

ANALISIS KELAYAKAN PEMILIHAN JURUSAN SISWA-SISWI SMA/SEDERAJAT DI ERA MERDEKA BELAJAR MENGGUNAKAN ALGORITMA NAÏVE BAYES

Maria Adventina Kolne¹, Yoseph Pius Kurniawan Kelen², Hevi Herlina Ullu³

^{1,2,3}Universitas Timor

(Program Studi Teknologi Informasi Fakultas Pertanian Sains dan Kesehatan Universitas Timor)

(Jl. Km. 09 Kelurahan Sasi, Kecamatan Kota Kefamenanu, telp. 0811 753 2015)

e-mail: ¹mariaakolne@gmail.com, ²yosepkelen@unimor.ac.id, ³heviherlina@unimor.ac.id

Abstrak

Di Era Merdeka Belajar menawarkan fleksibilitas yang lebih besar bagi siswa untuk memilih jurusan sesuai minat dan bakat, terutama bagi sekolah yang sudah menerapkan kurikulum merdeka belajar. Namun masih adanya kendala dalam proses pemilihan jurusan yaitu ada siswa masih bingung dalam memilih jurusan, dan juga sebagian siswa pilih jurusan tidak berdasarkan minat terutama pada mata pelajaran – mata pelajaran yang ada di jurusan tersebut sehingga mengakibatkan adanya kendala dalam proses belajar mengajar, dan proses pemilihan jurusan di SMAN PANTURA masih berbasis manual. Berdasarkan permasalahan yang ada membuat peneliti melakukan penelitian untuk mengetahui klasifikasi jurusan siswa. Metode yang digunakan dalam penelitian ini yaitu dengan menggunakan metode Naive Bayes bahasa pemrograman python, google colab. Data set yang dikumpulkan melalui hasil wawancara, observasi dan studi pustaka, sebanyak 119 siswa sebagai data training dan 91 data siswa sebagai data testing. Penelitian ini bertujuan untuk mengklasifikasi siswa di SMAN PANTURA kedalam jurusan yaitu IPA, BAHASA dan IPS. Setelah dianalisis menggunakan python Pada data training terdapat 3 jurusan yaitu : IPA=29 siswa, BAHASA=10 siswa, dan IPS=80 siswa dengan Nilai akurasi yang diperoleh dari data training yaitu 65%. Sedangkan pada data testing diklasifikasi dalam 3 jurusan yaitu : IPA = 1 siswa, BAHASA = 14 siswa dan IPS = 76 siswa. Dari hasil klasifikasi yang diperoleh dapat disimpulkan bahwa jika data yang digunakan berjumlah banyak maka tingkat nilai akurasi atau kualitas dari hasil klasifikasi semakin baik dan hasil penelitian menunjukkan bahwa penelitian ini dapat memberikan kontribusi atau nilai tambah bagi para guru dan siswa dalam hal mengklasifikasi jurusan siswa.

Kata kunci: Analisis kelayakan, Naive Bayes, Python, Google Colab, SMA Negeri PANTURA.

Abstract

The Independent Learning Era offers greater flexibility for students to choose majors according to their interests and talents, especially for schools that have implemented the independent learning curriculum. However, there are still obstacles in the process of choosing a major, namely that some students are still confused about choosing a major, and also some students choose a major not based on interest, especially in the subjects in that major, resulting in obstacles in the teaching and learning process and the selection process. majors at SMAN PANTURA are still manual based. Based on existing problems, researchers conducted research to find out the classification of students' majors. The method used in this research is using the Naive Bayes method in the Python programming language, Google Colab. The data set was collected through interviews, observations and literature studies, as many as 119 students as training data and 91 students' data as testing data. This research aims to classify students at SMAN PANTURA into majors, namely science, language and social studies. After analyzing using python, the training data contained 3 majors, namely: science=29 students, LANGUAGE=10 students, and social studies=80 students with the accuracy value obtained from the training data being 65%. Meanwhile, the testing data is classified into 3 departments, namely: science = 1 student, LANGUAGE = 14 students and social studies = 76 students. From the classification results obtained, it can be concluded that if a large amount of data is used, the level of accuracy or quality of the classification

results will be better and the research results show that this research can provide a contribution or added value for teachers and students in terms of classifying students' majors.

Keywords: Feasibility analysis, Naive Bayes, Python, Google Colab, PANTURA State High School.

1. PENDAHULUAN

Di era Merdeka Belajar, pendidikan Indonesia mengalami perubahan mendasar menuju sistem yang lebih dinamis dan inklusif. Salah satu aspek penting dari transformasi ini adalah dalam pemilihan jurusan bagi siswa-siswi SMA/sederajat. Keputusan ini memerlukan analisis yang cermat untuk memastikan kesesuaian antara minat siswa dengan pilihan jurusan yang tersedia, terutama bagi sekolah yang sudah menerapkan Kurikulum Merdeka Belajar. Salah satu sekolah yang sudah menerapkan Kurikulum Merdeka Belajar adalah SMA Negeri Pantura.

