
PENERAPAN METODE NAÏVE BAYES UNTUK PENENTUAN JURUSAN SISWA - SISWI PADA SMAN 9 KENDARI

Nur Khafifah¹⁾

¹⁾Program Studi Matematika, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Halu Oleo, Kendari, Indonesia
Email: nurkhafifah298@gmail.com

Gusti Ngurah Adhi Wibawa^{2,a)}, Arman^{1,b)}, Wayan Somayasa^{1,c)}, Ruslan^{2,d)} dan La Gubu^{1,e)}

¹⁾Program Studi Matematika, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Halu Oleo, Kendari, Indonesia

²⁾Program Studi Statistika, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Halu Oleo, Kendari, Indonesia

Email: ^{a)}gnawibawa@gmail.com, ^{b)}arman.mtmk@uho.ac.id, ^{c)}wayan.somayasa@uho.ac.id,
^{d)}rushlan@gmail.com, dan ^{e)}lagubu@uho.ac.id

ABSTRAK

Penjurusan siswa-siswa disuatu sekolah merupakan suatu proses penempatan siswa-siswi ke dalam jurusan tertentu sesuai dengan kemampuan dan keinginan siswa-siswi. SMAN 9 Kendari adalah salah satu sekolah menengah atas negeri yang ada dikota kendari. Adapun pelajaran yang diberikan meliputi semua mata pelajaran wajib sesuai kurikulum yang berlaku. Saat ini SMAN 9 Kendari memiliki pemilihan jurusan yaitu Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS) Dan Ilmu Bahasa Dan Budaya (IBB). Terdapat tiga jurusan yang harus di pilih siswa-siswi salah satunya dengan mekanisme yang sudah ditentukan pihak sekolah seperti tes tertulis. Hal tersebut yang nantinya akan menjadi parameter menentukan klasifikasi siswa-siswi akan jurusan/ peminatannya. Pada penelitian ini, menggunakan penerapan metode *Naïve Bayes* untuk melakukan klasifikasi nilai dan menghasilkan rekomendasi jurusan siswa-siswi. Untuk klasifikasi jurusan siswa-siswi, variabel yang digunakan ialah Nilai Matematika SMP (X_1), Nilai Bahasa Indonesia SMP (X_2), Nilai Bahasa Inggris SMP (X_3), Nilai IPA SMP (X_4), Nilai IPS SMP (X_5), dan Nilai Ujian Sekolah SMP (X_6). Berdasarkan hasil penelitian yang dilakukan sistem penjurusan siswa- siswi dapat memberikan rekomendasi jurusan dengan nilai Akurasi adalah 55,55%, nilai sensitifitas adalah 75% dan spesifisitas adalah 25%.

Kata kunci: Penjurusan siswa-siswi, *Naïve Bayes*, klasifikasi.

ABSTRACT

Student majoring in a school is a process of placing students into certain majors according to the abilities and desires of students. SMAN 9 Kendari is one of the public high schools in Kendari city. The lessons given include all compulsory subjects according to the applicable curriculum. Currently SMAN 9 Kendari has a selection of majors namely Natural Sciences (IPA), Social Sciences (IPS) and Language and Culture (IBB). There are three majors that students must choose one of them with a mechanism that has been determined by the school such as a written test. This will later become a parameter to determine the classification of students for their majors / specializations. In this research, using the application of the Naïve Bayes method to classify scores and produce recommendations for students' majors. For the classification of students' majors, the variables used are Junior High School Mathematics Score (X_1), Junior High School Indonesian Score (X_2), junior high school English score (X_3), junior high school science score (X_4), junior high school social studies score (X_5), and junior high school exam score (X_6). Based on the results of the research conducted, the student majoring system can provide recommendations for majors with an accuracy value of 55.55%, sensitivity value is 75% and specificity is 25%.

