

Implementasi *Data Mining* dengan Metode *Naive Bayes* Untuk Memprediksi Penerimaan Siswa Baru di MTS NU Islamiyah Asembagus

Firman Santoso¹, Sunardi², Hifni Zainul Lukman^{3✉}

^{1, 2, 3} Teknologi Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 15-08-2023

Direvisi : 27-08-2023

Diterima : 30-08-2023

Kata Kunci:

Data Mining, *Naive Bayes*,
Prediksi.

Keywords :

Data Mining, Naive Bayes,
predictions.

Corresponding Author :

Hifni Zainul Lukman

Teknologi Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia

Jl. KHR. Syamsul Arifin No. 1-2, Sukorejo, Sumberejo, Kec. Banyuputih, Kabupaten Situbondo, Jawa Timur 68374

Email: hifnizainul@gmail.com

ABSTRAK

Penerimaan siswa baru adalah langkah awal yang signifikan dalam memasuki dunia pendidikan. Salah satunya di MTs NU Islamiyah, sebuah lembaga pendidikan di Desa Trigonco, Kecamatan Asembagus. Meski mampu menerima siswa baru setiap tahunnya, fluktuasi jumlah siswa yang masuk dan tidak menjadi tantangan bagi lembaga ini. Fenomena ini mempengaruhi perencanaan kebijakan di masa mendatang, termasuk persiapan sarana dan prasarana yang optimal. Penelitian ini bertujuan membangun sistem prediksi penerimaan siswa baru di MTS NU Islamiyah dengan metode Naive Bayes menggunakan Python. Tujuannya ialah meningkatkan akurasi prediksi jumlah siswa baru yang akan masuk setiap tahun. Data siswa yang digunakan sebanyak 623 data training dan 82 data testing digunakan sebagai dasar prediksi. Metode Naive Bayes digunakan untuk mengklasifikasikan data dengan menghitung probabilitas berdasarkan data historis. Hasil prediksi yang diperoleh bahwa prediksi tidak masuk class precision sebesar 0% dan Support 11 model memiliki performa tidak baik dalam memprediksi class tidak masuk. Sedangkan untuk prediksi yang masuk class precision sebesar 87% dan support 71. model memiliki performa yang baik dalam memprediksi kelas yang masuk.

ABSTRACT

Acceptance of new students is an important first step in entering the world of education. One of them is at MTs NU Islamiyah, an educational institution in Trigonco Village, Asembagus District. Even though it is able to accept new students every year, fluctuations in the number of incoming students are not a challenge for this institution. This phenomenon influences future policy planning, including the optimal preparation of facilities and infrastructure. This study aims to build a prediction system for new student admissions at MTS NU Islamiyah using the Naive Bayes method using Python. The aim is to improve the prediction accuracy of the number of new students who enter each year. The student data used is 623 training data and 82 test data used as the basis for predictions. The Naive Bayes method is used to classify data by calculating probabilities based on historical data. The prediction results obtained are predictions not included in the 0% precision class and the Support 11 model has a poor performance in predicting the excluded class. As for predictions that are included in the class, the precision is 87% and the support is 71. The model has good performance in predicting the class that is entered.

PENDAHULUAN

Perkembangan pesat dalam teknologi informasi telah mengubah cara berbagai perusahaan, lembaga, dan bahkan bidang pendidikan beroperasi. di mana teknologi informasi telah memberikan dampak positif pada berbagai sektor, salah satunya kegiatan penerimaan siswa baru yang dilakukan pada setiap tahunnya (Teknologi et al., 2020).

Penerimaan siswa baru merupakan proses pendaftaran kepada siswa baru untuk masuk sekolah dan gerbang awal yang dilalui peserta didik didalam objek Pendidikan (Irawan & Damanik, 2022) salah satunya di lembaga Pendidikan MTS NU Islamiyah yang terletak di Desa Trigonco Kecamatan Asembagus.

