

**PENERAPAN ALGORITME NAÏVE BAYES UNTUK
PEMILIHAN SISWA JURUSAN IPA DAN IPS**

SKRIPSI



Disusun oleh
Gusti Rila Praja
18.11.2270

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENERAPAN ALGORITME NAÏVE BAYES UNTUK
PEMILIHAN SISWA JURUSAN IPA DAN IPS**

SKRIPSI

Untuk Memenuhi Sebagian Persyaratan
Mencapai Gelar Sarjana
Pada Program Studi Informatika



Disusun oleh
Gusti Rila Praja
18.11.2270

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

**PENERAPAN ALGORITME NAÏVE BAYES UNTUK PEMILIHAN
SISWA JURUSAN IPA DAN IPS**

yang dipersiapkan dan disusun oleh

Gusti Rila Praja

18.11.2270

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Agustus 2022

Dosen Pembimbing,

Theopilus Bayu Sasongko, S.Kom., M.Eng.

NIK. 190302375

PENGESAHAN

SKRIPSI

**PENERAPAN ALGORITME NAÏVE BAYES UNTUK PEMILIHAN
SISWA JURUSAN IPA DAN IPS**

yang dipersiapkan dan disusun oleh

Gusti Rila Praja

18.11.2270

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dina Maulina, M.Kom.

NIK. 190302250

Bety Wulan Sari, M.Kom.

NIK. 190302254

Theopilus Bayu Sasongko, S.Kom., M.Eng.

NIK. 190302375

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Gusti Rila Praja
NIM : 18.11.2270

Menyatakan bahwa Skripsi dengan judul berikut:

PENERAPAN ALGORITME NAÏVE BAYES UNTUK PEMILIHAN SISWA JURUSAN IPA DAN IPS

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Agustus 2022 Yang

Menyatakan,



Gusti Rila Praja

MOTTO

“Sebaik-baik nya manusia adalah yang bermanfaat bagi manusia yang lain”

“Sukses adalah saat persiapan dan kesempatan bertemu”

“Kalah saat memulai pdada garis *start* bukan berarti tidak mempunyai kesempatan yang sama untuk mencapai garis *finish*”



PERSEMBAHAN

Dalam menyelesaikan skripsi ini banyak pihak yang berperan sehingga skripsi ini dalam terselesaikan dengan baik. Penulis mengucapkan terimakasih kepada :

1. Allah SWT dengan segala kuasa-Nya dan Rasulnya yang telah membimbing penulis dalam berikhtiar, memberikan rezeki, dan memberikan ilmu pengetahuan dalam menyelesaikan skripsi ini,
2. Bapak Drs. Agus Supriyadi dan Ibu Sugiati, S.Pd., selaku orang tua penulis yang telah memberikan doa, semangat dan motivasi kepada penulis,
3. Adik Asti Aulia Jayanti selaku adik yang telah memberikan semangat dalam skripsi,
4. Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng selaku dosen pembimbing yang telah membimbing dan mengarahkan penulis dalam menyelesaikan skripsi,
5. Mas Erwin Ginanjar Wijanarko, S.Kom., selaku saudara sepupu yang telah memberikan motivasi dan bantuan dalam skripsi ini,
6. Perkumpulan Sholihin Squad, Muftazar Ilham Yudistira, Al-lanang Cahaya Gusti Agung dan kawan-kawan yang telah menghibur penulis dalam mengerjakan skripsi,
7. Perkumpulan Antis the series, Rizki Ridha Atmaja, Faiz Al-haq Maulabeta Raya, S.H., Daniel Unggun Hendra Nugraha dan kawan-kawan yang telah menghibur penulis dalam masa-masa pembuatan skripsi,
8. Mas Adi Zaenul Mustaqim, S.Kom., yang telah membimbing dan membantu penulis dalam skripsi ini,
9. Diri Sendiri

KATA PENGANTAR

Puji syukur saya ucapkan atas kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Penerapan Algoritme *Naïve Bayes* untuk Pemilihan Siswa Jurusan IPA dan IPS”. Penulis mengucapkan terimakasih kepada pihak yang telah memberikan semangat, doa serta dukungan kepada penulis selama mengerjakan skripsi.