SMA Negeri PANTURA merupakan salah satu satuan pendidikan jenjang SMA di Desa Oepuah, Kecamatan Biboki Moenleu, Kabupaten Timor Tengah Utara, Nusa Tenggara Timur. Penentuan pemilihan jurusan pada SMA Negeri PANTURA merupakan program rutin yang setiap tahun ajaran baru diselenggarakan yang ditujukan kepada siswa baru yang akan mendaftar di SMA Negeri PANTURA. Jurusan di SMA Negeri PANTURA terdiri dari IPA, IPS, dan Bahasa. Proses pemilihan jurusan dilakukan dengan mempertimbangkan nilai tes tertulis, persetujuan orang tua dan nilai raport SMP siswa.

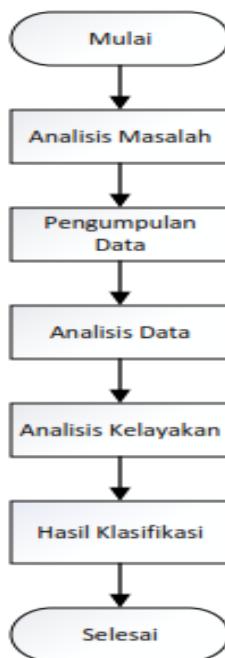
Namun Permasalahan yang sering terjadi pada saat pemilihan jurusan di SMA PANTURA adalah siswa masih kebingungan dalam memilih jurusan, dan juga sebagian siswa pilih jurusan tidak berdasarkan minat terutama pada mata pelajaran – mata pelajaran yang ada di jurusan tersebut sehingga mengakibatkan adanya kendala dalam proses belajar mengajar, dan proses pemilihan jurusan di SMA PANTURA masih berbasis manual. Oleh karena itu diperlukan suatu teknik dalam menganalisis data yang dapat membantu siswa dalam proses pemilihan jurusan. Teknik dalam menganalisis data, dapat menggunakan metode *Naïve Bayes*. Algoritma *Naïve Bayes* merupakan salah satu metode yang dapat digunakan untuk mengklasifikasi data. *Bayesian classification* merupakan algoritma pengklasifikasian statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu class. Algoritma *Naïve Bayes* juga merupakan salah satu metode dalam bidang data mining [1]

Data mining atau disebut juga *Knowladge Discovery in Database* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar. *Output* dalam data mining dapat dipergunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki pengambilan keputusan di masa yang akan datang[2]. Beberapa penelitian terdahulu yang telah berhasil menerapkan algoritma *Naïve Bayes* dalam klasifikasi yaitu, Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Menggunakan Metode *Naïve Bayes* [3], Penerapan Metode *Naïve Bayes* Dengan *Particle Swarm Optimization* (PSO) Untuk Pemilihan Peminatan Jurusan Pada SMK [4], Analisis Kinerja Algoritma C4.5 Dan *Naïve Bayes* Dalam Memprediksi Keberhasilan Sekolah Menghadapi UN [5], dan Perancangan Klasifikasi Algoritma Naive Bayes Pada Data Pemilihan Jurusan Siswa[6].

Dengan mengamati permasalahan yang ada dan mempelajari penelitian terdahulu yang telah memberikan solusi, maka diusulkan penelitian untuk menganalisis kelayakan jurusan siswa di era merdeka belajar menggunakan algoritma *naïve bayes*. Penelitian tersebut dilakukan untuk membantu dan memudahkan para guru dan siswa untuk melakukan analisis pemilihan jurusan.

2. METODE PENELITIAN

Gambar berikut merupakan ilustrasi yang memperlihatkan tahapan penelitian yang dilakukan oleh peneliti adalah sebagai berikut :



Gambar 1. Tahapan Penelitian

Penelitian ini dimulai dengan analisis masalah dengan cara melakukan wawancara dengan wakil kepala sekolah SMA NEGERI PANTURA dan membaca beberapa jurnal referensi yang berkaitan dengan permasalahan Kelayakan Pemilihan Jurusan Siswa. Setelah masalah ditemukan, Maka dilakukan pengumpulan data sesuai dengan analisa masalah. Proses pengumpulan data menggunakan beberapa teknik yaitu : wawancara, studi pustaka dan observasi. Data yang digunakan adalah data siswa kelas X dan XI SMA NEGERI PANTURA tahun pelajaran 2023/2024. Berdasarkan analisis masalah yang telah dilakukan maka perlu diambil data untuk bisa dianalisis seperti data nilai raport SMP, nilai tes masuk dan persetujuan orang tua siswa. Data yang sudah dianalisis dibagi menjadi dua kelompok data yaitu data *training* sebanyak 119 siswa dan data *testing* sebanyak 91 siswa. Setelah mengelompokkan data, maka dilakukan analisis kelayakan dengan metode *Naïve Bayes*. Setelah melakukan analisis kelayakan menggunakan metode *naïve bayes* maka akan di peroleh hasil klasifikasi berupa siswa sudah di klasifikasikan kedalam jurusan dan proses selesai.

