dapat memprediksi kecenderungan pilihan jurusan berdasarkan pola historis.

Beberapa penelitian sebelumnya telah menerapkan algoritma klasifikasi dalam pemetaan peminatan program studi mahasiswa menggunakan algoritma Decision Tree untuk klasifikasi minat jurusan siswa SMA [6], [7], [8], sementara dalam studi lain [9] memanfaatkan algoritma K-Nearest Neighbor (KNN) berbasis nilai akademik, dan studi [10] menggunakan algoritma C4.5 untuk klasifikasi minat studi. Namun, sebagian besar penelitian tersebut belum secara eksplisit mempertimbangkan variabel latar belakang pendidikan, seperti jenis sekolah dan jurusan asal sekolah, serta belum diimplementasikan dalam bentuk sistem klasifikasi yang dapat digunakan secara *real-time*. Penelitian ini berbeda dengan penelitian sebelumnya karena memanfaatkan algoritma Naïve Bayes untuk membangun sistem klasifikasi peminatan program studi secara real-time dengan mempertimbangkan atribut jenis kelamin, asal sekolah, dan jurusan asal sekolah sebagai variabel input utama. Algoritma Naïve Bayes merupakan teknik klasifikasi berbasis probabilistik yang efisien dalam menangani data kategorikal dan mampu memberikan hasil prediksi yang akurat meskipun dengan data pelatihan terbatas. Dengan menggunakan dataset penerimaan mahasiswa baru Fakultas Ilmu Komputer Universitas Tomakaka pada periode 2015/2016 hingga 2024/2025, penelitian ini tidak hanya memberikan kontribusi teoretis dalam pengembangan sistem rekomendasi peminatan berbasis data, tetapi juga kontribusi praktis melalui implementasi model klasifikasi ke dalam aplikasi web berbasis Flask. Sistem ini diharapkan dapat membantu institusi dalam memetakan minat mahasiswa, menyusun strategi promosi jurusan secara tepat sasaran, serta memberikan rekomendasi peminatan program studi secara real-time untuk mendukung pengambilan keputusan akademik secara tepat guna dan optimal.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis algoritma Naïve Bayes. Proses penelitian mengikuti tahapan Knowledge Discovery in Database (KDD), yang meliputi pengumpulan data historis, pra-pemrosesan, perancangan dan pelatihan model klasifikasi, evaluasi performa model, serta implementasi dalam bentuk aplikasi berbasis web [11]. Model dibangun dan dilatih menggunakan bahasa pemrograman Python 3.10 dengan bantuan pustaka *pandas*, *scikit-learn*, dan *Flask* sebagai *framework* implementasi web-nya.

Model klasifikasi ini dirancang agar dapat secara otomatis memprediksi jurusan pilihan calon mahasiswa berdasarkan variabel input: jenis kelamin, asal sekolah, dan jurusan asal sekolah. Diagram alur dan pseudocode digunakan untuk menggambarkan logika kerja algoritma.

2.2 Tahapan Penelitian

Data yang digunakan berjumlah 1.046 dataset yang bersumber dari Sekretariat Penerimaan Mahasiswa Baru (PMB) Universitas Tomakaka Mamuju dari tahun akademik 2015/2016 hingga 2024/2025. Proses pengolahan data mengikuti tahapan-tahapan KDD berikut:

a. Pemilihan Data *(Data Selection)*
   Sebelum tahap pencarian informasi pada proses KDD dimulai, perlu dilakukan pemilihan data (seleksi) dari kumpulan data operasional yang tersedia. Data yang telah dipilih tersebut kemudian disimpan pada berkas tersendiri, terpisah dari operasional database, untuk digunakan dalam proses data mining [12].
b. Pra-pemrosesan *(Pre-processing/Cleaning)*
   Dilakukan proses data *cleaning* yaitu proses pembersihan data yang akan dianalisis. mengisi data yang missing values di mana data yang diperoleh masih mengandung data yang bernilai null dan data duplikasi, tidak konsisten dan memperbaiki kesalahan pada data. Oleh karena itu, sebelum menganalisis data, harus dipastikan data yang dianalisis adalah data bersih, lengkap (tidak mengandung nilai null), dan tidak memiliki data duplikasi karena dapat memberi hasil yang tidak akurat. Pada proses ini dilakukan juga proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD [13].
c. Split Dataset (Data *Training* dan *Testing)*
   Naïve Bayes akan melakukan klasifikasi tetapi data bersih yang siap diolah tadi dibagi terlebih dahulu menjadi dua set data, yaitu data *training* dan data *testing*. Data pelatihan (*training*) digunakan untuk membangun atau melatih model *classifier*. Sedangkan adalah data uji (*testing*) digunakan untuk menguji *classifier* yang telah dibangun untuk melihat seberapa akurat hasil klasifikasinya [14].

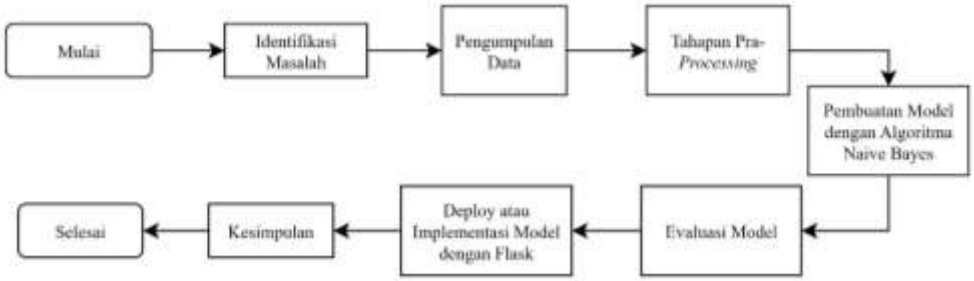
BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

d. Merancang dan Melatih *Classifier (Training and Testing)*
Tahap ketiga ialah Merancang dan Melatih *Classifier* dengan menerapkan Teorema Naïve Bayes.

e. Evaluasi *Classifier (Evaluation)*
Pada tahap kelima dilakukan evaluasi terhadap klasifikasi model, yaitu proses penilaian terhadap pola atau informasi yang dihasilkan dari proses data mining untuk memastikan apakah pola tersebut sesuai dengan kenyataan atau justru bertentangan dengan hipotesis awal, serta mengukur tingkat akurasi dari model tersebut [15]. Evaluasi dilakukan menggunakan metode Confusion Matrix untuk menghasilkan nilai akurasi, presisi, recall, dan F1-score.

f. *Deployment*
Tahap terakhir, model dilatih menggunakan pustaka scikit-learn pada Python. Variabel input dikodekan menggunakan *OneHotEncoder* agar dapat dibaca sebagai fitur kategorikal. Model yang terbentuk kemudian disimpan dalam pipeline klasifikasi untuk diintegrasikan ke dalam aplikasi Flask [16].