Keywords: *Students' majors, Naïve Bayes, classification.*

1. Pendahuluan

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang

menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. *Naïve Bayes* merupakan metode pembelajaran secara probabilitas statistik. Teorema Bayes sering digunakan dalam menduga probabilitas suatu kejadian. Metode Bayes adalah salah satu cara untuk mengatasi ketidakpastian data dan memudahkan untuk dianalisa serta membantu mendapatkan keputusan terbaik pada suatu permasalahan. Metode Bayes juga dapat diterapkan untuk memprediksi probabilitas kejadian yang akan datang berdasarkan pengalaman dimasa sebelumnya. Itulah alasan mengapa sangat penting mempelajari metode Bayes [1].

Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut:

1. Metode *Naïve Bayes* bekerja teguh (robust) terdapat data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (outliner). *Naïve Bayes* juga biasa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bias mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.[2].

Klasifikasi adalah suatu proses untuk menemukan sekelompok dari beberapa data kedalam beberapa kelas dengan tujuan untuk memperkirakan kelas dari suatu objek labelnya tidak diketahui, Klasifikasi dapat diterapkan dalam berbagai aspek sehingga seiring berjalannya waktu metode klasifikasi cukup banyak dikembangkan, Namun terdapat permasalahan yang sering ditemui dalam klasifikasi yaitu masalah ketidakseimbangan data. Ketidakseimbangan data terjadi ketika salah satu kelas memiliki jumlah yang jauh lebih besar dibanding kelas lainnya sehingga menyebabkan menurunnya kinerja klasifikasi pada kelas minoritas. Kinerja algoritma *machine learning* biasanya dievaluasi dengan akurasi hasil prediksi, namun hal ini tidak sesuai apabila terjadi ketidakseimbangan kelas. Metode *machine learning* cenderung memberi label berupa kelas mayoritas pada data yang

diprediksi dan mengabaikan kelas minoritas sehingga hanya akan menghasilkan akurasi hasil prediksi yang baik kelas mayoritas saja [3].

Penjurusan siswa merupakan suatu proses penempatan siswa-siswi ke dalam jurusan atau peminatan tertentu, sehingga siswa-siswi dapat menyerap semua mata pelajaran dengan optimal dan sesuai dari kemampuan yang dimilikinya. Penjurusan dilakukan karena terdapat perbedaan dari setiap individu yang membuat tingkat pemahaman setiap siswa-siswi dalam penerimaan informasi berbeda-beda. Sulitnya dalam mengklasifikasikan jurusan menjadi kendala para siswa-siswi dan menjadikan waktu yang kurang efisien karena harus menghitung satu per satu nilai tes yang sudah dilaksanakan sebelumnya [4].

2. Kajian Pustaka

2.1. Analisis Bayes

Bayes (hukum Bayes atau aturan Bayes) ditemukan oleh ilmuwan Inggris Thomas Bayes pada tahun 1763 yang menjelaskan tentang probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya pada kondisi yang mungkin terkait dengan peristiwa tersebut. Dalam teori probabilitas dan statistika, Teorema Bayes adalah sebuah teorema dengan dua penafsiran berbeda. Dalam penafsiran Bayes, Teorema ini menyatakan seberapa jauh derajat kepercayaan subjektif harus berubah secara rasional ketika ada petunjuk baru. Penerapan teorema *Bayes* untuk memperbarui kepercayaan dinamakan inferensi *Bayes*. Teorema *Bayes* merupakan teori kondisi yang memperhitungkan kemungkinan suatu kejadian (hipotesis) bergantung pada kejadian lain.

2.2. Naïve Bayes

Naïve Bayes merupakan salah satu algoritma yang terdapat pada taktik klasifikasi. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema *Bayes*.