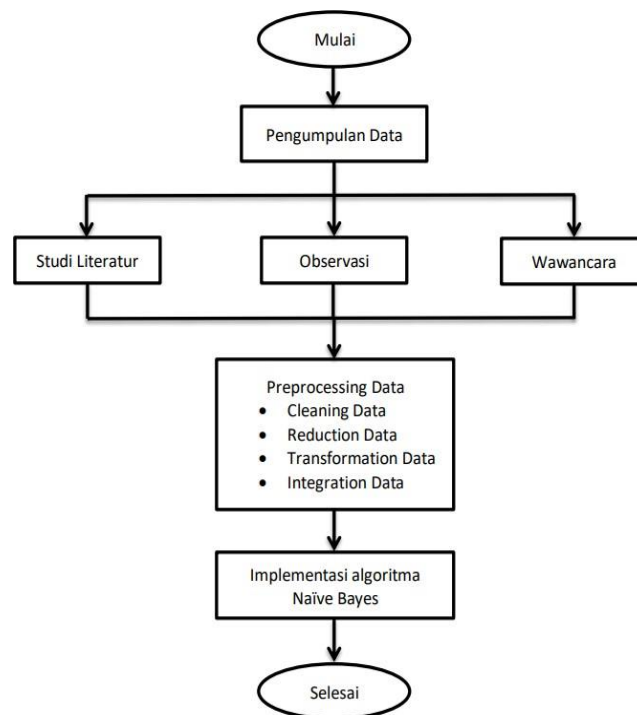
MTS NU Islamiyah pada setiap tahunnya mampu menerima siswa siswi baru. Kegiatan penerimaan siswa siswi baru ini merupakan titik awal untuk menentukan kelancaran tugas suatu sekolah, dengan dibantu tenaga pengajar serta dilengkapi sarana dan prasarana yang optimal dalam kegiatan belajar mengajar, menghasilkan siswa siswi yang terampil dan ber wawasan luas (Irawan & Damanik, 2022). Akan tetapi Pihak Lembaga memiliki kendala dalam penerimaan siswa baru karena jumlah siswa yang mendaftar pada pihak Lembaga fluktuasi setiap tahunnya.

Untuk mengatasi hal tersebut diperlukan suatu teknik dalam data mining yang dapat mengubah sesuatu yang tidak bermanfaat menjadi berharga dan dapat memberikan suatu informasi yang penting (Fahreza, 2022) dengan menggunakan metode *Naive Bayes*. Digunakan sebagai metode prediksi penerimaan siswa baru di MTS NU Islamiyah. Ada 5 peran data mining yaitu estimasi, prediksi, kalasifikasi, clustering, asosiasi (Maruli Sitompul et al., 2021).

Naive Bayes adalah metode yang dikembangkan oleh ilmuan inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes.(Imandasari et al., 2019) Oleh karena itu kriteria yang diperlukan untuk penerimaan siswa baru di MTS NU Islamiyah dengan menggumpulkan data siswa baru di tahun sebelumnya data ini akan di Implementasikan kedalam system Pemrograman dengan menggunakan bahasa Python untuk menghasilkan keterangan hasil prediksi. Dengan menerapkan metode *Naive Bayes* sebagai prediksi untuk memperkirakan jumlah siswa yang kemungkinan akan masuk dan juga dapat memberikan informasi yang lebih akurat dalam prediksi penerimaan siswa baru agar segala sesuatu dapat dipersiapkan oleh pihak sekolah.

METODE PENELITIAN

Pada tahapan metode penelitian terdapat beberapa langkah-langkah dalam metodem penelitia. Langkah awal yang dilaksanakan adalah pengumpulan data. Langkah Berikutnya proses pengolahan data dan tahapan terakhir yaitu tahapan pengujian (Informasi et al., 2022). Metode yang digunakan pada penelitian adalah metode penlitian deskriptif yaitu penelitian yang dilakukan terhadap variabel yang data-datanya belum ada sehingga perlu dilakukan proses manipulasi melalui pemberian treatment/perlakuan tertentu terhadap subjek penelitian yang kemudian diamati/diukur dampaknya (data yang akan dating) (Jaedun, 2011). Gambaran terlampir pada Tahapan Penelitian pada Gambar 1.



Gambar 1. Kerangka Penelitian

Pengumpulan Data

Teknik pengumpulan data merupakan teknik yang digunakan dalam mengumpulkan suatu informasi terkait yang dibutuhkan dalam melakukan penelitian. Adapun teknik pengumpulan data yang digunakan dalam penelitian ini (Irawan & Damanik, 2022) yaitu:

a. Observasi

Metode ini dilakukan dengan secara langsung pada objek yang akan diteliti (Senika & Iskandar, 2022) di MTS NU Islamiyah secara langsung untuk mendapatkan data siswa baru di tahun sebelumnya dari yang dari data yang valid yang dibutuhkan oleh peneliti.

b. Wawancara

Wawancara dilakukan kepada kepala sekolah terkait dalam penerimaan siswa baru untuk menggali informasi tentang masalah yang terjadi dengan penerimaan siswa baru yang akan masuk.