Penulis menyadari bahwa naskah skripsi ini masih terdapat kekurangan, dalam hal ini penulis mengharapkan kritik dan saran dari pembaca untuk melakukan perbaikan. Smoga skripsi ini dapat bermanfaat bagi pembaca.

Yogyakarta, 5 Agustus 2022

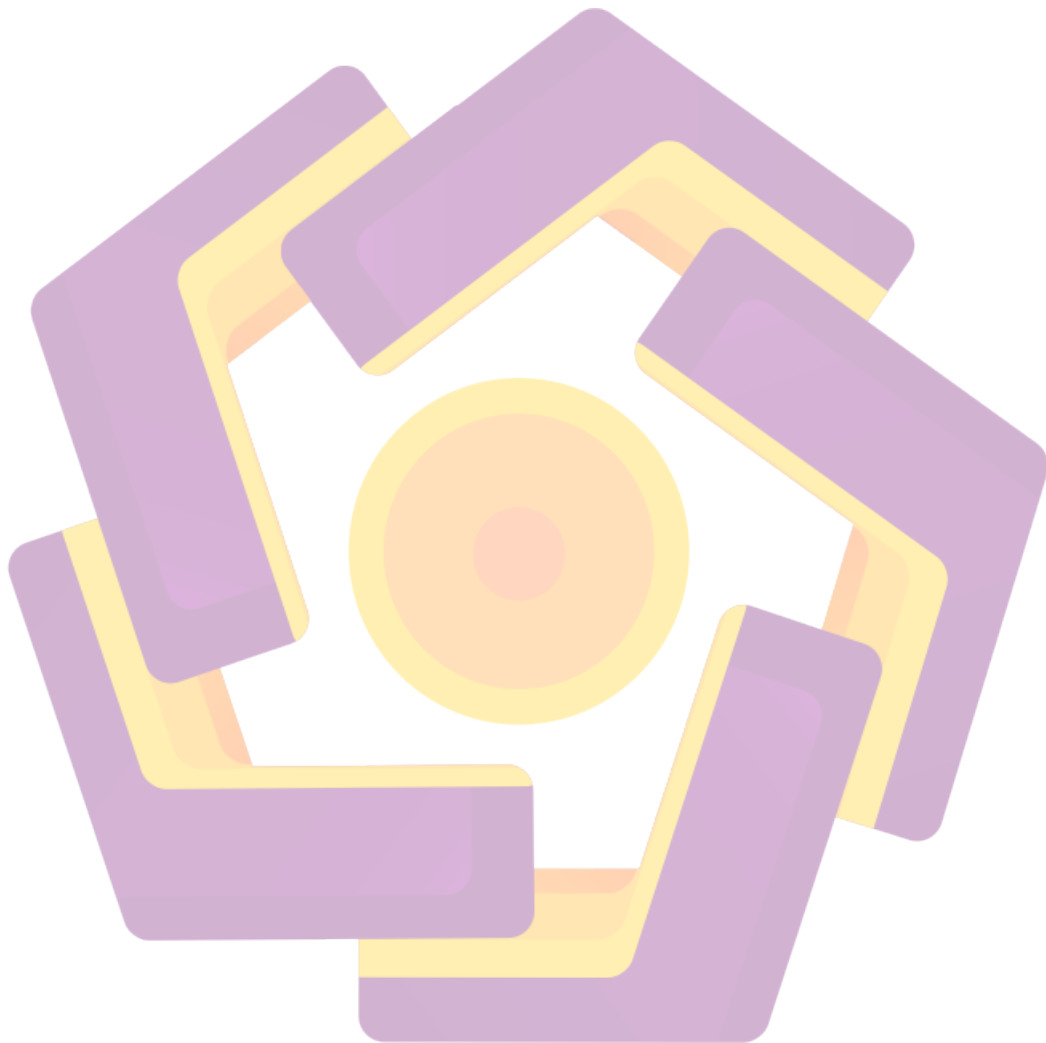
Penulis

DAFTAR ISI

JUDUL.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	Error! Bookmark not defined.
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
INTISARI.....	xiii
<i>ABSTRACT</i>	xiv
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian	4
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.6.1 Metode Pengumpulan Data.....	5
1.6.2 Metode Analisis	5
1.6.3 Metode Implementasi.....	5
1.6.4 Metode Testing.....	7
1.7 Sistematika Penulisan.....	8
BAB II.....	10
LANDASAN TEORI.....	10
2.1 Tinjauan Pustaka	10
2.2 Penjurusan Siswa.....	18

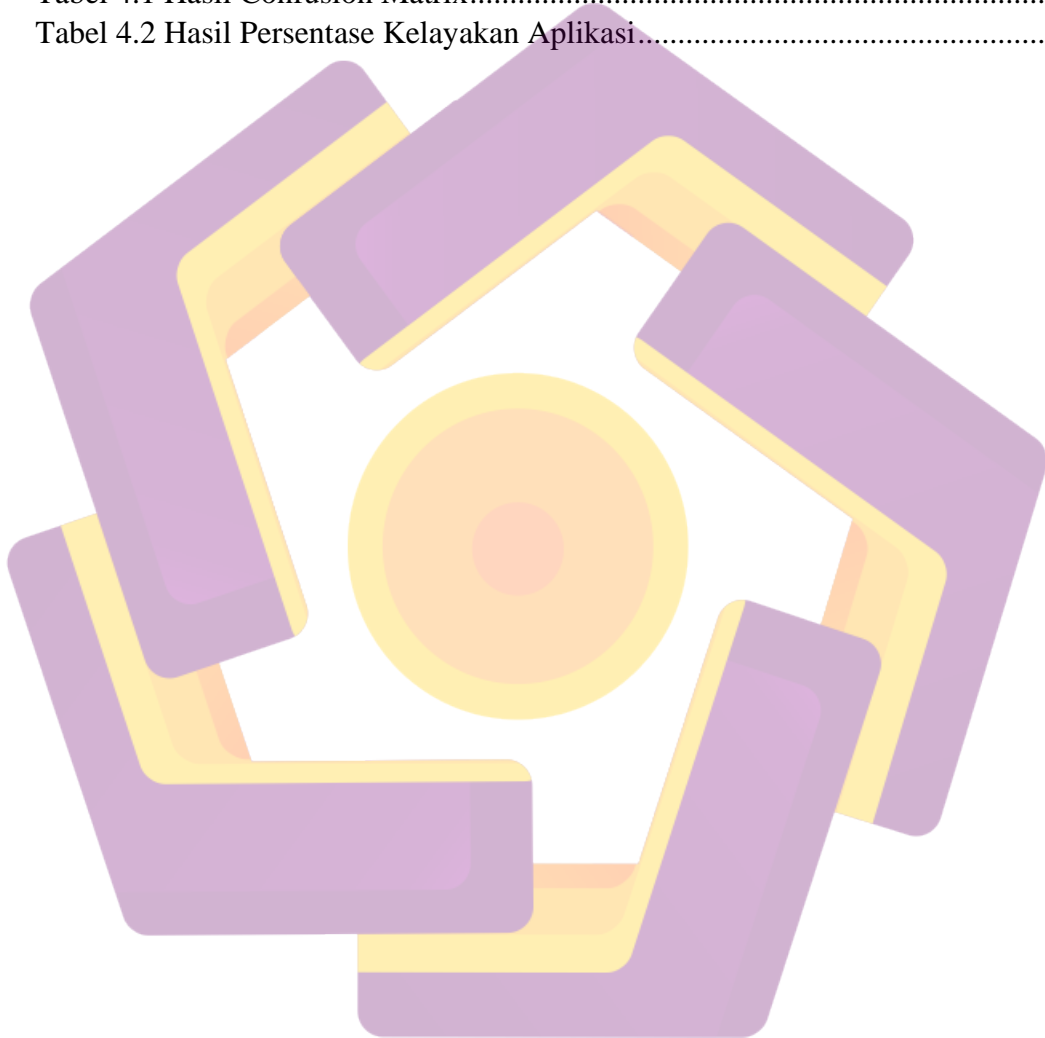