Alur dari Tahapan penelitian untuk melakukan data mining berikut:



Gambar 1. Tahapan Penelitian

2.3 Algoritma Naïve Bayes

Tahapan penelitian yang dilakukan meliputi tahapan data mining dalam menerapkan metode klasifikasi Naive Bayes. Istilah "data mining" ini digunakan cukup luas di industri Teknologi Informasi (TI). Ini sering diterapkan pada berbagai kegiatan pemrosesan data skala besar seperti pengumpulan, ekstraksi, pergudangan, dan analisis data. Naïve Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen terhadap fitur lainnya [17]. Algoritma ini terdiri dari beberapa tahapan sebagai berikut:

a. Menentukan Kelas
Identifikasi terlebih dahulu kelas (label) yang akan diprediksi. Misalnya: Sistem Informasi dan Teknik Informatika.

b. Menghitung Probabilitas Kelas *(Prior Probability)*
Rumusnya adalah $P(C_k) = N_k / N$
Keterangan:
- $P(C_k)$: probabilitas kelas ke-k (prior)
- N_k: jumlah data pada kelas ke-k
- N: total jumlah data

c. Menghitung Probabilitas Atribut Terhadap Kelas (*Likelihood*)
Untuk setiap fitur x_i, hitung menggunakan rumus $P(x_i \mid C_k) = N_{(xi,\ C_k)} / N_i C_k)$ [18]
Keterangan:
- $P(x_i \mid C_k)$: probabilitas nilai fitur x_i muncul dalam kelas C_k
- $N_{(xi,\ C_k)}$: jumlah data dengan nilai x_i dalam kelas C_k
- $N_i C_k)$: total data dalam kelas C_k

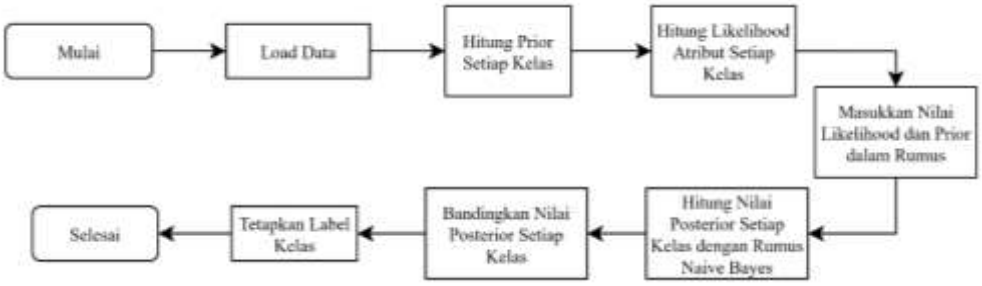
d. Menghitung Probabilitas Posterior (Menggunakan Teorema Naïve Bayes)
Tahapan ini, prior dan likelihood dikombinasikan untuk menghitung peluang akhir (posterior) dari setiap kelas, rumusnya $P(C_k \mid x) \propto P(C_k) \times \prod P(x_i \mid C_k)$ [18]
Keterangan:
- $x = (x_1, x_2, ..., x_n)$: fitur-fitur input
- $P(C_k \mid x)$: probabilitas data x termasuk ke dalam kelas C_k

e. Menentukan Kelas
Pilih kelas dengan nilai probabilitas posterior terbesar sebagai hasil klasifikasi dengan menggunakan rumus $\hat{y} = argmax\_C_k\ P(C_k \mid x)$ [18]

Proses klasifikasi algoritma Naïve Bayes digambarkan dalam bentuk flowchart diagram. Flowchart merupakan cara untuk menggambarkan proses kerja algoritma atau langkah-langkah pemecahan suatu masalah yang direpresentasikan

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

menggunakan simbol-simbol tertentu. Tahapan dari penggunaan algoritma Naïve Bayes ditunjukkan pada Flowchart Gambar 2 di bawah ini:



Gambar 2. Tahapan Algoritma Naïve Bayes

2.4 Pengujian dengan *Confusion Matrix*

Pada tahap ini pengujian model penelitian dilakukan dengan metode Confusion Matrix yang mempresentasikan hasil evaluasi model dengan menggunakan tabel matrik . Jika dataset terdiri dari 2 kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Evaluasi menggunakan confusion matrix menghasilkan nilai Akurasi, Precision, Recall, serta F1-Score. Akurasi dalam klasifikasi menunjukkan persentase data yang berhasil diklasifikasikan dengan tepat setelah pengujian hasil klasifikasi dilakukan. Presisi menggambarkan proporsi prediksi positif yang sesuai dengan kondisi positif pada data sebenarnya. Sementara itu, recall menunjukkan proporsi data positif yang dapat diprediksi secara benar positif.