Analisis

Tahap ini berupa tahap seleksi data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Melakukan seleksi dengan memperhatikan konsistensi data, missing value dan redundant data (Sinaga et al., 2022). Noeng Muhadjir (1998: 104) mengemukakan pengertian analisis data sebagai “upaya mencari dan menata secara sistematis catatan hasil observasi, wawancara, dan lainnya untuk meningkatkan pemahaman peneliti tentang kasus yang diteliti dan menyajikannya sebagai temuan bagi orang lain. Sedangkan untuk meningkatkan pemahaman tersebut analisis perlu dilanjutkan dengan berupaya mencari makna.” (Rijali, 2019). Dalam konteks penelitian menggunakan algoritma Naive yang akan digunakan dalam proses pelatihan dan pengujian algoritma.

Implementasi

Pada tahapan ini proses implementasi metode naïve bayes kedalam Bahasa programan python.

HASIL DAN PEMBAHASAN

Pada bab ini penelitian akan dijelaskan hasil dari penelitian yang telah dilakukan mengenai pengimplementasin data mining untuk memprediksi Penerimaan Siswa Baru di MTS NU Isalmiyah dengan menggunakan metode Naïve Bayes melalui observasi terhadap objek penelitian di lingkungan dan wawancara kepada kepala sekolah untuk pengumpulan data yang didapat sebagai berikut:

Hasil Observasi

Observasi terhadap objek yang dilakukan pada bulan juli di MTS NU Islamiyah Asembagus yang bertujuan untuk memberikan gambaran nyata pada proses pengumpulan data Siswa baru di tahun sebelumnya. Dari observasi tersebut juga peneliti ingin mendapatkan beberapa data yang dibutuhkan untuk penelitian ini, sehingga dapat mengolah data antara usia, jenis kelamin, asal sekolah, Penghasilan Ortu, dan alamat.

Hasil Wawancara

Wawancara disini dilakukan secara langsung oleh peneliti kepada Ka TU sebagai selaku yang memegang data dari siswa baru di lembaga tersebut bertempat di kantor lembaga tersebut dari beberapa wawancara hasilnya salah satunya:

- Siswa yang akan masuk pada tiap taunnya fluktuatif dan menjadi kendala bagi pihak sekolah mempersiapkan sarana dan prasarana.
- menyimpan banyak data dalam berbagai bentuk, baik hard file maupun soft file, tetapi tidak terlalu paham tentang teknologi informasi (IT) untuk mengelola data tersebut.

Analisis Data

Berdasarkan hasil penelitian yang sudah dilakukan didapat hasil dari setiap tahapan. Adapun hasil dari setiap tahapan tahapan yang sudah dilakukan adalah pengumpulan dataset, preprocessing data, implementasi naïve bayes, dan hasil uji coba yang sudah dilakukan dari metode tersebut.

Pengumpulan Dataset

Dari hasil pengumpulan dataset yang dilakukan oleh peneliti di MTS NU Islamiyah ini mendapatkan dataset yang didapat dari bagian KA TU lembaga tersebut. Data yang dipakai dalam penelitian ini adalah data dalam 10 tahun, Data yang telah didapat meliputi fitur yang melekat pada siswa baru tersebut, dan status dalam bentuk dokumen excel, dan akan diolah di google colaboratory dengan bahasa python yang nantinya akan di proses satu persatu. Adapun dataset seperti terlampir dalam table di bawah Tabel data set:

Tabel 1. Tabel Dataset

No	Nama	Jenis Kelamin	Usia	Tempat Lahir	Status Keluarga	Tahun	Alamat	Status
1	Nando Darmawan	L	12	Situbondo	Mampu	2014	Asembagus	Masuk
2	Oktavia Eka	P	11	Situbondo	Mampu	2014	Asembagus	Masuk

3	Putra Firmansyah	L	13	Situbondo	Tidak Mampu	2014	Jangkar	Tidak Masuk
4	Qoriyah Gita	P	13	Situbondo	Tidak Mampu	2014	Asembagus	Tidak Masuk
5	Rafi haris				Mampu			Masuk
6	Sari	L	13	Bondowoso	Mampu	2014	Asembagus	Masuk
7	Indriani Adi Jaya	P	13	Situbondo	Mampu	2014	Jangkar	Tidak Masuk
8	Candra Laksana	L	13	Situbondo	Tidak mampu	2014	Asembagus	Masuk
9	Dito Mandiri							Masuk
10	Ella Nanda	L	13	Situbondo	Mampu	2014	Jangkar	Masuk
701	----- Ananda I.K	P	13	Situbondo	----- Mampu	2014	Asembagus	----- -
702	Aufa Aulia	L	14	Situondo	Mampu	2023	Asembagus	?
703	N.A Bagus Dwiyanto	P	14	Situbondo	Tidak Mampu	2023	Asembagus	?
		L	14	Situbondo	Mampu	2023	Jangkar	?