2.3	Data Mining.....	20
2.4	Klasifikasi.....	21
2.5	Naive Bayes.....	22
2.5.1	<i>Multinomial Naive Bayes</i>	22
2.5.2	<i>Burnoulli Naive Bayes</i>	23
2.5.3	<i>Gaussian Naive Bayes</i>	23
2.6	Python.....	24
2.7	Flowchart.....	24
2.8	<i>Confusion Matrix</i>	27
2.9	Kurva ROC dan AUC.....	29
2.10	Google Colab.....	30
BAB III	31
ANALISIS DAN PERANCANGAN	31
3.1	Tinjauan Umum.....	31
3.2	Pengumpulan Data	33
3.3	<i>Data Selection</i>	45
3.4	<i>Pre-processing</i>	46
3.5	<i>Naive Bayes Clasification Model</i>	46
3.6	Perancangan Sistem.....	58
BAB IV	62
HASIL DAN PEMBAHASAN	62
4.1	Data Selection.....	62
4.2	Pre-processing	63
4.3	Naive Bayes Model	65
4.4	Pengujian	67
4.4.1	Pengujian <i>Confusion Matrix</i>	67
4.4.2	Pengujian <i>KFold Cross Validation</i>	70
4.4.3	Pengujian Aplikasi di Sekolah	71
4.5	Implementasi pada <i>Website</i>	73
4.6	Antarmuka Pengguna	75
BAB V	78
PENUTUP	78