True Positive (TP) merupakan jumlah record postif dalam dataset yang diklasifikasikan positif. True Negative (TN) merupakan jumlah record negative dalam dataset yang diklasifikasikan positif. False Positive (FP) merupakan jumlah record negatif dalam dataset yang diklasifikasikan positif. False Negative (FN) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif [19]. Berikut adalah persamaan model Confusion Matrix:

a. *Accuracy* dalam evaluasi metrik digunakan untuk mengukur tingkat akurasi pada sebuah algoritma dengan membagi keseluruhan data yang benar dengan jumlah keseluruhan data [19]. Rumus mencari akurasi pada model yang dibuat adalah sebagai berikut:
   Accuracy = (TP+TN)/ (TP + FP + TN + FN)
b. *Precision* dalam evaluasi metrik digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif [19]. Rumus mencari presisi pada model yang dibuat adalah sebagai berikut:
   Precision = TP/ (TP + FP)
c. *Recall/Sensitivity* dalam evaluasi metrik digunakan untuk menunjukkan presentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif [19]. Rumus mencari sensitivitas pada model yang dibuat adalah sebagai berikut:
   Recall/Sensitivity = TP/ (TP + FN)
d. *F1-Score* dalam evaluasi metrik digunakan untuk mengukur keseimbangan antara Precision dan Recall**,** dan menilai seberapa baik model klasifikasi menangani data yang tidak seimbang atau kesalahan FP/FN [19]**.** Rumus mencari skor F1 adalah sebagai berikut:
   F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

3. HASIL DAN PEMBAHASAN

3.1 Seleksi Data dan Variabel *(Data and Variable Selection)*

Dalam penelitian ini 1.046 data Penerimaan Mahasiswa Baru di Fakultas Ilmu Komputer Tahun Akademik 2015/2016 sampai 2024/2025 tidak semuanya digunakan. Untuk itu, perlu dilakukan seleksi data. Seleksi data yaitu mengambil data-data yang diperlukan saja dalam penelitian, dalam kasus ini berarti data dari atribut/variabel yang diperlukan dalam penelitian. Dari 1.046 data tersebut dilakukan proses seleksi data di mana dari 20 atribut yang ada, hanya menyisakan 5 atribut yang berpengaruh dalam klasifikasi dan prediksi peminatan jurusan. 5 atribut itu ialah: Nama Lengkap, Jenis Kelamin, Asal Sekolah, Jurusan Asal Sekolah dan Jurusan Yang Dipilih di mana Label Kelas yang ditetapkan adalah atribut Jurusan Yang Dipilih.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

Tabel 1. Kumpulan Data Mentah

No.	No. Tes	Nama Lengkap	Tempat Lahir	Tgl Lahir	Jenis Kelamin	Tahun Lulus	Asal Sekolah	…	Program Studi
1.	0524080	M Yunus	Kombiling	11/04/2005	Laki-Laki	2024	SMK Bina Karya	…	Teknik Informatika
2.	0524085	Syahrul	Saludengen	01/02/2003	Laki-Laki	2024	SMAN 1 Tommo	…	Sistem Informasi
3.	0524101	Fiska Fatana	Karama	27/04/2006	Perempuan	2024	SMKS Juan Pattaropura	…	Sistem Informasi
4.	0524073	Musdalifah	Mamuju	04/06/2006	Perempuan	2024	SMAN 2 Mamuju	…	Sistem Informasi
5.	0524075	Muhammad Syariie F	Mamuju	04/03/2006	Laki-Laki	2024	SMAN 1 Mamuju	…	Sistem Informasi
6.	0524067	Widia Andini	Bulu Bonggu	30/09/2005	Perempuan	2024	SMAN 1 Dapurang	…	Sistem Informasi
7.	0524094	Jefri	Dusun Sidal	02/01/2004	Laki-Laki	2023	SMKN 1 Dapurang	…	Teknik Informatika
8.	0524097	Ismail	Bambaloka	06/08/2003	Laki-Laki	2022	SMKS Bina Karya	…	Teknik Informatika
…	…	…	…	…	…	…	…	…	…
1046.	0524096	Saldi Apandi	Pulau Salissingan	25/07/2002	Laki-Laki	2020	SMKN 2 Majene	…	Teknik Informatika

Tabel 2. Hasil Seleksi Variabel

Jenis Kelamin	Asal Sekolah	Jurusan Asal Sekolah
Perempuan	SMK	TEKNIK
Laki-laki	SMA	IPA
	MA	IPS
	Lainnya	Lainnya

3.2 Pra-pemrosesan (*Pra-processing/ Cleaning*)

Pada dataset Penerimaan Mahasiswa Baru di Fakultas Ilmu Komputer terdapat beberapa missing values, pembersihan data dilakukan secara manual dengan mencari data yang nilainya sama (data duplikasi). Hasil dari mencari duplikasi data ada sebanyak 0 data, sehingga data yang digunakan masih berjumlah 1.046 data. Dengan demikian, data yang diolah tetap sebanyak 1.046 *record* data. Setelah bersih dari data duplikasi, selanjutnya adalah melakukan modifikasi untuk data yang kosong. Pemodifikasian dilakukan dengan memodifikasi/mengisi data yang kosong dengan data dominan atau terbanyak.

Tabel 3. Hasil Pembersihan Data

No	Nama Lengkap	Jenis Kelamin	Asal Sekolah	Jurusan Asal Sekolah	Jurusan Dipilih
1	A. Erwinda Maharani	Perempuan	MA	IPA	Teknik Informatika
2	Ahlas	Laki-Laki	SMK	Lainnya	Sistem Informasi
3	Ahmad Rayyan RW	Laki-Laki	SMA	IPS	Sistem Informasi
4	Akmal Hidayat	Laki-Laki	SMK	Teknik	Teknik Informatika
.....	.....	.....	.....	.....	.....
1.046	Yusni Paramita	Perempuan	SMK	Teknik	Sistem Informasi

Pada Record data Asal Sekolah data dimodifikasi ke dalam 3 kategori sekolah: SMK (Sekolah Menengah Kejuruan), SMA (Sekolah Menengah Atas) dan MA (Madrasah Aliyah), dan record data Jurusan Asal Sekolah dimodifikasi dan dibagi menjadi 4 kategori dengan Jurusan Asal Sekolah terbanyak yaitu: Teknik, IPS (Ilmu Pengetahuan Sosial), IPA (Ilmu Pengetahuan Alam), dan kategori Lainnya dimana kategori Lainnya adalah atribut minoritas yang memiliki record

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

Jurusan Asal Sekolah yang jauh lebih sedikit dibandingkan kategori yang paling sering muncul yaitu: Teknik, IPS, dan IPA. Contoh Jurusan Asal Sekolah kategori Lainnya adalah jurusan Tata Boga, jurusan Bahasa, dan lain-lain sebagainya yang sifatnya jarang. Atribut Jurusan Yang Dipilih adalah Label Kelas yang ditetapkan untuk klasifikasi peminatan program studi ini.

