

---

## **Sistem Rekomendasi Jurusan Menggunakan Algoritma *Naïve Bayes Gaussian* Berbasis Web**

Ken Bagus Panuluh Yudha Perkasa<sup>1</sup>, Febry Eka Purwiantono<sup>2\*</sup>

<sup>1</sup>STIKI Malang, Teknik Informatika, Jl. Raya Tidar No.100, Karangbesuki, Kec. Sukun, Kota Malang, Jawa Timur 65146, Indonesia

<sup>2</sup>STIKI Malang, Manajemen Informatika, Jl. Raya Tidar No.100, Karangbesuki, Kec. Sukun, Kota Malang, Jawa Timur 65146, Indonesia

---

### **Informasi Artikel**

Diterima: 05-12-2023

Direvisi: 15-12-2023

Diterbitkan: 25-12-2023

### **Kata Kunci**

*Sistem Rekomendasi; Algoritma Naive Bayes Gaussian; Web; Jurusan, Pendidikan Tinggi*

### **\*Email Korespondensi:**

*febry@stiki.ac.id*

### **Abstrak**

Penelitian ini bertujuan mengembangkan sebuah sistem rekomendasi jurusan berbasis web menggunakan algoritma Naive Bayes Gaussian guna mengatasi masalah kebingungan siswa dalam memilih jurusan di STIKI Malang. Terdapat keterbatasan informasi dan bimbingan karir yang membuat siswa lulusan SMA/K sulit menentukan pilihan jurusan yang sesuai dengan minat dan potensinya. Dalam penelitian ini, data latih dari 107 mahasiswa aktif dan lulusan STIKI Malang digunakan untuk memberikan rekomendasi berdasarkan berbagai atribut seperti jenis kelamin, jurusan, skill, hobi, alasan kuliah, alasan memilih program studi, ketertarikan terhadap matematika, dan ketertarikan terhadap bahasa Inggris. Metode Naive Bayes Gaussian berhasil mengklasifikasikan data dengan akurasi mencapai 87,85%, yang dapat mengatasi masalah ketidakpastian dalam pemilihan jurusan. Diharapkan sistem ini dapat membantu siswa lulusan SMA/K dalam memilih jurusan yang sesuai, mengurangi kesalahan pemilihan jurusan, dan memanfaatkan potensinya secara optimal.

### **Abstract**

*This study aims to develop a web-based department recommendation system using the Gaussian Naive Bayes algorithm to address the issue of student confusion in selecting majors at STIKI Malang. Limited career guidance and information pose challenges for high school graduates in making informed decisions about their suitable majors based on interests and potentials. In this research, training data from 107 active students and graduates are utilized to provide recommendations based on various attributes such as gender, current major, skills, hobbies, reasons for pursuing higher education, program selection motives, interest in mathematics, and interest in English. The Gaussian Naive Bayes method successfully classifies continuous data with an accuracy of 87,85%, effectively dealing with the uncertainty in major selection. It is hoped that this system will assist high school graduates in choosing appropriate majors, reducing major selection errors, and optimizing potential.*

---

## 1. Pendahuluan

Melanjutkan pendidikan merupakan impian banyak pelajar di Indonesia. Setelah menyelesaikan program di SMA/K, siswa akan memutuskan untuk memilih universitas dan jurusan yang sesuai dengan aspirasi masing-masing. Namun, masih banyak siswa yang bingung dalam mengambil keputusan untuk memilih jurusan yang tepat. Banyak lulusan SMA/K berencana untuk melanjutkan ke universitas, tetapi lulusan tersebut tidak tahu jurusan apa yang akan dipilih, sehingga cenderung memilih jurusan sendiri tanpa bantuan dalam mengambil keputusan (Herdiansah 2020).