5.1 Kesimpulan.....	78
5.2 Saran.....	79
DAFTAR PUSTAKA	80
LAMPIRAN.....	84



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian.....	13
Tabel 2.2 Simbol pada Flowchart	25
Tabel 2.3 Tabel Confusion Matrix	27
Tabel 3.1 Informasi Atribut Dataset	33
Tabel 4.1 Hasil Confusion Matrix.....	69
Tabel 4.2 Hasil Persentase Kelayakan Aplikasi.....	71



DAFTAR GAMBAR

Gambar 4.1 Dataset Asli	62
Gambar 4.2 Source code untuk data selection	62
Gambar 4.3 Hasil data selection.....	63
Gambar 4.4 Code untuk Mencari Data Kosong	64
Gambar 4.5 Mengisi Data Kosong pada Tipe Data Object.....	64
Gambar 4.6 Mengisi Data Kosong pada Tipe Data Numerik	64
Gambar 4.7 Convert Nilai Psikologi.....	64
Gambar 4.8 Convert Minat Siswa.....	65
Gambar 4.9 Pembagian data	65
Gambar 4.10 Naive Bayes Model	66
Gambar 4.11 Hasil Perbandingan Akurasi.....	66
Gambar 4.12 Source code pada Confusion Matrix	67
Gambar 4.13 Hasil Confusion Matrix.....	68
Gambar 4.14 Source code untuk ROC dan AUC.....	69
Gambar 4.15 Kurva ROC.....	70
Gambar 4.16 Pengujian KFold Cross Validation.....	70
Gambar 4.17 Hasil Pengujian KFold Cross Validation	70
Gambar 4.18 Skala Kategori Kelayakan Hasil Kuisisioner.....	72
Gambar 4.19 Klasifikasi Peminatan Siswa Menggunakan Flask.....	74
Gambar 4.20 Tampilan Website	76
Gambar 4.21 Hasil Prediksi Minat Siswa IPA.....	77
Gambar 4.22 Hasil Prediksi Minat Siswa IPS	77