Persamaan dari Teorema *Naïve Bayes* adalah

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Untuk menjelaskan Metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena

itu, Metode *Naïve Bayes* diatas disesuaikan sebagai berikut :

$$P(C|X_1 \dots X_n) = \frac{P(C)P(X_1 \dots X_n|C)}{P(C|X_1 \dots X_n)}$$

Dimana variabel C merepresentasikan kelas, sementara variabel $X_1 \dots X_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*prosterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas secara global (disebut juga *evidence*).

$$\text{Posterior} = \frac{\text{prior} * \text{likelihood}}{\text{evidence}}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|X_1 \dots X_n)$ menggunakan aturan perkalian sebagai berikut :

$$P(C|X_1 \dots X_n) = P(C)P(X_1 \dots X_n|C)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Sehingga perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naïf), bahwa masing-masing petunjuk (X_1, X_2, \dots, X_n) saling bebas (*independent*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sabagai berikut :

$$P(X_i|X_j) = \frac{P(X_i \cap X_j)}{P(X_j)} = \frac{P(X_i)P(X_j)}{P(X_j)} = P(X_i)$$

untuk $i \neq j$, sehingga

$$P(X_i|C, X_j) = P(X_i|C)$$

2.2.1 Karakteristik Naïve Bayes

Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut:

1. Metode *Naïve Bayes* bekerja teguh (*robust*) terdapat data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (outliner). *Naïve Bayes* juga biasa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan predeksi.

2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bias mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

2.3.1 Kelebihan dan Kekurangan Naïve Bayes

Adapun Kelebihan dan Kekurangan Metode *Naïve Bayes*:

a. Kelebihan Metode *Naïve Bayes*

Metode *Naïve Bayes* memiliki beberapa kelebihan yaitu sebagai berikut :

1. Alur kerja dari algoritma yang mudah dipahami oleh pengguna.
2. Memiliki struktur kode yang tidak rumit/serhana.
3. Memiliki rentang waktu pendek dalam perhitungan.
4. Dapat menangani masalah dengan data kuantitatif maupun data diskrit.
5. Memiliki titik noise yang kokoh dan kuat terhadap atribut yang tidak relevan.
6. Algoritma *Naïve Bayes* hanya memerlukan data kecil untuk melakukan proses klasifikasi.
7. Nilai yang hilang tidak termasuk dalam perhitungan prediksi.

b. Kekurangan Metode *Naïve Bayes*

Metode *Naïve Bayes* memiliki beberapa kekurangan yaitu sebagai berikut:

1. Satu probabilitas saja tidak bias mengukur seberapa dalam tingkat keakuratannya. Dengan kata lain, kurang bukti untuk membuktikan kebenaran jawaban yang dihasilkan dari teori ini.
2. Tidak berlaku jika probabilitas kondisionalnya adalah 0 (nol), apabila nol maka probabilitas prediksi akan bernilai nol juga
3. Mengansumsikan variabel bebas.

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menunjukkan jumlah frekuensi dan kombinasi dari dataset yang diberikan. Algoritma *Naïve Bayes Classifier* merupakan sebuah metode yang mengandalkan probabilitas dari record. Dalam melakukan klasifikasi dengan Algoritma *Naïve Bayes Classifier* tersebut memerlukan proses pencarian probabilitas yang di gunakan untuk melihat prediksi di masa yang akan datang berdasarkan data dari masa lampau. Dalam menjalankan proses tersebut tentunya diperlukan sebuah dataset. Pada Algoritma *Naïve Bayes Classifier*, dataset tersebut dibagi menjadi dua yaitu data training dan data testing.

Model statistik merupakan salah satu model yang terpercaya sangat andal sebagai pendukung

pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu metode yang menggunakan konsep probabilitas adalah Naïve Bayesian Classifier. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Apabila diberikan k atribut yang saling bebas (independence), nilai probabilitas dapat diberikan sebagai berikut :

$$P = (x_1, \dots, x_k | C) = P(x_1 | C) \dots P(x_k | C)$$

Jika atribut ke- i bersifat diskrit, maka $P(x_i | C)$ diestimasi sebagai frekuensi relatif dari sampel yang memiliki nilai x_i sebagai atribut ke- i dalam kelas C . Namun, jika atribut ke- i bersifat kontinu, maka $P(x_i | C)$ diestimasi dengan fungsi Densitas Gauss.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

2.3.1 Kelebihan Naïve Bayes Classifier

Keuntungan penggunaan Metode *Naïve Bayes Classifier* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel *independent*, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

2.4 Klasifikasi

Klasifikasi merupakan sebuah cara untuk menemukan properti yang sama pada himpunan objek di dalam sebuah data dan mengklasifikasikannya kedalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan, bertujuan untuk menentukan model dari *training set* yang membedakan kelas yang sesuai, model tersebut selanjutnya digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui. Beberapa model klasifikasi di gunakan dalam *Data Mining*.