Lulusan SMA/K tidak semua mendapatkan bimbingan karir dari guru. Hal inilah yang membuat lulusan SMA/K bingung dan sulit mengambil keputusan karena informasi yang tersedia sangat sedikit. Akibatnya, banyak lulusan SMA/K yang salah memilih jurusan kuliah atau hanya memilih jurusan atas saran teman dan kerabat, bahkan mengikuti trend yang tidak sesuai dengan minat, bakat dan bakatnya (Herdiansah 2020).

STIKI Malang adalah perguruan tinggi yang bergerak di bidang Informatika, pada perguruan tinggi ini terdapat 4 program studi, yaitu program studi Teknik Informatika, Sistem Informasi, Manajemen Informatika, dan Desain Komunikasi Visual. Siswa/i yang telah lulus dari SMA/K dan ingin melanjutkan pendidikan di perguruan tinggi STIKI Malang akan memutuskan untuk memilih program studi yang ada di STIKI Malang. Sering kali terjadi tamu yang datang ke ruangan Penerimaan Mahasiswa Baru STIKI Malang/calon mahasiswa baru di STIKI Malang masih bingung untuk menentukan jurusan.

Untuk membuat sistem pemilihan jurusan, dibutuhkan sebuah metode atau algoritma. Seperti yang dilakukan oleh (Aditya and Purwiantono 2020) menggunakan metode SPK (Sistem Penunjang Keputusan) yaitu *AHP (Fuzzy-Analytical Hierarchy Process)* untuk pemilihan jurusan pada perguruan tinggi negeri, namun pada metode ini tidak terdapat nilai akurasi, oleh sebab itu peneliti mengusulkan algoritma Naïve Bayes Gaussian pada bidang ilmu Data Mining Klasifikasi seperti penelitian yang dilakukan oleh (Prioyo, Sari and Mardiana 2023) dan (Ramadhani, Effendy and Darmawan 2022) yang dapat memunculkan dan menghasilkan nilai akurasi di atas 70%.

Naïve Bayes Gaussian merupakan algoritma berbasis nilai kontinu dengan konsep probabilitas yang dapat digunakan untuk menentukan jenis dokumen dan dapat mengolah data dalam jumlah besar dengan akurasi tinggi. Tingkat kinerja pasti yang dihasilkan oleh metode Naive Bayes Gaussian bergantung pada data pelatihan yang dipilih, akurasi akan dimaksimalkan jika data pelatihan mencakup sebagian besar atau bahkan semua data yang disimpan (Mujahidin, Prasetyo, and Utomo 2022). Sebenarnya ada algoritma Naïve Bayes jenis lain yaitu Naïve Bayes Multinomial Text. Namun algoritma Naïve Bayes Multinomial Text hanya cocok pada data yang bersifat diskrit sedangkan Naïve Bayes Gaussian sangat cocok pada data continue seperti yang ada dalam penelitian ini. Kelas dokumen tidak hanya ditentukan oleh kata yang muncul saja tapi juga ditentukan oleh jumlah kemunculannya (Ashari et al. 2020). Pada penelitian ini penulis menggunakan 8 parameter yang digunakan dalam metode Naïve Bayes Gaussian ini, yaitu jenis kelamin, jurusan, skill, hobi, alasan kuliah, alasan ambil program studi, ketertarikan terhadap matematika, dan ketertarikan terhadap bahasa inggris.

Hal yang telah dipaparkan tersebutlah yang menjadikan latar belakang penulis untuk melakukan penelitian ini, dengan adanya sistem rekomendasi jurusan akan sangat membantu Siswa/i lulusan SMA/K dalam memutuskan untuk memilih jurusan yang sesuai dengan minat, bakat, dan kemampuan yang dimiliki. Sistem ini diharapkan dapat memberikan rekomendasi atau saran bagi calon mahasiswa untuk memilih jurusan atau program studi yang akan dipilih.

## 2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah sebagai berikut:

### 2.1 Pengumpulan Data

Teknik yang digunakan untuk mengumpulkan data dan informasi pada penelitian ini adalah *survey* kepada mahasiswa STIKI Malang dari Angkatan 2018 – 2022 untuk digunakan sebagai *sample data training* pada aplikasi.