Preprocessing Data

Data Preprocessing bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya. Tahapan ini dilakukan pada data mentah untuk menghilangkan data yang bermasalah atau inkonsisten. Data yang bermasalah seperti data yang noise atau yang mengandung error (Alghifari & Juardi, 2021). Preprocessing data adalah langkah pertama dalam data mining untuk mengubah raw data (bentuk mentah) ke dalam data yang lebih mudah dipahami. Beberapa langkah yang dilakukan pada saat preprocessing data yaitu tokenisasi yaitu proses pemisahan suatu rangkaian karakter berdasarkan pada karakter spasi yang mungkin saja terjadi penghapusan karakter tertentu. (Jasri et al., 2022)

Tabel 2. Aturan Asosiasi

Alamat	Jarak
Asembagus	< 1 KM
Jangkar	> 1 KM

Implementasi Naïve Bayes

Setelah tahap information preprocessing di atas dilakukan, selanjutnya informasi yang telah disiapkan dieksekusi untuk pemeriksaan dengan menggunakan teknik Naïve Bayes. Naïve Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes (Senika & Iskandar, 2022) yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian seperti (1). Rumus Naïve Bayes (Suyanto, 2019).

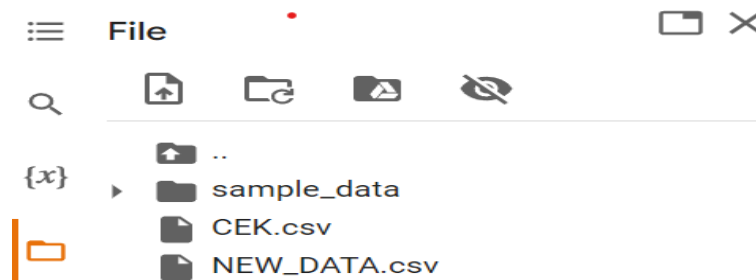
$$P(H|X) = \frac{P(H) \cdot P(X)}{P(X)} \quad (\text{Suntoro \& Kom, n.d.})$$

Keterangan :

X : data dengan class yang belum diketahui
 H : hipotesis data X merupakan suatu class spesifik
 $P(H|X)$: probabilitas hipotesis H berdasar kondisi (posteriori probability)
 $P(H)$: probabilitas hipotesis H (prior probability)
 $P(X|H)$: probabilitas X berdasar kondisi pada hipotesis H
 $P(X)$: probabilitas dari X

Perhitungan Menggunakan Google Colaboratory

Langkah pertama dalam proses perhitungan Naïve Bayes yaitu mengunggah dataset yang sudah di siapkan sebelumnya oleh peneliti ke dalam Google Colaboratory.



Gambar 1. Import Dataset

Setelah proses pengunggahan dataset selesai, lalu dilanjutkan dengan proses import library apa saja yang akan digunakan saat proses analisis yang kita lakukan seperti pada Segmen Program 1 dibawah ini.

Segemen Program 1. Import Library

```

1. #importing the libraries
2. import numpy as np
3. import pandas as pd
4. from sklearn.naive_bayes import GaussianNB
5. from sklearn.model_selection import train_test_split
6. from sklearn import metrics
7. import joblib
8. import time

```

setelah proses Import Library selesai yaitu memanggil dataset di google colaboratory dengan keseluruhan kode tersebut bertujuan untuk membaca data dari file Excel 'Cek.xlsx' ke dalam DataFrame dan menampilkan 10 baris pertama dari DataFrame tersebut. segemen Program 2 dan Gambar 2 dibawah ini:

Segmen program 2. Membaca Data Excel

```

1. # Membaca data dari file excel ke dalam DataFrame
2. nama_file_xlsx = 'Cek.xlsx'
3. data = pd.read_excel(nama_file_xlsx)
4.
5. data.head(10)

```


	No	Nama	JENIS KELAMIN	Usia	Tempat Lahir	Tgl Lahir	Tahun	JARAK	STATUS KELUARGA	STATUS
0	1	NANDO DARMAWAN	L	13	SITUBONDO	2001-06-07 00:00:00	2014	< 1 KM	MAMPU	MASUK
1	2	OKTAVIA EKA	L	12	SITUBONDO	2002-06-26 00:00:00	2014	> 1 KM	TIDAK MAMPU	TIDAK MASUK
2	3	PUTRA FIRMANSYAH	P	13	SITUBONDO	2001-12-15 00:00:00	2014	< 1 KM	MAMPU	MASUK
3	4	QORIYAH GITA	P	13	SUKABUMI	2001-12-23 00:00:00	2014	< 1 KM	MAMPU	MASUK
4	5	RAFI HARIS	L	13	SITUBONDO	2001-08-28 00:00:00	2014	< 1 KM	MAMPU	MASUK
5	6	SARI INDRIANI	P	12	SITUBONDO	2002-05-14 00:00:00	2014	< 1 KM	MAMPU	MASUK
6	7	ADI JAYA	L	12	SITUBONDO	2002-01-08 00:00:00	2014	< 1 KM	MAMPU	MASUK
7	8	BELLA KURNIAWATI	P	13	SITUBONDO	2001-05-05 00:00:00	2014	< 1 KM	MAMPU	MASUK
8	9	CANDRA LAKSANA	L	12	SITUBONDO	2002-05-30 00:00:00	2014	< 1 KM	TIDAK MAMPU	MASUK
9	10	DITO MANDIRI	L	13	BONDOWOSO	2001-10-24 00:00:00	2014	< 1 KM	MAMPU	MASUK