INTISARI

Pendidikan merupakan aspek penting di masyarakat sebagai salah satu faktor penunjang keberhasilan suatu keluarga dan negara, pada era modern saat ini dalam sektor pendidikan terdapat banyak tantangan dan masalah yang harus dihadapi. Masalah yang dialami adalah ketika peserta didik memasuki masa sekolah SMP (Sekolah Menengah Pertama) dan melanjutkan sekolah ke jenjang yang lebih tinggi yaitu sekolah SMA (Sekolah Menengah Atas). Dengan menggunakan system kurikulum 2013, pemilihan jurusan dimulai saat siswa memasuki kelas 1/X SMA. Pemilihan jurusan tersebut adalah jurusan IPA dan IPS. Nilai mata pelajaran selama peserta didik sekolah di SMP menjadi parameter dalam pemilihan jurusan IPA dan IPS. Banyaknya jumlah siswa menjadi kendala tersendiri pada sekolah maupun guru karena sistem penjurusan yang masih menggunakan cara manual.

Untuk mengatasi masalah penjurusan yang masih menggunakan cara manual, dilakukan klasifikasi menggunakan teknik data mining. Algoritme yang digunakan penulis disini yaitu algoritme *Gaussian Naïve Bayes* untuk memprediksi peminatan siswa dengan mengklasifikasi siswa minat IPA atau minat IPS. Dataset terdiri dari 214 data siswa dan 2 label yaitu minat siswa IPA dan IPS. Dilakukan proses seleksi data dalam memilih atribut. Proses penelitian pada model pembagian data dengan format 80% untuk data *training* dan 20% sebagai data *testing*.

Hasil Pengujian membuktikan menggunakan algoritme *Multinomial Naive Bayes* mendapatkan akurasi sebesar 53,48%, algoritme *Bernoulli Naive Bayes* sebesar 44,18% dan *Gaussian Naive Bayes* sebesar 86,04%. Evaluasi program menggunakan metode *confusion matrix* dan kurva ROC. Menggunakan algoritme *Gaussian Naïve Bayes* akurasi yang didapatkan sebesar 86,04% dan AUC sebesar 85,86%.

Kata Kunci: Pemilihan Jurusan, Klasifikasi, *Multinomial Naive Bayes*, *Bernoulli Naive Bayes*, *Gaussian Naive Bayes*, *Confusion Matrix*, *ROC*.

ABSTRACT

Education is an important aspect of society as one of the factors supporting the success of a family and a country. In today's modern era in the education sector, there are many challenges and problems that must be faced. The problem experienced is when students enter junior high school (junior high school) and continue their education to a higher level, namely high school (high school). By using the 2013 curriculum system, the selection of majors begins when students enter class 1/X SMA. The choice of majors is between the science and social studies majors. The value of subjects while students are in junior high school is a parameter in the selection of science and social studies majors. A large number of students is a separate obstacle for schools and teachers because the major systems are still using the manual method.

To overcome the problem of majors that still use the manual method, classification is carried out using data mining techniques. The algorithm used by the author here is the Gaussian Naïve Bayes algorithm to predict student interest by classifying students' interests in science or social studies interests. The dataset consists of 214 student data and 2 labels, namely the interests of science and social studies students. The data selection process is carried out in selecting attributes. The data sharing model research process with 80% format for training data and 20% format for testing data.

The test result prove that using the Multinomial Naïve Bayes algorithm gets an accuracy of 53.48%, the Bernoulli Naïve Bayes algorithm is 44.18% and Gaussian Naïve Bayes is 86.04%. Program evaluation uses the confusion matrix and ROC curve methods. Using the Gaussian Naïve Bayes algorithm, the accuracy obtained is 86.04% and AUC is 85.86%.

Keywords: *Selection of Science and Social Studies Majors, Classification, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Gaussian Naïve Bayes, Confusion Matrix, ROC.*