### 2.2 Analisa Data

Data yang diperoleh adalah data jenis kelamin, jurusan di SMA/K, *skill*, hobi, alasan berkuliah, alasan mengambil prodi, minat matematika, minat bahasa inggris, dan juga asal program studi dari mahasiswa STIKI Malang Angkatan 2018 – 2022. Data tersebut akan digunakan dalam perhitungan prediksi untuk rekomendasi jurusan menggunakan Naïve Bayes Gaussian. Metode Naïve Bayes Gaussian merupakan algoritma berbasis nilai kontinu dengan konsep probabilitas yang dapat digunakan untuk menentukan jenis dokumen dan dapat mengolah data dalam jumlah besar dengan akurasi tinggi. Tingkat kinerja pasti yang dihasilkan oleh metode Naive Bayes Gaussian bergantung pada data pelatihan yang dipilih, akurasi akan dimaksimalkan jika data pelatihan mencakup sebagian besar atau bahkan semua data yang disimpan. (Mujahidin, Prasetyo, and Utomo 2022). Adapun tahapan perhitungan algoritma *Naïve Bayes Gaussian*: (1) Mencari Probabilitas pada tiap atribut/variabel. (2) Menghitung bobot dari setiap kelas/*class*. (3) Menghitung akurasi dari setiap data untuk mengklasifikasikan data dalam set pengujian dan menghitung seberapa sering prediksi yang cocok dengan label asli. (4) Membuat *confussion matrix*. (5) Menghitung nilai akurasi. (6) Uji coba data baru.

### 2.3 Algoritma Naïve Bayes Gaussian

Untuk melakukan perhitungan manual, ada 6 tahap yang harus dilakukan yaitu:

#### 1. Mencari probabilitas tiap atribut

Pada tahap ini, kita perlu menghitung probabilitas tiap atribut dalam dataset berdasarkan distribusi Gaussian. Dalam konteks Naive Bayes, kita menganggap semua fitur saling independen (bersifat "naif"), jadi kita bisa menghitung probabilitas tiap atribut secara terpisah.

$$P(x_i|c) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right) \quad (1)$$

Keterangan:

$P(x_i|c)$  adalah probabilitas bahwa atribut  $x_i$  memiliki nilai tertentu dalam kelas  $c$ .

$\mu_{c,i}$  adalah rata-rata atribut  $x_i$  dalam kelas  $c$ .

$\sigma_{c,i}$  adalah deviasi standar atribut  $x_i$  dalam kelas  $c$ .

#### 2. Menghitung bobot dari setiap class

Bobot dalam konteks ini merujuk pada probabilitas a priori dari setiap kelas dalam dataset. Ini adalah probabilitas murni dari sebuah instance yang dipilih secara acak menjadi dari kelas tertentu tanpa mempertimbangkan fitur-fitur yang ada.

$$P(c) = \frac{\text{Jumlah instance data dalam kelas } c}{\text{Total jumlah instance data dalam semua kelas}} \quad (2)$$

### 3. Menghitung akurasi setiap data

Di tahap ini, kita akan menerapkan model yang telah dilatih untuk mengklasifikasikan data dalam set pengujian dan menghitung seberapa sering prediksi cocok dengan label asli. Hasil ini akan digunakan untuk menghitung akurasi.

### 4. Membuat *Confusion Matrix*

*Confusion Matrix* adalah tabel yang menjelaskan kinerja algoritma klasifikasi. Ini akan memberikan gambaran detail tentang seberapa baik dan buruk model dalam mengklasifikasikan data, termasuk *True Positives*, *False Positives*, *True Negatives*, dan *False Negatives*.

### 5. Menghitung nilai akurasi

Akurasi adalah ukuran umum dari kinerja model dan dihitung dari *Confusion Matrix*. Akurasi adalah rasio dari jumlah prediksi yang benar (*True Positives* dan *True Negatives*) dibagi dengan jumlah total prediksi.