Klasifikasi adalah suatu fungsionalitas data mining yang menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek didalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian. Pada tahap pembelajaran, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis *training data*. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $Y = F(X)$ dimana Y adalah kelas prediksi dan X adalah *tuple* yang ingin diprediksi kelasnya. Selanjutnya pada tahap pengklasifikasian, model

yang telah dihasilkan akan digunakan untuk melakukan pengklasifikasian. Klasifikasi proses pencarian sekumpulan model yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu obyek yang belum diketahui kelasnya. Naïve Bayes merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naïve Bayes* adalah Teorema *Bayes*, yaitu melakukan klasifikasi dengan melakukan perhitungan nilai probabilitas.

Menurut (Oktaviana, 2016) menjelaskan bahwa pada proses pengklasifikasian terdapat dua proses yang harus dilakukan yaitu:

1. Proses Latih (*Training*)

Dalam proses ini akan digunakan data latih set atau data sampel yang telah diketahui label-label atau atribut dari sampel tersebut untuk membangun model.

2. Proses Uji (*Testing*)

Dalam proses uji ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses latih maka dibangun data yang disebut dengan data uji untuk mengklasifikasikan label-labelnya.

Menurut (Sulaksono, dkk, 1945) Ada dua jenis model klasifikasi yaitu :

1. Pemodelan deskriptif (*descriptive modelling*) yaitu model klasifikasi yang dapat berfungsi sebagai suatu alat penjelasan untuk membedakan objek-objek dalam kelas-kelas yang berbeda.
2. Pemodelan prediktif (*predictive modelling*) yaitu klasifikasi yang dapat digunakan untuk memprediksi label kelas *record* yang tidak diketahui.

Proses klasifikasi didasarkan pada empat komponen yaitu:

1. Kelas yaitu variabel dependen yang berupa kategorik yang mempresentasikan label yang terdapat pada objek.
2. *Predictor* yaitu variabel independen yang direpresetasikan oleh karakteristik atribut data.
3. *Training dataset* yaitu satu set data yang berisi nilai dari kedua komponen diatas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
4. *Testing dataset* yaitu data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

Ketepatan klasifikasi suatu *classifier* secara keseluruhan dapat diukur melalui akurasi dan secara khusus melalui sensitivitas untuk kelas positif serta spesifisitas untuk kelas negative. Untuk mengevaluasi kinerja suatu metode klasifikasi dilakukan dengan *confusion matrix*. *Confusion matrix* dapat dilihat pada Tabel 2. 1

Tabel 2. 1 Confusion Matrix

Class	Predictive Positive	Predictive Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Keterangan dari masing-masing istilah pada Tabel 2. 1 adalah sebagai berikut:

- TP : *True Positive*, yaitu hasil prediksi dan kondisi sebenarnya sama-sama positif.
- TN : *True Negative*, yaitu hasil prediksi dan kondisi sebenarnya sama-sama negatif.
- FP : *False Positive*, yaitu hasil prediksi namun sebenarnya positif.
- FN : *False Negative*, yaitu hasil prediksi namun sebenarnya negatif.

Akurasi, sensitivitas dan spesifisitas dapat dihitung dengan rumus berikut:

a. Akurasi

Akurasi merupakan jumlah perbandingan data yang benar diklasifikasikan dengan jumlah keseluruhan data (Syahidah, 2014).