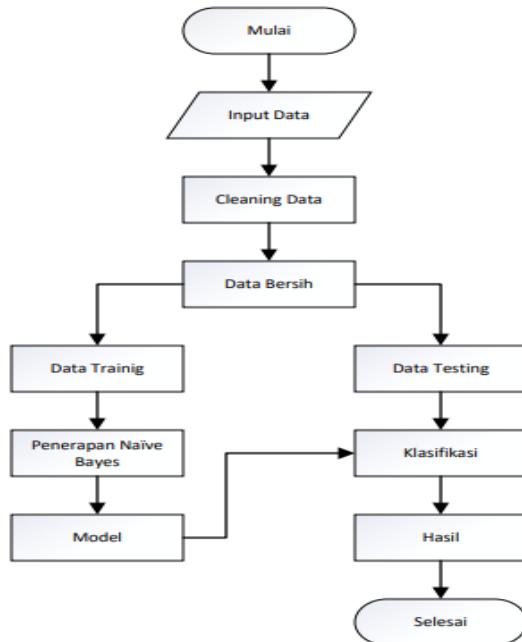
2.1. SMA Negeri PANTURA

SMAN PANTURA adalah salah satu satuan pendidikan dengan jenjang SMA di Oepuah, Kec. Biboki Moenleu, Kab. Timor Tengah Utara, Nusa Tenggara Timur yang didirikan pada tanggal 21 Desember 2012. Dalam menjalankan kegiatannya, SMAN PANTURA berada di bawah naungan Kementerian Pendidikan dan Kebudayaan.

SMAN PANTURA terdapat tiga jurusan yaitu : IPA, IPS, dan Bahasa. Dalam Proses Pemilihan Jurusan masih dilakukan secara manual sehingga mengakibatkan siswa masih kebingungan dalam memilih jurusan, dan sebagian siswa memilih jurusan tidak berdasarkan minatnya, terutama pada mata pelajaran jurusan tersebut, sehingga mengalami kendala dalam proses belajar mengajar.

2.2. Analisis Kelayakan

Berikut ini merupakan tahapan analisis kelayakan klasifikasi jurusan siswa sebagai berikut :



Gambar 2. Analisis Kelayakan

Dimulai dari menginput data, setelah itu melakukan *cleaning* data atau menghapus data untuk membersihkan data yang duplikat, data kosong, tidak valid atau tidak relevan, sehingga memperoleh data bersih yang sudah dibagi menjadi data *train* dan data *test*. Pada Data *train* berisi 119 data siswa dengan 11 *Fitur* diantaranya : Nama Siswa, Nilai Raport SMP (Agama, Pkn, Bindo, Binggris, Mat, Ipa, Ips), Nilai Tes Masuk, Persetujuan Ortu Dan Jurusan. Selanjutnya Data *Train* akan diklasifikasi (*Training*) dengan menggunakan metode *naïve bayes*. Metode ini akan digunakan untuk menghitung Akurasi, *Precision*, dan *Recall*nya. Selanjutnya gunakan *Model* tersebut untuk klasifikasi data *testing*. Data *Test* berisi 91 data siswa dengan 11 fitur yang sama seperti Data *Train*. Setelah melakukan klasifikasi pada data *testing*, maka dapat memperoleh hasil klasifikasi jurusan.

2.3. Data Mining

Data mining adalah penguraian (yang tidak sederhana) dari sekumpulan data menjadi informasi yang memiliki potensi secara implisit (tidak/jelas) yang sebelumnya tidak diketahui. Data mining atau disebut juga *Knowledge Discovery in Database* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaitan data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar[6].

Data mining merupakan ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki[7].

2.4. Naïve Bayes

Algoritma Naïve Bayes merupakan metode untuk menyelesaikan permasalahan klasifikasi. Konsep yang digunakan oleh naïve bayes yakni menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan dalam melakukan pengklasifikasian [8]. Teorema keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Teorema Bayes berfungsi untuk menentukan atau menghitung probabilitas dari sebuah hipotesis berdasarkan probabilitas prior. Prior Probabilitas adalah nilai dari probabilitas yang sejak awal diyakini benar, sebelum melakukan pengamatan dan analisis terhadap data yang dimiliki [9]. Teorema Bayes juga memiliki asumsi *class conditional independent*, yang artinya nilai dari atribut dari kelas yang ada bersifat independent atau tidak saling memberikan pengaruh. Misalkan X adalah sampel data yang label kelasnya tidak diketahui sedangkan H adalah hipotesis bahwa Y termasuk dalam kelas H. Dengan *training* data X dan Probabilitas posteriori dari hipotesis H maka formula klasifikasi dengan Teorema Bayes lainnya sebagai berikut[10] :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.2)$$

Dimana :

$P(H|X)$: Probabilitas posterior dari kelas H terhadap X atau probabilitas dari suatu kejadian X pada kelas H

$P(X|H)$: Probabilitas posterior X dikondisikan pada kelas H atau probabilitas dari kondisi X dan apabila berada di kelas H

$P(X)$: Probabilitas dari kelas X atau probabilitas prior dari X

$P(H)$: Probabilitas dari kelas H atau probabilitas prior dari H

2.5. Bahasa Pemrograman Python

Python adalah sebuah bahasa pemrograman yang sangat populer yang memiliki banyak keuntungan dalam mendukung pemrograman berorientasi objek dan dapat berjalan pada berbagai platform sistem operasi seperti komputer pribadi, Macintosh, dan UNI [11]. Beberapa keunggulan dari bahasa pemrograman Python adalah sebagai berikut :

1. Program dapat dikembangkan dengan cepat dengan sedikit kode yang diperlukan.
2. Python mendukung multiplatform, artinya dapat digunakan di berbagai sistem operasi.
3. Bahasa pemrograman Python relatif mudah dipelajari.
4. Python memiliki sistem pengelolaan memori otomatis.
5. Python adalah bahasa pemrograman berorientasi objek.