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total jumlah data pengujian}} \times 100\% \quad (3)$$

### 6. Uji Coba Data Baru

Setelah model dilatih dan dievaluasi, tahap selanjutnya adalah penerapannya pada data baru. Ini memungkinkan kita untuk melihat bagaimana model akan berfungsi pada data yang sebelumnya tidak pernah dilihat dan memberikan indikasi kinerja model pada dunia nyata.

## 3. Hasil dan Pembahasan

### 3.1 Data Survey

Berdasarkan hasil *survey* kepada mahasiswa STIKI Malang Angkatan 2018-2022, didapatkan data data jenis kelamin, jurusan di SMA/K, *skill*, hobi, alasan berkuliah, alasan mengambil prodi, minat matematika, minat bahasa inggris, dan juga asal program studi. Data ini akan digunakan untuk perhitungan pada kecocokan prediksi pada tiap *class*.

### 3.2 Data Testing

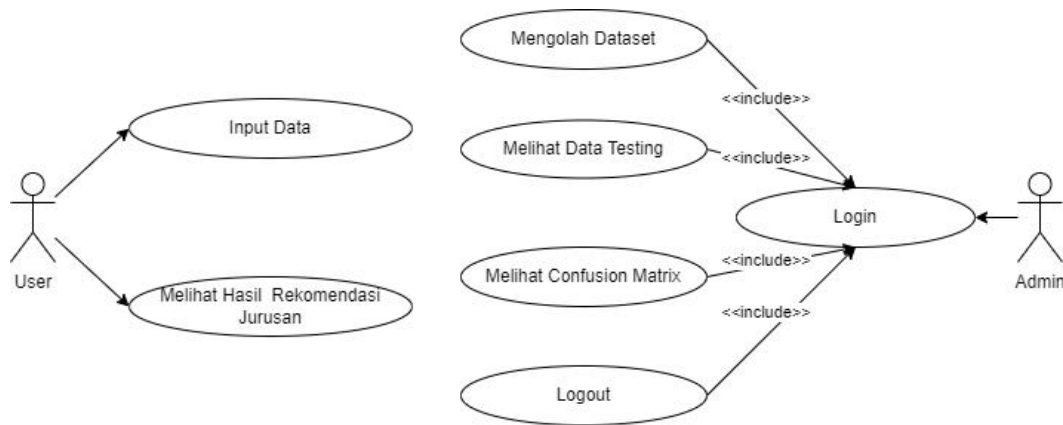
Berikut merupakan sampel dari salah satu mahasiswa STIKI untuk *data testing*.

Tabel 1. Sample Data Testing salah satu mahasiswa

Atribut	Nilai	Nilai Prediksi			
		TI	SI	MI	DKV
Jenis Kelamin	Laki				
Jurusan	IPA				
<i>Skill</i>	Web Programming				
Hobi	Tidur				
Alasan Kuliah	Tidak di terima SBMPTN	0.808799362	0.085255567	0.026017977	0.07992709
Alasan Ambil Prodi	Tertarik dengan Prodi				
Matematika	Suka				
Bahasa Inggris	Suka				