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP}$$

b. Sensitifitas

Sensitifitas merupakan fungsi dari kelas data positif yang diklasifikasikan dengan benar dibandingkan dengan keseluruhan data kelas actual positif (Syahidah, 2014).

$$Sensitivitas = \frac{TP}{TP + FN}$$

2.5. Penjurusan Siswa-Siswi

SMAN 9 Kendari adalah salah satu sekolah menengah atas negeri yang ada di kota kendari. Adapun pelajaran yang diberikan meliputi semua mata pelajaran wajib sesuai kurikulum yang berlaku. Saat ini SMAN 9 Kendari memiliki pemilihan jurusan yaitu Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS) Dan Ilmu Bahasa dan Budaya (IBB). Terdapat tiga jurusan yang harus di pilih siswa-siswi salah satunya dengan mekanisme yang sudah ditentukan pihak sekolah seperti tes tertulis. Hal tersebut yang nantinya akan menjadi parameter menentukan klasifikasi siswa-siswi akan jurusan/ peminatannya.

Penjurusan siswa merupakan suatu proses penempatan siswa-siswi ke dalam jurusan atau peminatan tertentu, sehingga siswa-siswi dapat menyerap semua mata pelajaran dengan optimal dan sesuai dari kemampuan yang dimilikinya. Penjurusan dilakukan karena terdapat perbedaan dari setiap

individu yang membuat tingkat pemahaman setiap siswa-siswi dalam penerimaan informasi berbeda-beda. Sulitnya dalam mengklasifikasikan jurusan menjadi kendala para siswa-siswi dan menjadikan waktu yang kurang efisien karena harus menghitung satu per satu nilai tes yang sudah dilaksanakan sebelumnya (Mafakhir dan Solichin, 2020).

Tabel 2. 2 Skala Penilaian

Skala Nilai	Nama Skala
81-100	Sangat Baik (A)
61-80	Baik (B)
41-60	Cukup (C)
21-40	Kurang (D)
0-20	Sangat Kurang (E)

3. Hasil dan Pembahasan

3.1. Metode Naïve Bayes

Naïve Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode *Naïve Bayes* data *string* yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria nilai data *string* maupun kriteria dengan nilai data numerik.

3.2. Membaca Data Training

Untuk menentukan data yang dianalisis dengan Metode *Naïve Bayes* maka langkah pertama yang dilakukan adalah membaca data *training*. Data *training* akan digunakan untuk melatih algoritma dalam pencarian model yang sesuai. Apabila data *training* yang digunakan 75% dan data *testing* 25%. Jumlah data *training* apabila dijumlah dengan data *testing* akan menjadi dataset tersebut. Jumlah persetasi pada data *training* nantinya akan ditentukan berdasarkan perulangan pada proses Program R agar data yang digunakan pada proses itu tetap maka menggunakan dataset yang sama (Karyadiputra & Hijriana, 2018).

Adapun data training yang digunakan yaitu data yang sudah di bagi 75% menggunakan Program R yang dapat dilihat pada Tabel 3. 1. berikut :

Siswa	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Siswa 1	IPA	A	A	A	A	A	A
Siswa 2	IPA	B	A	A	A	A	A
Siswa 3	IPA	A	A	A	A	A	A
Siswa 4	IPA	A	A	A	A	A	A
Siswa 5	IPA	A	A	A	A	A	A

Siswa 6	IPA	B	A	B	B	B	B
Siswa 7	IPA	A	A	A	A	A	A
Siswa 8	IPA	A	A	A	A	A	A
Siswa 9	IPA	A	A	A	A	A	A
Siswa 10	IPA	B	A	A	B	A	A
Siswa 11	IPA	A	A	A	B	A	A
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Siswa 36	IBB	B	A	A	B	B	B

3.3. Menentukan Probabilitas

Berdasarkan data siswa-siswi diketahui jumlah data *training* adalah 36 data dan terdapat 3 jurusan yaitu IPA, IPS dan IBB. Dari data *training*, ditentukan nilai probabilitas dari tiap-tiap variabel kedalam kategori. sebagai berikut:

Tabel 3. 2 Variabel Y

Variabel (Y)	C	P(C)
IPA	16	0,4444
IPS	10	0,2778
IBB	10	0,2778
JUMLAH	36	1

3.3.1 Menentukan Probabilitas untuk Setiap Kategori Nilai

Probabilitas untuk variabel X_i diperoleh menggunakan rumus :

$$P(X_i = Y|C = Y) = \frac{P(X_i = Y \cap C = Y)}{P(C = Y)}$$

- Nilai Matematika (X_1)

$$P(X_1 = Baik|C = IPA) = \frac{7}{16} = 0,4375$$

$$P(X_1 = Baik|C = IPS) = \frac{8}{10} = 0,8$$

$$P(X_1 = Baik|C = IBB) = \frac{7}{10} = 0,7$$

- Nilai Bahasa Indonesia (X_2)

$$P(X_2 = Sangat Baik|C = IPA) = \frac{15}{16} = 0,9375$$

$$P(X_2 = Sangat Baik|C = IPS) = \frac{4}{10} = 0,4$$

$$P(X_2 = Sangat Baik|C = IBB) = \frac{7}{10} = 0,7$$

- Nilai Bahasa Inggris (X_3)

$$P(X_3 = Sangat Baik|C = IPA) = \frac{13}{16} = 0,8125$$

$$P(X_3 = Sangat Baik|C = IPS) = \frac{1}{10} = 0,1$$

$$P(X_3 = Sangat Baik|C = IBB) = \frac{7}{10} = 0,7$$

- Nilai IPA (X_4)

$$P(X_4 = Sangat Baik|C = IPA) = \frac{11}{16} = 0,6875$$

$$P(X_4 = Sangat Baik|C = IPS) = \frac{1}{10} = 0,1$$

$$P(X_4 = Sangat Baik|C = IBB) = \frac{4}{10} = 0,4$$

- Nilai IPS (X_5)

$$P(X_5 = Sangat Baik|C = IPA) = \frac{13}{16} = 0,8125$$

$$P(X_5 = Sangat Baik|C = IPS) = \frac{4}{10} = 0,4$$

$$P(X_5 = Sangat Baik|C = IBB) = \frac{7}{10} = 0,7$$

- Nilai Ujian Sekolah (X_6)

$$P(X_6 = Sangat Baik|C = IPA) = \frac{13}{16} = 0,8125$$

$$P(X_6 = Sangat Baik|C = IPS) = \frac{3}{10} = 0,3$$

$$P(X_6 = Sangat Baik|C = IBB) = \frac{5}{10} = 0,5$$

Dengan cara yang sama dapat dihitung nilai peluang setiap kelas dan peluang bersyarat masing-masing kategori nilai disetiap kelas. Secara ringkas dengan peluangnya.

3.4. Pengujian Metode Naïve Bayes

Pengujian Metode *Naïve Bayes* kali ini digunakan data *testing*, adapun data *testing* yang digunakan dapat dilihat pada Tabel 3. 4 berikut :

Siswa	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Siswa 1		B	A	A	A	A	A
Siswa 2		B	B	B	B	B	B
Siswa 3		A	A	A	A	A	A
Siswa 4		A	A	A	B	A	A
Siswa 5		B	B	B	A	A	B
Siswa 6		B	A	B	B	A	A
Siswa 7		B	A	A	A	A	A
Siswa 8		B	A	A	B	A	A
Siswa 9		B	A	B	A	A	A

Dengan menggunakan rumus yang telah ada, untuk menentukan masing-masing jurusan dari setiap data siswa-siswi yang akan diuji