2.6. Google Colaboratory

Colaboratory, atau “Colab” merupakan produk dari *Google Research*. Colab memungkinkan siapa saja menulis dan mengeksekusi kode *python* arbitrer melalui browser, dan sangat cocok untuk *machine learning*, analisis data, serta pendidikan. Secara lebih teknis, Colab merupakan layanan *notebook Jupyter* yang dihosting dan dapat digunakan tanpa penyiapan, serta menyediakan akses gratis ke *resource* komputasi termasuk GPU. *Resource Colab* tidak dijamin dan sifatnya terbatas, serta batas penggunaannya terkadang berfluktuasi. Hal ini diperlukan agar Colab dapat menyediakan *resource* secara gratis[12].

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

Implementasi metode *Naïve bayes* untuk menganalisis jurusan siswa di SMAN PANTURA dilakukan untuk mengklasifikasi jurusan siswa secara otomatis dan sistematis. Tahapan awal yang dilakukan yaitu melakukan pengumpulan data. Data yang digunakan yaitu data siswa kelas X dan XI SMAN PANTURA tahun pelajaran 2023/2024 yaitu data nilai raport SMP (agama, pkn, bindo, binggris, mat, ipa, ips), nilai_tes dan persetujuan_ortu siswa. Setelah data dikumpulkan tahapan selanjutnya yaitu membagi data menjadi dua yaitu data *training* dan data *testing*. Data *training* yang digunakan adalah data siswa kelas XI yang berjumlah 119 siswa dan data *testing* yang digunakan adalah data siswa kelas X yang berjumlah 91 siswa. Setelah itu melakukan analisis data pada data *training* dan data *testing* dengan bahasa pemrograman *python* menggunakan metode *naive bayes*. Kontribusi dari implementasi metode *Naïve bayes* dalam mengklasifikasi jurusan siswa dapat memberikan manfaat yang berarti bagi para guru dan siswa di sekolah. Dari hasil klasifikasi yang diperoleh dari penelitian ini, para guru dapat mengklasifikasi jurusan siswa memerlukan perhatian khusus seperti siswa memiliki nilai kurang dari standar Kriteria Ketuntasan Minimal (KKM).

3.2 Hasil

3.2.1 Import library Python

Tahapan pertama yang dilakukan yaitu mengimport library yang dibutuhkan dalam proses *Naïve Bayes* untuk mengklasifikasi jurusan siswa. Proses klasifikasi dilakukan menggunakan *software google colaboratory*. Library yang digunakan yaitu *numpy*, *pandas*, *matplotlib* dan *sklearn*. Pada library *sklearn* terdapat beberapa modul yang digunakan yaitu *GaussianNaiveBayes*, *confusion_matriks*, *classification_report* dan *accuracy_score*. Berikut merupakan *coding* yang digunakan untuk mengimport library python.

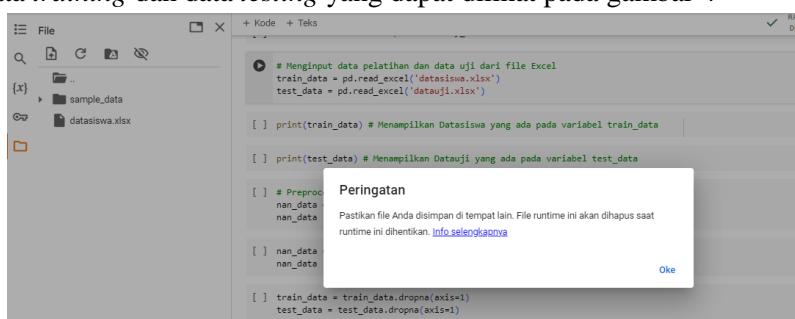
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
```

Gambar 3. Coding import library python

Pada gambar 3 merupakan pengcondingan yang digunakan untuk mengklasifikasi jurusan siswa, yang dimana *library pandas* berfungsi untuk mengambil dataset yang digunakan, *library numpy* berfungsi untuk melakukan komputasi matriks, *library matplotlib* dengan modul *pyplot* digunakan untuk *plotting* atau membuat plot, *library sklearn* dengan modul *preprocessing* diimport *Label Encoder* untuk melakukan proses pemisalan dari bentuk data *string* menjadi integer/angka untuk beberapa data dibeberapa fitur, *library sklearn* dengan modul *model_selection* diimport *train_test_split* untuk melakukan proses pembagian data kedalam bentuk data *training* dan data *testing*, *library sklearn* dengan modul *preprocessing* diimport *GaussianNB* untuk menjalankan algoritma *naïve bayes* menghitung probabilitas dari suatu data kontinyu terhadap kelas tertentu, *library sklearn* dengan modul diimport *classification_repot* untuk mengevaluasi hasil prediksi dari model klasifikasi dengan menampilkan berbagai metriks seperti *precision*, *recall*, *f1-score*, dan *support* untuk setiap kelas. *library sklearn* dengan modul diimport *accuracy_score* digunakan untuk menghitung akurasi model klasifikasi, yaitu proporsi prediksi yang benar dibandingkan dengan total jumlah sampel.

3.2.2 Menginput Data dan Membaca Data

Tahap selanjutnya adalah menginput data *training* dan data *testing* yang akan dianalisis. Data *training* dan data *testing* yang diinput harus dalam format excel dan csv. Untuk menginput data dapat dilakukan dengan *upload* langsung file data ke penyimpanan sesi *python*. Berikut ini tampilan *upload* file data *training* dan data *testing* yang dapat dilihat pada gambar 4



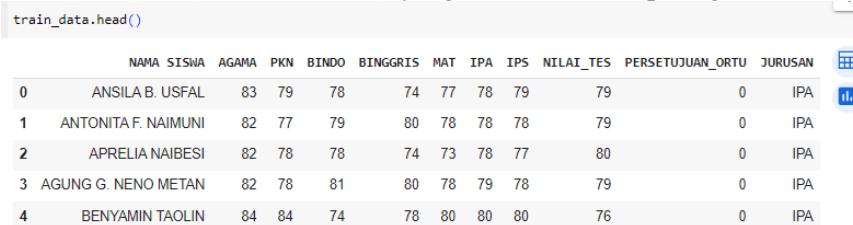
Gambar 4 Tampilan *upload* data *training* dan data *testing* pada *python*

Setelah data *training* dan data *testing* diinput kedalam bahasa *python*, maka berikut ini kode yang digunakan untuk membaca dan menampilkan data *training* dan data *testing*.