### 3.3 Use Case Diagram

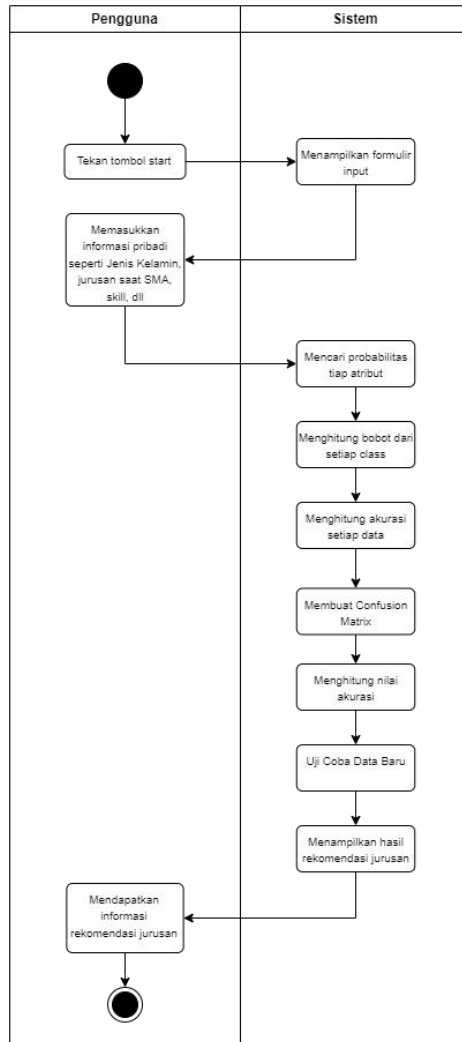
*Use Case* adalah teknik untuk menangkap persyaratan fungsional dari sistem baru atau yang dimodifikasi. Setiap *Use Case* terdiri dari satu atau lebih skenario yang menjelaskan bagaimana sistem berinteraksi dengan pengguna atau sistem lain untuk mencapai tujuan bisnis tertentu. Teknik ini tidak menjelaskan cara kerja bagian dalam sistem atau implementasinya. Yang ditampilkan adalah langkah-langkah yang dilakukan pengguna untuk menggunakan perangkat lunak (Artina 2006). Gambar 1 menunjukkan bagaimana hubungan antara aktor sebagai *User* dan juga *Admin*. *User* dapat menginputkan data sesuai keinginannya dan akan mendapat hasil rekomendasi jurusan sesuai dengan data yang di-input-kan. Sedangkan *Admin* dapat melakukan *login* ke dalam sistem terlebih dahulu sebelum dapat melakukan mengolah *dataset*, melihat *data testing*, melihat hasil perhitungan *confusion matrix*, dan juga *logout*.



Gambar 1. Use Case Diagram

### 3.4 Activity Diagram

Activity diagram menggambarkan berbagai aliran aktivitas dalam sistem yang sedang di rancang, bagaimana masing-masing aliran berawal, decision yang mungkin terjadi dan bagaimana mereka berakhir (Ismail 2020). Untuk *User*, proses dimulai dengan menekan tombol start, lalu sistem akan menampilkan formulir untuk di isi oleh *User*. Setelah user mengisi formulir, sistem akan melakukan perhitungan *Naive Bayes* untuk menghitung prediksi rekomendasi jurusan sesuai data yang sudah di isi oleh *User*.



Gambar 2. Activity Diagram User

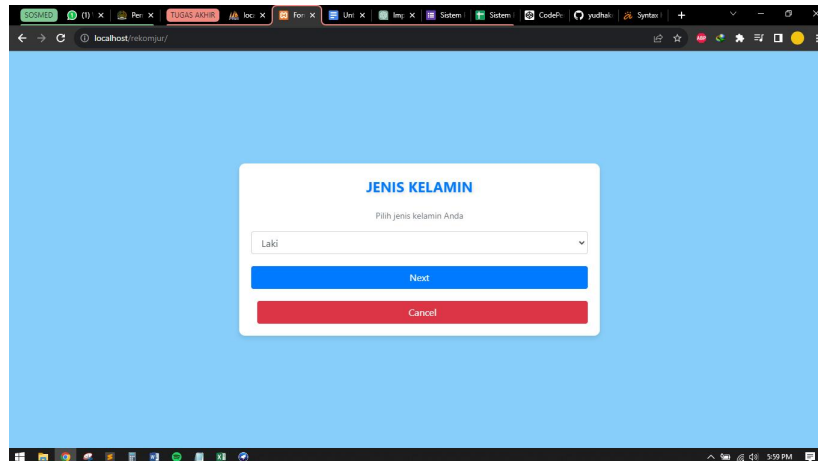
### 3.5 Implementasi

Berikut adalah hasil implementasi setelah *website* dibuat:



Gambar 3. Halaman Awal User

Halaman awal untuk *User* seperti berikut. *User* dapat menekan tombol *start* untuk memulai mengisi data pribadi dari *user*.