3.3 Merancang dan Melatih *Classifier (Training and Testing)*

Data yang telah melalui tahapan pra-pemrosesan yang siap diolah tadi dibagi terlebih dahulu dibagi menjadi dua set data, yaitu data training dan data testing. Pembagian data dilakukan dengan menggunakan prinsip Pareto yaitu prinsip 80% : 20%. Sebanyak 80% dari data yang disiapkan bertindak sebagai data latih dan 20% sisanya bertindak sebagai data uji. Pada kasus ini, jumlah data yang akan diolah sebanyak 1.046 data, sehingga 80% atau sejumlah 836 *record* data bertindak sebagai data latih, dan 20% sisanya yaitu sebanyak 210 *record* data bertindak sebagai data uji.

Gambar 3. Distribusi Data *Training* dan Data *Testing*

a. Menghitung probabilitas kelas *(Prior Probability)*

Pertama menentukan probabilitas kelas dengan menghitung probabilitas masing-masing kelas yang ada. Pada studi kasus ini terdapat 2 kelas, yaitu kelas Sistem Informasi dan kelas Teknik Informatika terhadap atribut label kelas, Jurusan Yang Dipilih. Rumus dan hasil perhitungannya berikut:

Tabel 4. Hasil Perhitungan Probabilitas Kelas

No	Deskripsi	Jumlah	Perhitungan	Σ
1	P(Jurusan yang Dipilih= Sistem Informasi)	433	433/836	0.51703
2	P(Jurusan yang Dipilih= Teknik Informatika)	403	403/836	0.48297
	Total	836		

Pada Tabel 4 memperlihatkan hasil perhitungan probabilitas kelas Sistem Informasi dan Teknik Informatika dengan membagi jumlah keseluruhan kelas Sistem Informasi dan Teknik Informatika kemudian dibagi dengan total keseluruhan data training 80% tadi atau sebanyak 836 data.

b. Menghitung Probabilitas Atribut Terhadap Kelas *(Likelihood)*

Kedua menentukan probabilitas atribut terhadap masing-masing kelas, dalam studi kasus ini, akan dilakukan perhitungan probabilitas setiap atribut/kategori terhadap kelas Sistem Informasi dan kelas Teknik Informatika.

1. Menghitung Probabilitas Atribut Terhadap kelas Sistem Informasi

Pertama menentukan probabilitas masing-masing atribut terhadap kelas Sistem Informasi. Hasil perhitungannya dapat terlihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Probabilitas Atribut Terhadap Kelas Sistem Informasi

No	Description	Jumlah	Perhitungan	Σ
1	P(JenisKelamin= Laki-laki|SI)	168	168/433	0.38799
2	P(JenisKelamin= Perempuan|SI)	265	265/433	0.61201
3	P(AsalSekolah= SMK|SI)	149	149/433	0.34411
4	P(AsalSekolah= SMA|SI)	220	220/433	0.50808
5	P(AsalSekolah= MA|SI)	58	58/433	0.13395
6	P(AsalSekolah= Lainnya|SI)	6	6/433	0.01386

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

7	P(JurusanAsalSekolah= Teknik|SI)	26	26/433	0.06005
8	P(JurusanAsalSekolah= IPA|SI)	105	105/433	0.24249
9	P(JurusanAsalSekolah= IPS|SI)	179	179/433	0.41339
10	P(JurusanAsalSekolah= Lainnya|SI)	123	123/433	0.28406

Tabel 5 adalah hasil perhitungan probabilitas *likelihood P(xi| SI)* untuk masing-masing nilai atribut terhadap kelas Sistem Informasi dengan membagi jumlah keseluruhan setiap atribut dari total 433 data kelas SI (Sistem Informasi).

2. Menghitung Probabilitas Atribut Terhadap kelas Teknik Informatika

   Kedua menentukan probabilitas masing-masing atribut terhadap kelas Teknik Informatika. Hasil perhitungannya dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Probabilitas Atribut Terhadap Kelas Teknik Informatika.

No	Description	Jumlah	Perhitungan	Σ
1	P(JenisKelamin= Laki-laki|TI)	305	305/403	0.75682
2	P(JenisKelamin= Perempuan|TI)	98	98/403	0.24318
3	P(AsalSekolah= SMK|TI)	191	191/403	0.47395
4	P(AsalSekolah= SMA|TI)	132	132/403	0.32754
5	P(AsalSekolah= MA|TI)	66	66/403	0.16377
6	P(AsalSekolah= Lainnya|TI)	14	14/403	0.03474
7	P(JurusanAsalSekolah= Teknik|TI)	191	191/403	0.47395
8	P(JurusanAsalSekolah= IPA|TI)	178	178/403	0.44169
9	P(JurusanAsalSekolah= IPS|TI)	13	13/403	0.03226
10	P(JurusanAsalSekolah= Lainnya|TI)	21	21/403	0.05211

Tabel 6 adalah hasil perhitungan probabilitas *likelihood P(xi / TI)* untuk masing-masing nilai atribut terhadap kelas Teknik Informatika dengan membagi jumlah keseluruhan setiap atribut dari total 403 data kelas TI (Teknik Informatika).

c. Menghitung Probabilitas Posterior dan Klasifikasi Data

Tahapan ini menentukan probabilitas posterior untuk menetapkan label kelas, tahapan ini adalah langkah utama dalam proses klasifikasi data menggunakan Naive Bayes. Di bawah ini dilakukan pengujian satu data untuk melihat hasil klasifikasi peminatan Program Studi setelah perhitungan model diperoleh. Data uji yang akan dihitung dapat dilihat pada Tabel 7 di bawah ini:

Tabel 7. Data Uji

Nama Lengkap	Jenis Kelamin	Asal Sekolah	Jurusan Asal Sekolah	Jurusan Yang Dipilih
Muna	Perempuan	SMA	IPA	?