```
train_data = pd.read_excel('datasiswa.xlsx')
test_data = pd.read_excel('datauji.xlsx')
```

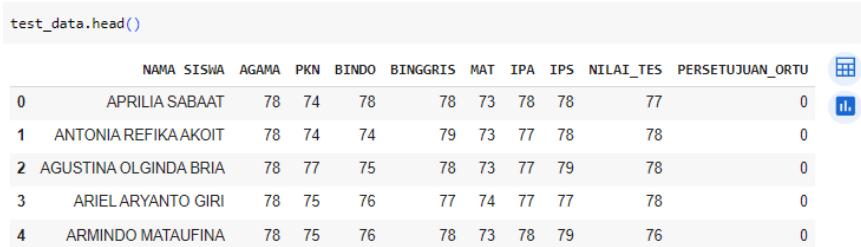
Gambar 5. coding python untuk membaca data *training* dan data *testing*

Setelah kode pada gambar 5 dijalankan akan maka akan muncul data *training* dan data *testing* yang digunakan dimana data *training* dan data *testing* yang akan ditampilkan merupakan *header* atau 5 baris pertama dari data *training* dan data *testing* yang akan muncul seperti gambar berikut :



	NAMA_SISWA	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAI_TES	PERSETUJUAN_ORTU	JURUSAN
0	ANSILA B. USFAL	83	79	78	74	77	78	79	79	0	IPA
1	ANTONITA F. NAIMUNI	82	77	79	80	78	78	78	79	0	IPA
2	APRELIA NAIBESI	82	78	78	74	73	78	77	80	0	IPA
3	AGUNG G. NENO METAN	82	78	81	80	78	79	78	79	0	IPA
4	BENYAMIN TAOLIN	84	84	74	78	80	80	80	76	0	IPA

Gambar 6. Tampilan data *training*



	NAMA_SISWA	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAI_TES	PERSETUJUAN_ORTU	JURUSAN
0	APRILIA SABAAT	78	74	78	78	73	78	78	77	0	IPA
1	ANTONIA REFIKA AKUIT	78	74	74	79	73	77	78	78	0	IPA
2	AGUSTINA OLGINDA BRIA	78	77	75	78	73	77	79	78	0	IPA
3	ARIELARYANTO GIRI	78	75	76	77	74	77	77	78	0	IPA
4	ARMINDO MATAUFINA	78	75	76	78	73	78	79	76	0	IPA

Gambar 7. Tampilan data *testing*

3.2.3 Cleaning Data / Preprocessing Data

Tahapan selanjutnya yaitu *preprocessing* data. Pada tahapan ini dilakukan pembersihan data, penanganan data yang kosong atau hilang (*missing values*). Selain itu juga, pada tahapan ini kita dapat menyeleksi fitur atau atribut yang akan digunakan. Misalnya pada proses klasifikasi hanya berlaku untuk fitur yang berjenis angka atau numerik, maka data yang memiliki tipe data teks akan dihapus atau di *drop*. Berikut coding yang digunakan untuk *drop* kolom yang tidak digunakan :

```
train_data = train_data.dropna(axis=1)
train_data = train_data.drop(['NAMA_SISWA', 'JURUSAN'], axis=1)
```

Gambar 8. Coding hapus kolom yang tidak digunakan pada data *training*