Gambar 4. Halaman Input Data User

Halaman berikutnya adalah halaman untuk memilih jenis kelamin dari *User*. Setelah data sesuai, *User* dapat menekan tombol *next* untuk melanjutkan mengisi data dan apabila ingin membatalkan pengisian, *User* dapat menekan tombol *cancel*.

### 3.6 Confussion Matrix

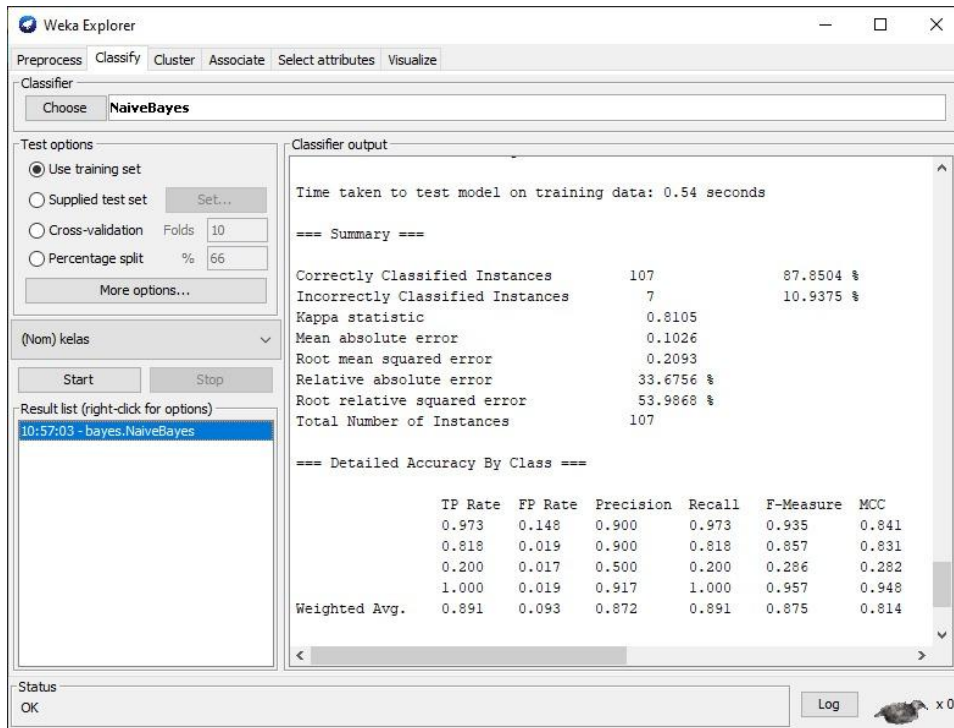
Berdasarkan hasil penelitian, didapat hasil perhitungan *confussion matrix* sebagai berikut:

Tabel 2. Hasil Confussion Matrix

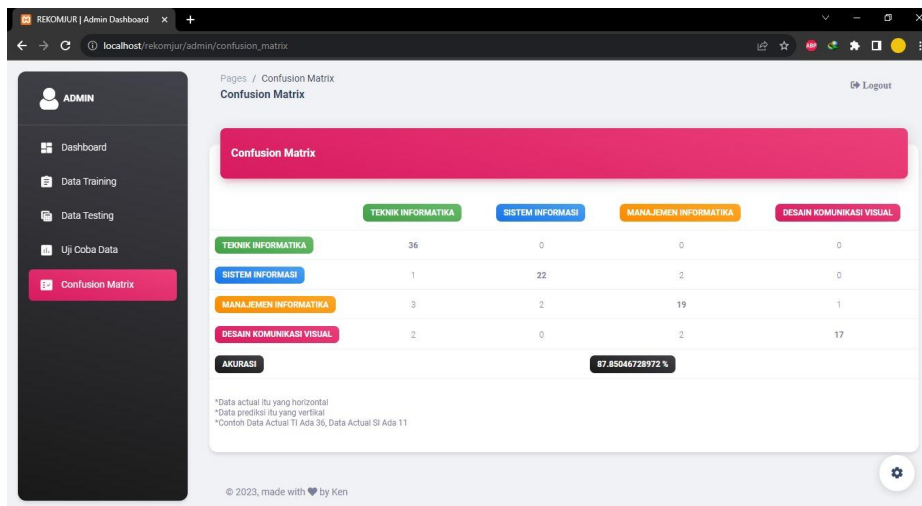
<i>Confussion Matrix</i>				
Class	TI	SI	MI	DKV
TI	36	0	0	0
SI	1	22	1	0
MI	3	2	19	1
DKV	2	0	2	17
<b>Akurasi</b>			87.8504672897196%	