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

3.4.1 Nilai $P(C)$ Berdasarkan Data Training

$$P(IPA) = \frac{16}{36} = 0,4444$$

$$P(IPS) = \frac{10}{36} = 0,2778$$

$$P(IBB) = \frac{10}{36} = 0,2778$$

3.4.2 Menentukan Jurusan Siswa-Siswi

Menghitung probabilitas bersyarat sesuai kondisi yang diberikan. Misalnya dilakukan perhitungan pada data *testing* ke – 1 dimana diketahui nilai-nilainya dapat dilihat pada Tabel 4. 4. Setelah di peroleh nilai probabilitas setiap kriteria yang dapat di lihat pada Tabel 4.3. kemudian dilakukan perkalian nilai kriteria sesuai dengan *likelihood* pada kategori IPA, IPS dan IBB Maka dilakukan perhitungan sebagai berikut :

$$\begin{aligned} P(Y_{IPA}|X_1, X_2, X_3, X_4, X_5, X_6) &= P(Y_{IPA}) \prod_{i=1}^6 P(X_i|Y_{IPA}) \\ &= P(Y_{IPA})P(X_1|Y_{IPA})P(X_2|Y_{IPA})P(X_3|Y_{IPA})P(X_4|Y_{IPA})P(X_5|Y_{IPA})P(X_6|Y_{IPA}) \\ &= 0,4444 * 0,4375 * 0,9375 * 0,8125 * 0,8125 * 0,6875 \\ &= 0,0672 \end{aligned}$$

$$\begin{aligned} P(Y_{IPS}|X_1, X_2, X_3, X_4, X_5, X_6) &= P(Y_{IPS}) \prod_{i=1}^6 P(X_i|Y_{IPS}) \\ &= P(Y_{IPS})P(X_1|Y_{IPS})P(X_2|Y_{IPS})P(X_3|Y_{IPS})P(X_4|Y_{IPS})P(X_5|Y_{IPS})P(X_6|Y_{IPS}) \\ &= 0,2228 * 0,8 * 0,4 * 0,1 * 0,1 * 0,4 * 0,3 \\ &= 0,0001 \end{aligned}$$

$$P(Y_{IBB}|X_1, X_2, X_3, X_4, X_5, X_6) = P(Y_{IBB}) \prod_{i=1}^6 P(X_i|Y_{IBB})$$

$$\begin{aligned} &= P(Y_{IBB})P(X_1|Y_{IBB})P(X_2|Y_{IBB})P(X_3|Y_{IBB})P(X_4|Y_{IBB})P(X_5|Y_{IBB})P(X_6|Y_{IBB}) \\ &= 0,2228 * 0,7 * 0,7 * 0,7 * 0,4 * 0,7 * 0,5 \\ &= 0,0107 \end{aligned}$$

Untuk menghasilkan nilai probabilitas maka dilakukan normalisasi terhadap *Likelihood* tersebut sehingga jumlah nilai yang diperoleh sama dengan 1.

a. Jurusan IPA

$$\frac{0,0672}{0,0672 + 0,0001 + 0,0107} = 0,8615$$

b. Jurusan IPS

$$\frac{0,0001}{0,0672 + 0,0001 + 0,0107} = 0,0013$$

c. Jurusan IBB

$$\frac{0,0107}{0,0672 + 0,0001 + 0,0107} = 0,1372$$

Dari nilai probabilitas yang diperoleh maka dapat disimpulkan dari data yang diuji diatas diketahui nilai probabilitas Jurusan IPA dengan nilai 0,8615 lebih besar dari pada Jurusan IPS dengan nilai 0,0013 dan Jurusan IBB dengan nilai 0,1372.

3.5. Hasil Klasifikasi Naïve Bayes

Hasil klasifikasi dari data *testing* dapat dilihat pada Tabel 3. 5 yang menggunakan software *R Studio* dalam proses klasifikasi.