```
test_data = test_data.dropna(axis=1)
test_data = test_data.drop('NAMA_SISWA', axis=1)
```

Gambar 9. Coding hapus kolom yang tidak digunakan pada data *testing*

Setelah kode di atas dijalankan maka langkah selanjutnya yaitu menampilkan data *training* dan data *testing* baru yang digunakan sebagai berikut :

	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAT_TES	PERSETUJUAN_ORTU	
0	83	79	78		74	77	78	79	79	0
1	82	77	79		80	78	78	78	79	0
2	82	78	78		74	73	78	77	80	0
3	82	78	81		80	78	79	78	79	0
4	84	84	74		78	80	80	80	76	0
...
114	86	81	82		80	82	79	80	81	0
115	75	70	74		73	75	78	79	79	0
116	78	70	74		73	75	75	77	79	0
117	78	76	75		73	75	77	79	77	0
118	82	77	78		77	77	78	79	79	0

119 rows × 9 columns

Gambar 10. Tampilan data *training* baru

	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAT_TES	PERSETUJUAN_ORTU	
0	78	74	78		78	73	78	78	77	0
1	78	74	74		79	73	77	78	78	0
2	78	77	75		78	73	77	79	78	0
3	78	75	76		77	74	77	77	78	0
4	78	75	76		78	73	78	79	76	0
...
86	78	80	78		78	73	78	77	76	0
87	81	75	76		78	77	77	79	78	0
88	78	75	75		78	76	79	79	79	0
89	78	76	74		78	73	78	77	80	0
90	78	74	78		77	77	78	76	78	0

91 rows × 9 columns

Gambar 11. Tampilan data *testing* baru

Dari hasil *preprocessing* data yang dilakukan jumlah *dataset* sebanyak 119 data *train* dan 91 data *test* sudah bersih dan semua data memiliki nilai atau bermakna dan data yang dihasilkan dari hasil *preprocessing* siap untuk digunakan dalam proses klasifikasi.

3.2.4 Data *Training*

Sebelum melakukan tahapan Naïve bayes, data *Training* yang telah bersih dilakukan analisis menggunakan model atau Naïve bayes agar dapat mempelajari pola atau struktur dari data yang digunakan dan data *training* tersebut dikelompokkan/diklasifikasi ke dalam jurusan berdasarkan data-data tersebut. Analisis data *training* dapat ditampilkan dengan coding seperti gambar berikut :

```
train_data = pd.read_excel('datasiswa.xlsx')
x_train = train_data.drop(['NAMA SISWA', 'JURUSAN'], axis=1)
y_train = train_data['JURUSAN']
```

Gambar 12. Coding untuk menampilkan data *training*

Setelah kode pada gambar 12 dijalankan maka akan muncul tampilan data *training* seperti gambar berikut :

	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAI_TES	PERSETUJUAN_ORTU
0	83	79	78	74	77	78	79	79	0
1	82	77	79	80	78	78	78	79	0
2	82	78	78	74	73	78	77	80	0
3	82	78	81	80	78	79	78	79	0
4	84	84	74	78	80	80	80	76	0
..
114	86	81	82	80	82	79	80	81	0
115	75	70	74	73	75	78	79	79	0
116	78	70	74	73	75	75	77	79	0
117	78	76	75	73	75	77	79	77	0
118	82	77	78	77	77	78	79	79	0

[119 rows x 9 columns]

Gambar 13. Tampilan data training

3.2.5 Penerapan Naïve Bayes

Tahap selanjutnya setelah training data yaitu proses Naïve bayes untuk analisis training data. Berikut ini coding naïve bayes sebagai berikut :

```
model = GaussianNB()
model.fit(x_train, y_train)
```

Gambar 14. Coding naïve bayes

Setelah model *naïve bayes* berhasil untuk digunakan menganalisis data pelatihan. Tahap selanjutnya menerapkan model *naïve bayes* untuk memprediksi data *training* *x_train*. Berikut ini *coding* memprediksi data *training* *x_train* sebagai berikut:

```
y_predict = model.predict(x_train)
y_predict
```

Gambar 15. Coding untuk prediksi data *x_train*

Setelah *coding* pada gambar 15. dijalankan maka akan muncul *output* tampilan dari hasil klasifikasi sebagai berikut :

```
array(['IPS', 'BAHASA', 'IPS', 'BAHASA', 'IPA', 'IPA', 'BAHASA',
       'IPS', 'IPS', 'IPS', 'IPS', 'IPA', 'IPA', 'IPA',
       'BAHASA', 'IPA', 'IPA', 'IPS', 'IPS', 'IPA', 'IPA', 'IPA',
       'IPA', 'IPS', 'IPA', 'IPS', 'IPS', 'IPS', 'IPS', 'IPS',
       'IPS', 'BAHASA', 'IPA', 'BAHASA', 'IPS', 'IPS', 'BAHASA', 'IPS',
       'IPA', 'IPA', 'IPS', 'BAHASA', 'IPS', 'IPA', 'IPS', 'IPA',
       'IPA', 'IPS', 'IPS', 'IPS', 'IPS', 'IPA', 'IPS', 'IPS',
       'IPS', 'IPS', 'IPS', 'IPS', 'IPS', 'IPS', 'IPS', 'IPS'],
      dtype='|<U6')
```

Gambar 16. Output tampilan prediksi data *x_train*

Dari gambar 16. menunjukkan hasil prediksi *x_train* dari 119 baris data siswa menggunakan *GaussianNB*.

3.2.6 Model

Setelah menerapkan metode *naïve bayes* maka langkah selanjutnya yaitu gunakan model tersebut untuk mengklasifikasi jurusan pada data *testing* siswa yang belum memiliki jurusan. Berikut ini adalah *coding* model yang digunakan untuk mengklasifikasi jurusan pada data uji siswa.