### 3.7 Uji Coba

Pada tahap terakhir, yaitu dilakukan perhitungan dengan menggunakan aplikasi WEKA dan uji coba terhadap setiap fungsi dari fitur yang ada pada *website*. Berikut hasil dari uji coba perhitungan *confussion matrix* dengan menggunakan aplikasi WEKA.



Gambar 5. Hasil Confussion Matrix Menggunakan Aplikasi WEKA



Gambar 6. Hasil Confussion Matrix Website

Setelah dilakukan pencocokan antara *Confussion Matrix* menggunakan aplikasi WEKA dan *website*, Dilakukannya uji coba fungsi menggunakan *black box testing*. Uji coba ini dilakukan agar diharapkan dapat menghasilkan sistem sesuai dengan perancangan sebelumnya. Berikut adalah hasil uji coba *website* menggunakan metode *black box* yang telah dilakukan:



Tabel 3. Black Box User

No	Pengujian	Prosedur	Hasil yang diharapkan	Keterangan
1.	Mulai Program	Klik tombol <i>start</i>	Pindah ke halaman <i>input jenis kelamin</i>	valid
2.	Tombol <i>next</i> pada <i>input</i> jenis kelamin	Klik tombol <i>next</i>	Pindah ke halaman <i>input</i> jurusan	valid
3.	Tombol <i>next</i> pada <i>input</i> jurusan	Klik tombol <i>next</i>	Pindah ke halaman <i>input skill</i>	valid
4.	Tombol <i>next</i> pada <i>input skill</i>	Klik tombol <i>next</i>	Pindah ke halaman <i>input hobby</i>	valid
5.	Melanjutkan ke <i>section input hobby</i>	Klik tombol <i>next</i>	Pindah ke halaman <i>input</i> alasan kuliah	valid
6.	Melanjutkan ke <i>section input</i> alasan kuliah	Klik tombol <i>next</i>	Pindah ke halaman <i>input</i> alasan prodi	valid
7.	Melanjutkan ke <i>section input</i> alasan kuliah	Klik tombol <i>next</i>	Pindah ke halaman <i>input</i> matematika	valid
8.	Melanjutkan ke <i>section input</i> matematika	Klik tombol <i>next</i>	Pindah ke halaman <i>input</i> bahasa inggris	valid
9.	Melanjutkan ke <i>section input</i> bahasa inggris	Klik tombol <i>next</i>	Pindah ke halaman hasil rekomendasi jurusan dan menampilkan rekomendasi	valid
10.	Mengulangi program	Klik tombol coba lagi	Pindah ke halaman mulai program	valid

Tabel 4. Black Box Admin

No	Pengujian	Prosedur	Hasil yang diharapkan	Keterangan
1.	<i>Login</i>	Memasukkan username dan password.	Berhasil melakukan <i>login</i>	valid
2.	Pilih menu <i>Data Training</i>	Klik tombol <i>Data Training</i>	Menampilkan <i>Data Training</i>	valid
3.	Pilih menu <i>Data Testing</i>	Klik tombol <i>Data Testing</i>	Menampilkan <i>Data Testing</i>	valid
4.	Pilih menu <i>Confusion Matrix</i>	Klik tombol <i>Confusion Matrix</i>	Menampilkan data dan hasil serta akurasi	valid