1. Pertama dilakukan perhitungan terhadap kelas Sistem Informasi terlebih dahulu menggunakan rumus menghitung probabilitas posterior. Pehitungan dituliskan sebagai betikut:
   = P(Jurusan yang Dipilih= Sistem Informasi) * P(JenisKelamin= Perempuan|SI) * P(AsalSekolah= SMA|SI) * P(JurusanAsalSekolah= IPA|SI)
   = 0,51703 * 0,61201 * 0.50808 * 0.24249
   = 0,03898523
2. Selanjutnya dilakukan perhitungan terhadap kelas Teknik Informatika untuk dibandingkan dengan hasil perhitungan terhadap kelas Sistem Informasi. Pehitungan dituliskan sebagai betikut:
   = P(Jurusan yang Dipilih= Teknik Informatika) * P(JenisKelamin= Perempuan|TI) * P(AsalSekolah= SMA|TI) * P(JurusanAsalSekolah= IPA|TI)
   = 0,48297 * 0,24318 * 0.32754 * 0.44169
   = 0,01699142
3. Selanjutnya dilakukan perbandingan posterior antara hasil perhitungan terhadap kelas Sistem Informasi dengan hasil perhitungan terhadap kelas Teknik Informatika. Pada hasil hitung manual data uji di atas, hasil perhitungan terhadap kelas Sistem Informasi lebih tinggi dibandingkan dengan hasil perhitungan terhadap kelas Teknik Informatika, dengan nilai 0,03898523 lebih besar dari 0,01699142.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

4. Menetapkan hasil klasifikasi data. Maka Mahasiswa atas nama Muna tersebut diklasifikasikan ke dalam program studi Sistem Informasi sesuai dari rumus menentukan kelas, yaitu mencari *argmax* atau nilai tertinggi dari hasil perhitungan terhadap kedua kelas.

3.4 Hasil Analisis

a. Analisis Probabilitas Posterior Berdasarkan Variabel Input

Tabel 8. Analisis Probabilitas Posterior Berdasarkan Variabel Jenis Kelamin

Jenis Kelamin	P(SI | x)	P(TI | x)	Kecenderungan
Perempuan	0.61	0.24	Cenderung SI
Laki-laki	0.39	0.76	Cenderung TI

Tabel 9. Analisis Probabilitas Posterior Berdasarkan Variabel Asal Sekolah

Asal Sekolah	P(SI | x)	P(TI | x)	Kecenderungan
SMK	0.34	0.47	Cenderung TI
SMA	0.51	0.33	Cenderung SI
MA	0.13	0.16	Cenderung TI
Lainnya	0.01	0.03	Cenderung TI

Tabel 10. Analisis Probabilitas Posterior Berdasarkan Variabel Jurusan Asal Sekolah

Jurusan Asal Sekolah	P(SI | x)	P(TI | x)	Kecenderungan
Teknik	0.06	0.47	Cenderung TI
IPA	0.24	0.44	Cenderung TI
IPS	0.41	0.03	Cenderung SI
Lainnya	0.28	0.05	Cenderung SI

b. Hasil Analisis

Hasil perkiraan probabilitas posterior menunjukkan bahwa pelajar perempuan cenderung memilih program studi Sistem Informasi, sedangkan pelajar laki-laki cenderung memilih Teknik Informatika. Hal ini dapat disebabkan oleh persepsi bahwa Sistem Informasi memiliki tingkat kesulitan teknis yang lebih rendah dan orientasi kerja yang lebih fleksibel, sehingga lebih diminati oleh pelajar perempuan. Sebaliknya, Teknik Informatika dengan karakteristik materi yang lebih teknis dan berbasis pemrograman mendalam lebih sesuai dengan minat dan keterampilan teknis yang umumnya dimiliki oleh siswa laki-laki. Dari sisi variabel asal sekolah, lulusan SMK cenderung memilih Teknik Informatika karena sebagian besar mereka memiliki latar belakang pendidikan teknik komputer jaringan atau rekayasa perangkat lunak, sehingga memiliki kesinambungan jalur keilmuan. Sedangkan lulusan SMA, khususnya dari jurusan IPS, cenderung memilih Sistem Informasi karena materi yang diberikan lebih mendukung pengolahan data dan sistem informasi manajemen yang bersifat multidisiplin. Analisis ini dapat menjadi acuan bagi institusi dalam merancang strategi promosi program studi yang lebih tepat sasaran. Misalnya, promosi Teknik Informatika dapat difokuskan pada kejuruan sekolah-sekolah dengan jurusan teknik, sedangkan promosi Sistem Informasi dapat difokuskan pada SMA dan Madrasah Aliyah dengan jurusan IPS atau IPA.