```
model = GaussianNB()  
model.fit(x_train, y_train)
```

Gambar 17. *Coding model*

3.2.6 Klasifikasi Jurusan Pada Data *Testing*

Setelah model *naïve bayes* berhasil diterapkan maka langkah selanjutnya adalah lakukan proses klasifikasi data *testing* yang disimpan pada variabel *x_test*. Berikut ini *coding* klasifikasi data *testing* (*x_test*) sebagai berikut :

```
y_predict = model.predict(x_train)  
y_predict
```

Gambar 18. Coding klasifikasi x_{-test}

Pada gambar 18 menunjukkan bahwa akan dilakukan klasifikasi pada x_{test} yang disimpan dalam variabel y_{ped} . Setelah *coding* pada gambar 18 dijalankan maka akan muncul *output* seperti pada gambar 19 berikut :

Gambar 19. *Output* tampilan klasifikasi x_{-test}

3.2.7 Hasil Klasifikasi

Setelah melakukan klasifikasi pada data *testing* maka akan diperoleh hasil dari klasifikasi tersebut. Berikut ini *coding* untuk menampilkan hasil klasifikasi data *testing* sebagai berikut :

```
x_test['predict'] = y_pred
```

Gambar 20. Coding untuk menampilkan hasil klasifikasi x_{test}

Setelah *coding* pada gambar 20 berhasil dijalankan maka langkah selanjutnya adalah *coding* untuk menampilkan hasil klasifikasi *y_pred* yang disimpan dalam variabel *x test['predict']*.

x_test

Gambar 21. Coding untuk menampilkan x_{-test}

Setelah *coding* pada gambar 21 dijalankan maka akan muncul *output* seperti pada gambar 21 berikut :

	AGAMA	PKN	BINDO	BINGGRIS	MAT	IPA	IPS	NILAI_TES	PERSETUJUAN_ORTU	predict
0	78	74	78		78	73	78	78	77	0 IPS
1	78	74	74		79	73	77	78	78	0 IPS
2	78	77	75		78	73	77	79	78	0 IPS
3	78	75	76		77	74	77	77	78	0 IPS
4	78	75	76		78	73	78	79	76	0 IPS
...
86	78	80	78		78	73	78	77	76	0 IPS
87	81	75	76		78	77	77	79	78	0 IPS
88	78	75	75		78	76	79	79	79	0 IPS
89	78	76	74		78	73	78	77	80	0 IPS
90	78	74	78		77	77	78	76	78	0 BAHASA

91 rows x 10 columns

Gambar 22. Output tampilan *x_train*

Gambar 22 menunjukkan siswa sudah diklasifikasi kedalam jurusan masing-masing berdasarkan kriteria-kriteria yang digunakan. Setelah data siswa diklasifikasi kedalam jurusan langkah selanjutnya yaitu menampilkan hasil klasifikasi dalam bentuk plot. Berikut *coding* yang digunakan untuk menampilkan data setelah diklasifikasi.

```
import matplotlib.pyplot as plt

# Data yang akan divisualisasikan
categories = ['IPA', 'BAHASA', 'IPS']
values = [1, '14', '76']

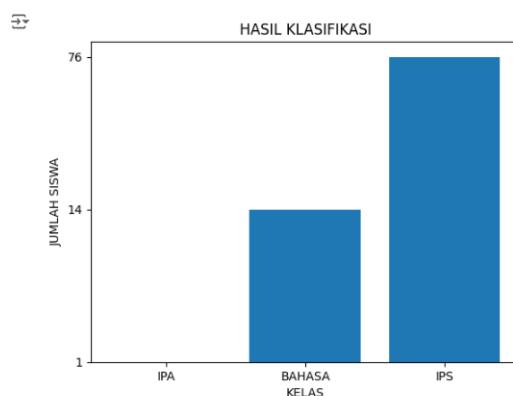
# Membuat diagram batang
plt.bar(categories, values)

# Menambahkan judul dan label
plt.title('HASIL KLASIFIKASI')
plt.xlabel('KELAS')
plt.ylabel('JUMLAH SISWA')

# Menampilkan diagram
plt.show()
```

Gambar 23. Coding untuk menampilkan plot hasil klasifikasi

Setelah *coding* pada gambar 23 dijalankan maka akan tampil *output* plot hasil klasifikasi seperti gambar berikut :



Gambar 24. Output setelah melakukan klasifikasi

Berdasarkan gambar 24 menunjukkan bahwa data *testing* yang sudah diklasifikasi dalam 3 jurusan yaitu : IPA = 1 siswa, BAHASA = 14 siswa dan IPS = 76 siswa. Setelah itu hitung nilai akurasi dari data *training* yang sudah diuji menggunakan metode *naïve bayes*. Berikut ini merupakan *coding* untuk menghitung nilai akurasi sebagai berikut :