Gambar 2. Dataset

Setelah cara yang paling umum memanggil dataset berikut, untuk lebih spesifik Setelah mengetahui kualitas setiap bagian, untuk setiap properti tidak semua akan dieksekusi dalam perhitungan Guileless Bayes sehingga penting untuk mengelola atau membuang segmen yang tidak diperlukan dalam proses Gullible Bayes nanti. Untuk bagian yang tidak digunakan, lebih spesifik yaitu No Nama, Usia, Tempat Lahir, Tgl Lahir dan Tahun seperti Segmen Program 3 dan gambar 3 dibawah ini:

Segmen Program 3. Drop Kolom yang Tidak Diperlukan

```

1. def drop_data():
2.     print("Proses Drop Kolom yang Tidak Diperlukan")
3.
4.     global data
5.     data = data.drop(columns=['No', 'Nama', 'Usia', '
    Tempat Lahir', 'Tgl Lahir', 'Tahun'])
6.     time.sleep(10)
7.     print("Drop Berhasil")
8.     time.sleep(5)
9.     print("Proses Selesai")
10. drop_data()

```

```

➤ Proses Drop Kolom yang Tidak Diperlukan
  Drop Berhasil
  Proses Selesai

```

Gambar 3. Proses Drop Kolom


Setelah tahap drop kolom selesai Pada tahap ini dilakukan pembersihan dataset (Eka et al., 2020) yang diharapkan dapat mengubah informasi objek menjadi numerik karena pada data subject tidak bisa dilakukan perhitungan menggunakan google colaboratory sehingga perlunya pengubahan ke numeric, seperti tahapan cleaning pada Segmen Program 4 dan Gambar 4 di bawah ini:

Segmen Program 4. Mengubah Katagorikal Menjadi Numerik

```

1. data_c = data.astype('category')
2. #ubah kategori menjadi numerik
3. data_c["Jenis Kelamin"] = data_c["Jenis Kelamin"].map(
    {'Laki - Laki': 1, 'Perempuan': 0})
4. data_c["Jarak"] = data_c["Jarak"].map({'< 1 Km': 1, '>
    1 Km': 0})
5. data_c["STATUS KELUARGA"] = data_c["STATUS KELUARGA"]
    .map({'Mampu': 1, 'Tidak Mampu': 0})
6. data_c["STATUS"] = data_c["STATUS"].map({'Masuk': 1, '
    Tidak Masuk': 0})
7. #data_c.tail(50) # Menampilkan Data
8. data_subset = data_c.iloc[1:10]
9.
10. # Menampilkan data yang dipanggil
11. print(data_subset)

```




	Jenis Kelamin	STATUS KELUARGA	Jarak	STATUS
1	0	1	1	1
2	0	1	1	0
3	0	0	1	1
4	1	1	1	0
5	0	0	1	0
6	1	1	1	0
7	0	1	1	1
8	1	1	0	1
9	1	0	0	0

Gambar 4. Hasil Dari Cleaning Data

Setelah proses Cleaning data selesai yaitu mengambil semua baris dan tiga kolom pertama dari DataFrame Dengan kata lain, ini mengambil tiga kolom pertama sebagai fitur (X), yang nantinya akan digunakan sebagai input untuk model, dan mengambil semua baris dari kolom keempat dari DataFrame dan menyimpannya dalam variabel Y. Ini mengasumsikan bahwa kolom keempat berisi label atau target yang ingin prediksi. Seperti Segmen Program 5 dan Gambar 5 dibawah ini:

Segmen Program 5. Memisahkan Lebel dan Fitur

```
1. #memisahkan data label dan fitur
2. X = data_c.iloc[:, :3].values
3. Y = data_c.iloc[:, 3].values
4.
5. Y = data_c.iloc[:, 3].values
6. Y
```



```
[1, 1, 0, 1, 0, ..., 1, 0, 1, 1, 1]
Length: 710
Categories (2, int64): [1, 0]
```