Tabel 11. Hasil Analisis Kecenderungan Pemilihan Jurusan berdasarkan Variabel

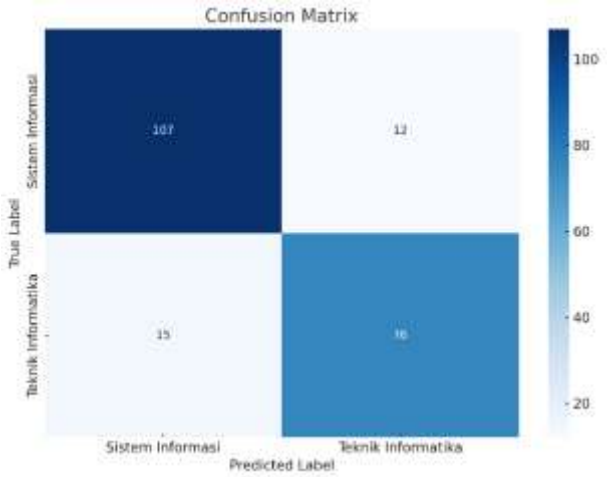
Variabel	Pengetahuan yang Diperoleh	Kontribusi bagi Pengambilan kebijakan
Jenis Kelamin	Mahasiswa perempuan cenderung memilih Sistem Informasi; laki-laki condong ke Teknik Informatika	Dapat menjadi dasar pendekatan promosi atau intervensi personalisasi
Asal Sekolah	Lulusan SMK lebih banyak memilih Teknik Informatika	Dapat menjadi dasar untuk desain kurikulum adaptif
Jurusan Asal Sekolah	Siswa dari jurusan Teknik sangat dominan memilih Teknik Informatika	Menggambarkan kesinambungan jalur keilmuan

3.5 Evaluasi *Classifier (Evaluation)*

Setelah semua proses melatih dan merancang model telah dilakukan, proses terakhir adalah melakukan pengujian terhadap sistem yang telah dibuat. Dalam data mining, tahap pengujian bertujuan untuk menghasilkan kinerja model yang

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

baik ketika diterapkan pada data riil. Confusion Matrix merupakan metode yang paling sering digunakan dalam penilaian klasifikasi model untuk mengetahui tingkat akurasi dari model data mining yang dikembangkan. Confusion matrix bekerja dengan membandingkan hasil klasifikasi dari *classifier* dengan label data yang sesungguhnya dan melakukan perhitungan yang dapat menghasilkan metrik keluaran seperti *recall, precision, f1-score* dan *accuracy* [20].



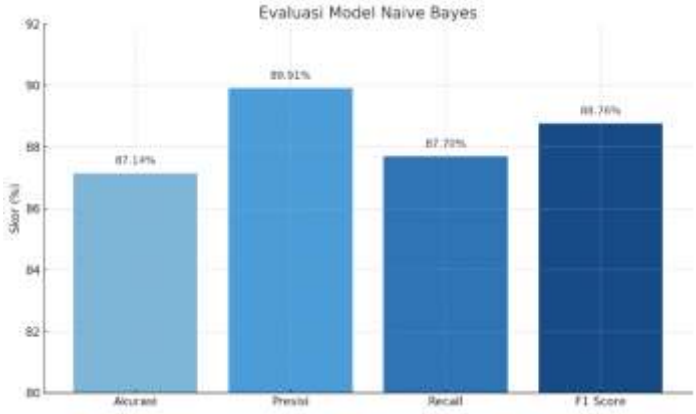
Gambar 4. Hasil Evaluasi *Classifier* menggunakan Confusion Matrix

Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu True Positif, True Negatif, False Positif, dan False Negatif. Label kelas Sistem Informasi direpresentasikan sebagai Positive dan label kelas Teknik Informatika direpresentasikan sebagai Negatif. Pada Gambar 4 di atas menampilkan hasil pengujian dari pengolahan data *testing* di mana n = 210 atau 20% data uji dari 1.046 data set yang ada. Berdasarkan hasil evaluasi performa model menggunakan *Confusion Matrix* diketahui nilai True Positive, True Negative, False Positive, dan False Negative. Uraian penjelasan gambar di atas sebagai berikut:

a. *True Positive* (TP) menginterpretasikan bahwa model memprediksi positif dan faktanya itu benar. Dalam studi kasus ini, model memprediksi mahasiswa memilih jurusan Sistem Informasi dan memang benar mahasiswa tersebut memilih jurusan Sistem Informasi. Dari hasil evaluasi performa model, didapati 107 data yang dikategorikan sebagai *True Positive*.
b. *True Negative* (TN) menginterpretasikan bahwa model memprediksi negatif dan faktanya itu benar. Model memprediksi mahasiswa memilih jurusan Teknik Informatika dan memang benar mahasiswa tersebut memilih jurusan Teknik Informatika. Dari hasil evaluasi model didapati 76 data.
c. *False Positive* (FP) menginterpretasikan bahwa model memprediksi positif dan faktanya itu salah. Model memprediksi mahasiswa memilih jurusan Sistem Informasi dan ternyata prediksinya salah, faktanya mahasiswa memilih jurusan Teknik informatika. Dari hasil evaluasi model didapati sebanyak 12 data. Dalam evaluasi model klasifikasi, FP sering disebut sebagai *Type* I *Error* atau adalah kesalahan tipe I.
d. *False Negative* (FN) menginterpretasikan bahwa model memprediksi negatif dan faktanya itu salah. Model memprediksi mahasiswa memilih jurusan Teknik Informatika dan ternyata prediksi salah, faktanya mahasiswa tersebut memilih jurusan Sistem Informasi. Dari hasil evaluasi didapati 15 data. Dalam evaluasi model klasifikasi atau prediksi FN merupakan kesalahan tipe 2 (*Type* II *Error*). Dalam kasus-kasus tertentu kesalahan tipe 2 ini sangat berbahaya. *Error* pada sebuah model akan sangat mempengaruhi pengambilan keputusan terhadap sebuah kasus. Dalam studi kasus ini karena mahasiswa diprediksi memilih jurusan Teknik Informatika tetapi faktanya mahasiswa tersebut memilih jurusan Sistem Informasi, maka hal ini akan mempengaruhi model dalam menentukan mahasiswa mana yang cocok di jurusan Teknik Informatika dan Sistem Informasi, hal ini akan mempengaruhi masa studi mahasiswa tersebut berdasarkan atribut yang telah dijadikan parameter untuk mempertimbangkan cocok tidaknya seorang mahasiswa berkuliah di program studi tersebut, di mana pembelajaran pada Program Studi Teknik Informatika lebih sedikit sulit dibandingkan jurusan Sistem Informasi. Hal ini akan mempengaruhi masa studi 4 tahun mahasiswa jika mengalami ketidakcocokan pemilihan program studi.