```
akurasi = classification_report(y_train, y_predict)
print(akurasi)
```

Gambar 25. Coding untuk menghitung nilai akurasi

Setelah *coding* pada gambar 25 dijalankan maka akan tampil *output* tampilan nilai akurasi seperti gambar berikut :

	precision	recall	f1-score	support
BAHASA	0.50	0.26	0.34	19
IPA	0.48	0.48	0.48	29
IPS	0.72	0.82	0.77	71
accuracy			0.65	119
macro avg	0.57	0.52	0.53	119
weighted avg	0.63	0.65	0.63	119

Gambar 26. OutpuBt tampilan hitung nilai akurasi

Berdasarkan gambar 26 diatas menunjukkan bahwa nilai akurasi dari data *training* setelah dianalisis sebesar : 0,65 % dari 119 siswa.

4. KESIMPULAN

Hasil mengklasifikasi jurusan siswa di SMAN PANTURA menggunakan metode *Naïve Bayes* terdapat 3 jurusan yaitu IPA, BAHASA dan IPS. Dari jumlah data yang digunakan sebanyak 119 data *training* dan 91 data *testing*. Setelah dianalisis menggunakan *python* Pada data *training* terdapat 3 jurusan yaitu : IPA=29 siswa, BAHASA=10 siswa, dan IPS=80 siswa dengan Nilai akurasi yang diperoleh dari data *training* yaitu 65%. Sedangkan pada data *testing* diklasifikasi dalam 3 jurusan yaitu : IPA = 1 siswa, BAHASA = 14 siswa dan IPS = 76 siswa. Dari hasil klasifikasi yang diperoleh dapat disimpulkan bahwa jika data yang digunakan berjumlah banyak maka tingkat nilai akurasi atau kualitas dari hasil klasifikasi semakin baik dan sarannya adalah Diharapkan menggunakan metode lain seperti metode *K-Means Clustering*, dan *Decision Tree* untuk klasifikasi dan mendapatkan metode yang lebih bagus, dari metode yang digunakan saat ini untuk mengklasifikasi jurusan siswa dan penelitian selanjutnya diharapkan, klasifikasi jurusan siswa menggunakan bahasa pemrograman *python*, *google colabulatory*, dapat dikembangkan berbasis web.

Daftar Pustaka

- [1] H. L. Sari and I. Y. Beti, “Penerapan Data Mining Dalam Pengelompokan Buku Yang Dipinjam Menggunakan Algoritma K-Means,” vol. 3, no. 6, pp. 925–933, 2023, doi: 10.30865/klik.v3i6.826.
- [2] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” vol. 3, pp. 46–56, 2024.
- [3] Y. Apridiansyah, N. D. M. Veronika, and E. D. Putra, “Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes,” *JSAI (Journal Scientific and Applied Informatics)*, vol. 4, no. 2. pp. 236–247, 2021. doi: 10.36085/jsai.v4i2.1701.
- [4] E. Mutiara, L. S. Ramdhani, R. Wajhillah, J. M. Hudin, and A. Gunawan, “Penerapan Metode Naïve Bayes Dengan PSO Untuk Pemilihan Peminatan Jurusan Pada SMK,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 8, no. 1. 2023. doi: 10.31294/ijcit.v8i1.15732.
- [5] Y. Angraini, S. Fauziah, and J. L. Putra, “Analisis Kinerja Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Keberhasilan Sekolah Menghadapi Un,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 5, no. 2. pp. 285–290, 2020. doi: 10.33480/jitk.v5i2.1233.
- [6] J. Fakhri, A. S. Sunge, and A. Turmudi zy, “Perancangan Klasifikasi Algoritma Naive Bayes Pada Data Pemilihan Jurusan Siswa,” *JTT (Jurnal Teknologi Terpadu)*, vol. 11, no. 2. pp. 260–

- 269, 2023. doi: 10.32487/jtt.v11i2.1823.
- [7] Y. P. K. Kelen, " Sistem Pendukung Keputusan : Konsep, Metode dan Implementasi" .
- [8] S. Triyanto, A. Sunyoto, and M. R. Arief, "Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 5, no. 2. pp. 109–117, 2021. doi: 10.35145/joisie.v5i2.1785.
- [9] A. Z. Mafakhir and A. Solichin, "Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta," *Fountain of Informatics Journal*, vol. 5, no. 1. p. 21, 2020. doi: 10.21111/fij.v5i1.4007.
- [10] M. E. Bere, K. Kelen, H. H. Ullu, and B. Baso, "Implementasi Algoritma Naive Bayes Classifier Terhadap Analisis Sentimen Kondisi Stunting di Indonesia Pada Media Sosial X," vol. 7, no. 4, pp. 598–605, 2024.
- [11] Adawiyah Ritonga and Yahfizham Yahfizham, "Studi Literatur Perbandingan Bahasa Pemrograman C++ dan Bahasa Pemrograman Python pada Algoritma Pemrograman," *J. Tek. Inform. dan Teknol. Inf.*, vol. 3, no. 3, pp. 56–63, 2023, doi: 10.55606/jutiti.v3i3.2863.
- [12] G. I. E. Soen, M. Marlina, and R. Renny, "Implementasi Cloud Computing dengan Google Colaboratory pada Aplikasi Pengolah Data Zoom Participants," *JITU: Journal Informatic Technology And Communication*, vol. 6, no. 1. pp. 24–30, 2022. doi: 10.36596/jitu.v6i1.781.



ZONAsi: Jurnal Sistem Informasi
Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)