Gambar 5. Hasil Dari Pemisahan Lebel

Pada tahap ini yaitu membagi dataset menjadi set pelatihan (untuk melatih model) dan set pengujian (untuk menguji kinerja model) dalam proporsi yang telah ditentukan. Pemisahan ini penting untuk menghindari overfitting dan untuk mengukur kemampuan model dalam melakukan prediksi pada data yang belum pernah dilihat sebelumnya. Seperti Segmen Program 6 dan dibawah ini:

Segmen Program 6. Membagi Dataset Pelatih dan Uji

```
1. #membagi dataset menjadi set pelatihan dan set tes
2. X_train, X_test, y_train, y_test = train_test_split(X,
    Y, test_size = 0.2, random_state = 710)
3. print(X_train.shape, y_train.shape)
```

Pada tahap ini membuat dan melatih model klasifikasi Gaussian Naive Bayes menggunakan data pelatihan X_train dan y_train. Model ini akan mengandalkan distribusi Gaussian untuk membuat prediksi berdasarkan probabilitas fitur-fitur yang diberikan. Setelah model dilatih, Anda dapat menggunakannya untuk memprediksi label pada data baru atau data pengujian seperti Segmen Program 7 dibawah ini:

Segmen Program 7. Melatih Model Klasifikasi GaussianNB

```
1. #membuat pengklasifikasi Gaussian
2. model = GaussianNB()
3. #melatih model dengan data pelatihan
4. model.fit(X_train, y_train)
```


pada tahap ini menyimpan model klasifikasi Gaussian Naive Bayes yang telah dilatih ke dalam sebuah file menggunakan pustaka joblib. Seperti Segmen Program 8 dibawah ini:

Segmen Program 8. Menyimpan Model Yang sudah di Latih

```
1. #menyimpan hasil data training dalam joblib
2. model_filename = 'trained_naive_bayes_model.joblib'
3. joblib.dump(model, model_filename)
```

pada tahap ini menggunakan model klasifikasi Gaussian Naive Bayes yang telah dilatih sebelumnya untuk memprediksi respons (label) pada data pengujian X_test, dan untuk mencetak (menampilkan) data pengujian X_test ke layar. Seperti Segmen Program 9 dibawah ini:

Segmen Program 9. Prediksi Respon Untuk Data Uji

```
1. #prediksi respons untuk data uji
2. y_pred = model.predict(X_test)
3. print(X_test)
```

tahap berikutnya menghitung akurasi model. Fungsi ini mengambil dua argumen: y_test yang berisi label sebenarnya dari data pengujian dan y_pred yang berisi label prediksi yang dihasilkan oleh model. Fungsi ini akan menghitung seberapa akurat prediksi model Anda dalam membandingkan dengan label sebenarnya seperti Segmen program 10 di bawah ini:

Segmen Program 10. Menghitung akurasi Model

```
1. # Model accuracy
2. print("Accuracy:", metrics.accuracy_score(y_test, y_pred))
```

tahap berikutnya ini akan memuat kembali model yang telah dilatih sebelumnya menggunakan joblib. kita membaca data baru dari file excel ke dalam DataFrame baru dan siap untuk digunakan dalam proses prediksi menggunakan model yang telah dilatih sebelumnya. seperti Segmen Program 11 dibawah ini:

Segmen Program 11. Membaca File Untuk di Uji

```
1. # Muat model Naive Bayes yang terlatih
2. model_filename = 'trained_naive_bayes_model.joblib'
3. loaded_model = joblib.load(model_filename)
4.
5. nama_file_xlsx_baru = 'Data_Uji1.xlsx'
6. data_baru = pd.read_excel(nama_file_xlsx_baru)
```

setelah melakukan beberapa langkah untuk mengubah kategori dalam DataFrame data_baru menjadi nilai numerik yang dapat digunakan untuk prediksi menggunakan model dan memilih tiga kolom pertama dari DataFrame yang sekarang berisi nilai numerik yang telah diubah. Kolom-kolom ini adalah "Jenis Kelamin", "Jarak", dan "STATUS KELUARGA". Anda mengambil nilai-nilai ini dan menyimpannya dalam array X, yang nantinya akan digunakan sebagai fitur dalam prediksi, seperti Segmen Program 12 di bawah ini:

Segmen Program 12. Mengubah Kategori Menjadi Numerik

```
1. df_ca = data_baru.astype('category')
2. # Convert categories to numeric
3. df_ca["Jenis Kelamin"] = df_ca["Jenis Kelamin"].map({'L': 1, 'P': 0})
4. df_ca["Jarak"] = df_ca["Jarak"].map({'< 1 Km': 1, '> 1 Km': 0})
5. df_ca["STATUS KELUARGA"] = df_ca["STATUS KELUARGA"].map({'Mampu': 1, 'Tidak Mampu': 0})
6.
7. X = df_ca.iloc[:, :3].values
```