Skor hasil evaluasi performa model menggunakan metode *Confusion Matrix*, nilai *accuracy, precision, f1-score*, dan *recall/sensitivity* dapat dilihat sebagai berikut:

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT



Gambar 5. Skor Evaluasi Performa Model Naïve Bayes

Berdasarkan hasil evaluasi performa model klasifikasi dalam studi kasus peminatan program studi di Fakultas Ilmu Komputer, Universitas Tomakaka, diperoleh nilai akurasi sebesar 87,14%. Nilai ini menunjukkan bahwa model memiliki tingkat ketepatan yang cukup baik dalam mengklasifikasikan pilihan jurusan mahasiswa secara keseluruhan. Selanjutnya, nilai presisi yang tinggi mencapai 89,91%, menunjukkan bahwa model memiliki tingkat ketepatan yang baik dalam mengklasifikasikan calon mahasiswa ke dalam jurusan yang benar. Presisi yang tinggi sangat penting apabila hasil prediksi akan digunakan untuk pengambilan keputusan strategis, seperti perencanaan kuota penerimaan per jurusan, distribusi sumber daya, atau penyusunan materi promosi jurusan yang lebih tepat sasaran. Dari sisi recall atau sensitivitas, model memperoleh nilai 87,70%, yang berarti dari seluruh mahasiswa yang benar-benar memilih jurusan tertentu, sekitar 88% berhasil dikenali dengan benar oleh model. Nilai ini krusial apabila intitusi ingin melakukan intervensi awal, misalnya dalam bentuk bimbingan atau pendekatan personal terhadap calon mahasiswa yang kecenderungannya kuat terhadap jurusan tertentu. Nilai F1-Score sebesar 88,76% mencerminkan keseimbangan antara presisi dan recall. Dalam konteks peminatan program studi, F1-Score menjadi metrik yang relevan karena model tidak hanya dituntut untuk akurat dalam memprediksi, tetapi juga adil dalam menjangkau seluruh kategori jurusan tanpa bias terhadap kelas tertentu.

Jika dibandingkan dengan studi [7] dan [8] yang menggunakan metode C.45 Decision Tree, model mereka hanya mencapai akurasi 12% hingga 84,27% saja, sementara model ini lebih tinggi sekitar 3%. Hal ini menunjukkan bahwa Naïve Bayes lebih cocok dalam konteks data kategorikal dengan variabel terbatas.

3.6 Implementasi Model ke dalam Program Aplikasi

Model klasifikasi Naïve Bayes yang telah dibangun kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Flask [21]. Proses klasifikasi dilakukan secara *real-time* melalui antarmuka yang dirancang agar ramah pengguna (*user-friendly*).

Gambar 6. Form Tampilan Website

Aplikasi ini menerima input berupa atribut kategorikal: Jenis Kelamin**,** Asal Sekolah, dan Jurusan Asal Sekolah, yang kemudian diproses untuk menghasilkan prediksi program studi secara langsung dalam bentuk teks dan visual

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

sederhana guna meningkatkan keterbacaan dan pemahaman hasil prediksi bagi pengguna, dapat dilihat pada Gambar 6. Implementasi sistem ini menggunakan pendekatan Naïve Bayes multinomial, dengan transformasi data menggunakan OneHotEncoder untuk menangani atribut kategorikal. Dataset yang digunakan bersumber dari data historis mahasiswa baru Fakultas Ilmu Komputer Universitas Tomakaka periode 2015–2024. Model dilatih dalam *pipeline* klasifikasi menggunakan Python versi 3.10, dan dikembangkan dalam lingkungan Flask karena fleksibilitasnya dalam integrasi serta kemudahan dalam *deployment* model *machine learning* ke aplikasi web [16].

Aplikasi ini memungkinkan pengguna untuk memasukkan data pribadi calon mahasiswa dan memperoleh hasil prediksi jurusan secara langsung seperti Gambar 7. Hasil prediksi ditampilkan secara otomatis melalui antarmuka web, menjadikan aplikasi ini sebagai alat bantu yang efektif dalam menganalisis kecenderungan peminatan calon mahasiswa. Aplikasi ini tidak hanya mendukung pemetaan minat sejak awal penerimaan mahasiswa baru, tetapi juga dapat digunakan dalam proses bimbingan akademik, penyusunan strategi promosi jurusan yang lebih terarah, serta optimalisasi alokasi sumber daya secara efisien berdasarkan tren pilihan program studi.

Gambar 7. Tampilan Website Setelah Melakukan Prediksi Jurusan

Untuk mendukung replikasi dan pengembangan lanjutan, proses klasifikasi disertai dengan *pseudocode* Python dari algoritma Naïve Bayes yang menggambarkan perhitungan probabilitas posterior dan penentuan kelas dappat dilihat pada Gambar 8 di bawah. Hal ini memberikan transparansi model dan dapat dijadikan referensi bagi peneliti lain maupun untuk keperluan audit akademik.

```
def naive_bayes_predict(x_input, model_params):
    hasil = {}
    for class_label in model_params['classes']:
        prior = model_params['prior'][class_label]
        likelihood = 1
        for feature in x_input:
            likelihood *=
model_params['likelihood'][class_label].get(feature, 1e-6)
        hasil[class_label] = prior * likelihood
    return max(hasil, key=hasil.get)
```

Gambar 8. Pseudocode Naïve Bayes dalam Python

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Naïve Bayes untuk mengklasifikasikan peminatan program studi mahasiswa baru di Fakultas Ilmu Komputer Universitas Tomakaka. Model yang dibangun menggunakan atribut jenis kelamin, asal sekolah, dan jurusan asal sekolah, dengan hasil evaluasi menunjukkan akurasi 87,14%, presisi 89,91%, recall 87,70%, dan F1-score 88,76%. Nilai-nilai tersebut mengindikasikan performa model yang baik dan seimbang, membuktikan bahwa Naïve Bayes efektif diterapkan dalam sistem rekomendasi berbasis data akademik di bidang pendidikan. Implementasi model ke dalam aplikasi web berbasis Flask memungkinkan pemberian rekomendasi peminatan secara real-time, sehingga bermanfaat dalam mendukung pengambilan keputusan akademik dan strategi

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 3, September 2025, Hal. 218-229
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.2142
https://journal.fkpt.org/index.php/BIT

promosi jurusan. Ke depan, penelitian dapat dikembangkan dengan memperluas atribut input, misalnya nilai akademik, hasil asesmen minat bakat, atau riwayat prestasi siswa untuk meningkatkan akurasi prediksi, serta menggunakan pendekatan klasifikasi lain seperti Random Forest atau SVM untuk perbandingan performa.