#### 4. Kesimpulan

Penelitian ini berhasil mengembangkan sebuah sistem rekomendasi jurusan berbasis web menggunakan algoritma Naive Bayes Gaussian sebagai solusi atas masalah kebingungan siswa dalam memilih jurusan di STIKI Malang. Dengan keterbatasan informasi dan bimbingan karir, siswa lulusan SMA/K sering kesulitan dalam menentukan jurusan yang cocok dengan minat dan potensinya. Sistem rekomendasi ini mampu memberikan solusi dengan akurasi 87,85% dalam mengidentifikasi jurusan yang sesuai berdasarkan atribut seperti jenis kelamin, jurusan sebelumnya, keterampilan, hobi, alasan kuliah, alasan memilih program studi, minat terhadap matematika, dan minat terhadap bahasa Inggris.

Hasil penelitian ini menunjukkan bahwa metode Naive Bayes Gaussian efektif dalam mengatasi ketidakpastian dalam pemilihan jurusan. Dengan adanya sistem ini, diharapkan siswa lulusan SMA/K dapat membuat keputusan yang lebih informatif dan sesuai dengan minat serta potensinya. Penggunaan teknologi rekomendasi dalam bidang pendidikan dapat menjadi langkah penting dalam membantu siswa mencapai kesuksesan akademik yang lebih baik.

## 5. Referensi

- Addin Aditya and Febry Eka Purwiantono. 2020. "The Application of Fuzzy-Analytical Hierarchy Process Method for Majors Selection at Public Universities." *JITE (Journal of Informatics and Telecommunication Engineering)* 3(2):240-251. doi: 10.31289/jite.v3i2.3245
- Artina, Nyimas. 2006. "Penerapan Analisis Kebutuhan Metode Use Case Metode." *Jurnal Ilmiah STMIK GI MDP Volume 2 N:1-6*.
- Ashari, Hamdhan, Deni Arifianto, Habibatul Azizah, and Al Faruq. 2020. "Perbandingan Kinerja Algoritma Multinomial Naïve Bayes (Mnb), Multivariate Bernoulli Dan Rocchio Algorithm Dalam Klasifikasi Konten Berita Hoax Berbahasa Indonesia Pada Media Sosial." 1-12.
- Hadi Priyono, Retno Sari and Tati Mardiana. 2022. "Klasifikasi Pemilihan Jurusan Sekolah Menengah Kejuruan Menggunakan Gradient Boosting Classifier." *JURNAL INFORMATIKA* 9(2):131~139.
- Ismail. 2020. "Perancangan Sistem Aplikasi Pemesanan Makanan Dan Minuman Pada Cafeteria NO Caffe Di Tanjung Balai Karimun Menggunakan Bahasa Pemrograman PHP Dan MySQL." *Jurnal Tikar* 1(2):192-206.
- Herdiansah, Arief. 2020. "Sistem Pendukung Keputusan Referensi Pemilihan Tujuan Jurusan Teknik Di Perguruan Tinggi Bagi Siswa Kelas Xii Ipa Menggunakan Metode Ahp." *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer* 19(2):223-34. doi: 10.30812/matrik.v19i2.579.
- Mujahidin, Syamsul, Bagus Prasetio, and Muchammad Chandra Cahyo Utomo. 2022. "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian Naïve Bayes." *Jurnal Vocational Teknik Elektronika Dan Informatika* 10(3):17-24.
- Nilam Ramadhania, Zainollah Effendy and Irwan Darmawan. 2022. "Penerapan Algoritma Naïve Bayes Classifier dan Fungsi Gaussian Untuk Penentuan Penjurusan Siswa Kelas X." *SMARTICS Journal* 8(1):14-21. doi: 10.21067/smartics.v8i1.6996