Pada tahap ini menggunakan model yang telah dimuat (`loaded_model`) untuk memprediksi kolom "STATUS" berdasarkan fitur-fitur yang ada dalam variabel X dan untuk mencetak (menampilkan) data fitur X, seperti Segmen Program 13 dibawah ini:

Segmen Program 13. Memprediksi Kolom STATUS yang Ada dalam Variable X

```
1. # Gunakan model yang dimuat untuk memprediksi kolom
   m "STATUS" berdasarkan fitur di X
2. y_pred = loaded_model.predict(X)
3.
4. # Print the new data and the predicted values for
   "STATUS"
5. print("Data_Uji:")
6. print(X)
```

Tahap berikutnya menunjukkan nilai prediksi yang dihasilkan oleh model untuk setiap sampel dalam data uji. Karena Anda telah memuat model yang telah dilatih sebelumnya dan menggunakan model tersebut untuk memprediksi kolom "STATUS" berdasarkan fitur-fitur di X. Segmen Program 14 dibawah ini:

Segmen Program 14. Prediksi Status

```
1. print("\nPredicted Status:")
2. print(y_pred)
```

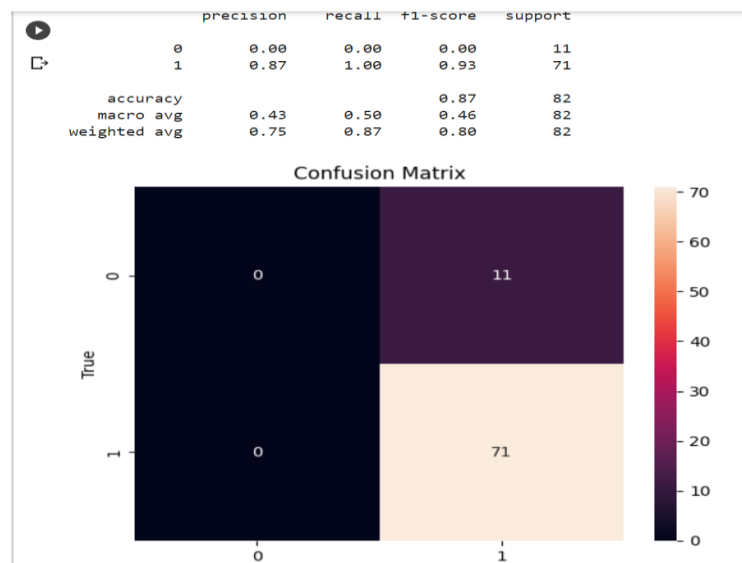
[illegible]

Gambar 7. Hasil Prediksi

Tahap berikutnya memvisualisasikan matriks (*confusion matrix*) Confusion matrix merupakan visualisasi untuk mengevaluasi dari kinerja model klasifikasi (Rizal et al., 2018). menggunakan pustaka Seaborn serta mencetak laporan klasifikasi, dan Dalam laporan klasifikasi, nilai yang disajikan adalah rata-rata presisi, recall, dan f1-score untuk setiap kelas, serta akurasi keseluruhan.

Segmen Program 15. Menvisualisasikan Matriks

```
1. \from sklearn import metrics
2. from sklearn.metrics import classification_report, conf
   usion_matrix
3. import seaborn as sns
4. matrix = confusion_matrix(y_pred, prediction_classifier
   )
5. sns.heatmap(matrix,annot=True, fmt="d")
6. plt.title('Confusion Matrix')
7. plt.xlabel('predicted')
8. plt.ylabel('True')
9. print(classification_report(y_pred, prediction_classifi
   er))
```



Gambar 8. Hasil dari Visualisasi

KESIMPULAN DAN SARAN

Kesimpulan

Dari penelitian ini menggunakan data training sebanyak 628 dan data testing sebanyak 82 dari 623 data siswa yang mendaftar dengan menggunakan 4 atribut yang sama. Berdasarkan hasil uji coba yang sudah dilakukan menggunakan naïve bayes adalah variable satu dengan yang lain memiliki keterkaitan dalam menganalisis siswa yang akan masuk. Jumlah atribut yang digunakan juga mempengaruhi keakurasian terhadap model algoritma yang digunakan. Setelah analisis selesai dilakukan didapat hasil untuk siswa yang masuk dan tidak masuk antara lain. Untuk tidak masuk mendapatkan *precision* 0.00 *recall* sebesar 0.00, *f1-score* sebesar 0.00%, *support* 11. Sedangkan untuk yang Masuk mendapatkan *precision* sebesar 87%, *recall* sebesar 100%, *f1-score* sebesar 93% , dan *support* sebesar 71. Dan mendapatkan tingkat akurasi sebesar 80% yang baik untuk hasil proses analisis.