REFERENCES

[1] A. Harningsih, "Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Mahasiswa Memilih Program Studi di Perguruan Tinggi dalam Perspektif Ekonomi Islam (Studi Terhadap Mahasiswa Angkatan 2017, Fakultas Ekonomi dan Bisnis Islam Universitas Islam Negeri Raden Intan Lampung)," Lampung, 2019.

[2] M. Saputro, "Analisis Faktor-Faktor yang Mempengaruhi Keputusan Mahasiswa dalam Memilih Program Studi," *Jurnal Pendidikan Informatika dan Sains*, vol. 6, no. 1, pp. 83–94, Jun. 2017.

[3] W. Seneru *et al.*, *Eksplorasi dalam Penilaian Belajar*. Yayasan Cendikia Mulia Mandiri, 2024.

[4] F. Mahardika, F. Supriadi, and A. Guntara, "Analisis Distribusi Minat Mahasiswa pada Konsentrasi Informatika Menggunakan Pendekatan Data-Driven Decision Making," *Inti Nusa Mandiri*, vol. 19, no. 2, pp. 278–287, Feb. 2025, doi: 10.33480/inti.v19i2.6347.

[5] A. R. Cahyono, A. Iskandar, and A. Rifqi, "Analisis Data Inventaris pada PT. Global Samudera Kreasi untuk Optimalisasi Pengelolaan dan Prediksi Kualitas Barang Menggunakan Algoritma Support Vector Machine Dan Naïve Bayes," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 3, p. 3863, Jun. 2025.

[6] B. Q. Husaini and J. Jemakmun, "Penerapan Algoritma Decision Tree C45 untuk Klasifikasi Penjurusan Siswa," *Jurnal Teknlogi Informatika dan Komputer MH. Thamrin* , vol. 9, no. 1, pp. 455–470, Mar. 2023, doi: 10.37012/jtik.v9i1.1512.

[7] A. Ulfa, D. Winarso, and E. Arribei, "Sistem Rekomendasi Jurusan Kuliah bagi Calon Mahasiswa Baru Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Muhammadiyah Riau)," *Jurnal Fasilkom*, vol. 10, no. 1, pp. 61–65, Apr. 2020.

[8] S. Sugiyanti and F. Muhammad Fauzi, "Analisis Data Mining dalam Prediksi Pemilihan Jurusan pada SMA Manggala Menggunakan Algoritma C 4.5," *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, vol. 8, no. 2, pp. 304–311, Jun. 2025, doi: 10.36595/misi.v5i2.

[9] P. M. Gunawan, "Sistem Pendukung Keputusan Pemilihan Bidang Minat Mahasiswa Menggunakan Algoritma KNN Berbasis Web," Ponorogo, 2024.

[10] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 11, no. 2, pp. 131–145, Oct. 2021, doi: 10.34010/jamika.v11i2.5506.

[11] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *Implementasi Data Mining (Clastering, Association, Prediction, Estimation, Classfication)*, Pertama. Indramayu: Penerbit Adab, 2023.

[12] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naive Bayes*, Pertama. Bandung: Kreatif Industri Nusantara, 2020.

[13] N. Purwati, H. Kurniawan, and S. Karmila, *Data Mining Volume 1*, Pertama., vol. 1. Banyumas: Zahira Media Publisher, 2021.

[14] M. Ardiana, P. N. Rahayu, T. Brian, I. I. Widyastuti, and S. Wibowo, "Binary Number Classification Using Naïve Bayes for Digit Recognition," *Jurnal Sistem Telekomunikasi, Elektronika, Sistem Kontrol, Power System & Komputer*, vol. 5, no. 1, pp. 25–34, Jan. 2025, doi: 10.32503/jtecs.v5i1.6740.

[15] Y. Nurdin, K. Saddami, and N. Nasaruddin, *Pengenalan Praktis Supervised Machine Learning: Dengan Jupyter Notebook*, Pertama. Banda Aceh: USK Press, 2025.

[16] H. Hasanka, "Deploy Scikit-Learn Machine Learning Models with Flask," Medium. Accessed: Jul. 10, 2025. [Online]. Available: https://hiranh.medium.com/deploy-scikit-learn-machine-learning-models-with-flask-b6d6413b019a

[17] D. Berarr, "Bayes' Theorem and Naive Bayes Classifier," *Encyclopedia of Bioinformatics and Computational Biology (Second Edition)*, vol. 1, pp. 483–494, Mar. 2025, doi: https://doi.org/10.1016/B978-0-323-95502-7.00118-4.

[18] K. Gohari *et al.*, "A Bayesian Latent Class Extension of Naive Bayesian Classifier and Its Application to the Classification of Gastric Cancer Patients," *BMC Med Res Methodol*, vol. 23, no. 1, Aug. 2023, doi: 10.1186/s12874-023-02013-4.

[19] A. Tharwat, "Classification Assessment Methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.

[20] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika: Jurnal Sistem Komputer*, vol. 11, no. 1, pp. 59–66, Apr. 2022, doi: 10.34010/komputika.v11i1.4350.

[21] A. C. Darmawan, "Pengembangan Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5," Yogyakarta, 2023.

[22] Andani, S. R., & Karim, A. (2025). Enhancing Lung Cancer Detection: Optimizing CNN Architectures through Hyperparameter Tuning. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 9(4), 944-954.