Saran

dapat menggunakan tambahan variabel yang dapat mempengaruhi tingkat keakurasiannya dari penelitian yang dilakukan. Dapat juga menggunakan metode yang lain, atau memakai 2 metode atau lebih sekaligus sebagai perbandingan dari metode yang digunakan tersebut dan menentukan metode mana yang lebih baik untuk dipakai

UCAPAN TERIMA KASIH

Penulis tidak dapat melakukan penelitian dan penyelesaian jurnal ini tanpa nikmat yang telah Allah berikan kepada kami. Berkat Allah SWT jurnal ini dapat kami selesaikan. Kepada kedua orang tua dan seluruh keluarga di rumah yang selalu mendukung kami baik dari segi material maupun non material. Ucapkan juga kepada dosen pembimbing kami ka prodi dan seluruh civitas akademika fakultas sains & teknologi universitas ibrahimy yang selalu memberikan arahan, membantu persiapan, dan menasehati kami dalam menyelesaikan jurnal ini.

REFERENSI

Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*,

- 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Eka, S., Putri, Y., & Surahman, A. (2020). *Penerapan Model Naive Bayes Untuk Memprediksi Potensi Pendaftaran Siswa Di Smk Taman Siswa Teluk Betung Berbasis Web*. 1(1), 93–99.
- Fahreza, A. (2022). Penerapan Data Mining dengan Metode Single Moving Average dalam Pengolahan Data Penerimaan Siswa Baru. *Proceeding Seminar Nasional Ilmu Komputer*, 2(1), 25–34.
- Imandasari, T., Irawan, E., Perdana Windarto, A., Wanto, A., & Tunas Bangsa Pematangsiantar Jln Jendral Sudirman Blok No, S. A. (2019). Prosiding Seminar Nasional Riset Information Science (SENARIS) Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, September, 750–761.
- Informasi, S., Darwan Ali, U., Batu Berlian, J., & Tengah, K. (2022). Penerapan Metode Naive Bayes Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangsian. *Jurnal JUPITER*, 14(2), 64–70.
- Irawan, E., & Damanik, B. E. (2022). *Teknik Klasifikasi dalam Memprediksi Penerimaan Siswa Baru Menggunakan Metode Naïve Bayes Classification Techniques in Predicting New Student Admission Using the Naïve Bayes Method*. 1(3). <https://doi.org/10.55123/jomlai.v1i3.963>
- Jaedun, A. (2011). Oleh : Amat Jaedun. *Metodologi Penelitian Eksperimen*, 0–12.
- Jasri, M., Wijaya, A., & Sunggara, R. (2022). Penerapan Data Mining untuk Klasifikasi Penyakit Demam Berdarah Dengue (DBD) Dengan Metode Naïve Bayes (Studi Kasus Puskesmas Taman Krocok). *SMARTICS Journal*, 8(1), 35–42. <https://doi.org/10.21067/smartics.v8i1.7375>
- Maruli Sitompul, A., Tunas Bangsa, S., Utara, S., Tunas Bangsa, A., & Jln Sudirman Blok No, I. A. (2021). Teknik Data Mining Dalam Prediksi Jumlah Siswa Baru Dengan Algoritma Naive Bayes. *KESATRIA : Jurnal Penerapan Sistem Informasi (Komputer Dan Managemen)*, 2(2), 108–117.
- Rizal, S., Studi, P., Informatika, T., & Yudharta, U. (2018). *Penerapan Algoritma Naïve Bayes Untuk Prediksi Penerimaan*. 10(1), 14–21.
- Senika, A., & Iskandar, D. (2022). *Implementasi Metode Naïve Bayes Dalam Penilaian Kinerja Sales Marketing Pada PT . Pachira Distrinusa*. 6, 701–709. <https://doi.org/10.30865/mib.v6i1.3331>
- Sinaga, S., Sembiring, R. W., & Sumarno, S. (2022). *Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru*. 1(1), 55–64.
- Suntoro, J., & Kom, M. (n.d.). *DATA MINING*.
- Suyanto, D. (2019). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data* (1st ed.). Informatika Bandung.
- Teknologi, J., Mura, I., Alamsyah, M. N., Alamsyah, M. N., Informatika, P. S., Insan, U. B., & Bayes, A. N. (2020). *Penerapan Algoritma Naive Bayes Untuk Prediksi Penerimaan Fakultas Komputer*. 12(02), 156–165.