Tabel 3.5 Hasil Klasifikasi

Siswa	Nilai Probabilitas			Kelas Asli	Kelas Prediksi
	IPA	IPS	IBB		
1	0,8615	0,0013	0,1372	IPA	IPA
2	0,0021	0,9638	0,0341	IPA	IPS
3	0,9371	0,011	0,0618	IPA	IPA
4	0,8201	0,0021	0,1778	IPA	IPA
5	0,0339	0,5593	0,4068	IPS	IPS
6	0,2904	0,3569	0,3527	IPS	IPS
7	0,8615	0,0013	0,1372	IBB	IPA
8	0,5931	0,0194	0,3875	IBB	IPA
9	0,6327	0,1346	0,2327	IBB	IPA

Hasil klasifikasi serta performa klasifikasi digunakan untuk mengevaluasi kinerja suatu metode

klasifikasi. Dengan bantuan software *R Studio* diperoleh hasil klasifikasi serta performa klasifikasi yang dapat dilihat pada Tabel 3. 6

Observasi	Prediksi		
	IBB	IPA	IPS
IBB	0	3	0
IPA	0	3	1
IPS	0	0	2

Berdasarkan Tabel 4.6 kemudian dilakukan perhitungan untuk mengetahui akurasi, sensitifitas dan spesifisitas sebagai berikut:

$$\text{Akurasi} = \frac{3 + 2}{3 + 1 + 2 + 3} \times 100 = 55,55\%$$

$$\text{Sensitifitas} = \frac{3}{3 + 1} \times 100 = 75\%$$

$$\text{Spesifisitas} = \frac{1}{1 + 3} \times 100 = 25\%$$

Dari hasil klasifikasi diatas dapat dilihat bahwa prediksi kelas (IPA) dari 4 data *testing* sebanyak 3 data diprediksi dengan tepat dan 1 data diprediksi tidak tepat. Kelas (IPS) dari 2 data *testing* sebanyak 2 yang di prediksi dengan tepat dan 0 data diprediksi tidak tepat. Kelas (IBB) dari 3 data testing sebanyak 0 yang di prediksi dengan tepat dan 3 data diprediksi tidak tepat. Jadi, dari 9 data *testing* sebanyak 5 data diprediksi dengan tepat dan 4 data diprediksi dengan tidak tepat.

4. Kesimpulan dan Saran

Berdasarkan hasil pembahasan yang dapat disimpulkan sebagai berikut:

Dari hasil penelitian maka dapat disimpulkan bahwa berdasarkan analisis dan pembahasan yang telah dilakukan pada data *training* dan data *testing* yaitu dengan perbandingan 75% dan 25% dimana data *training* sebanyak 36 siswa-siswi dan data *testing* sebanyak 9 siswa-siswi. Dengan hasil akurasi sebesar 55,55% artinya performa dari hasil klasifikasi ini dapat dikatakan tepat karena melebihi 50% , sensitifitas sebesar 75% dan spesifisitas sebesar 25% artinya evaluasi sistem untuk klasifikasi agar menghasilkan nilai yang baik.

Berdasarkan kesimpulan di atas, maka penulis memberikan saran sebagai berikut:

Berdasarkan hasil penelitian ini, disarankan untuk peneliti selanjutnya menggunakan lebih banyak data yang dimasukkan kedalam sistem, dikarenakan dengan lebih banyaknya data yang dimasukkan maka akan menghasilkan prediksi yang

mendekati sebenarnya dalam penentuan jurusan siswa-siswi agar diperoleh hasil akurasi yang maksimal, serta menggunakan metode klasifikasi lainnya sebagai perbandingan. Di karenakan nilai akurasi pada penelitian ini kurang maksimal maka disarankan untuk menggunakan pembagian data *training* 90% dan data *testing* 10% sehingga menghasilkan nilai akurasi 75% artinya performa dari hasil akurasi ini dapat di gunakan karena semakin besar hasil akurasi yang didapatkan performa klasifikasi baik.

Ucapan Terima Kasih

Penelitian ini dapat terlaksana dengan baik atas bantuan dari berbagai pihak. Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada pembimbing 1 Bapak Dr. Gusti Ngurah Adhi Wibawa, S.Si., M.Si pembimbing 2 Dr. Arman, S.Si., M.Si., dan para penguji yang memberikan saran, kritikan dan ide sehingga penelitian dapat terlaksana dengan baik.

Daftar Pustaka

- [1] A. M. Ainul dan I. T. Utami. (2018) ‘Penerapan Pemodelan Analisis Regresi Linear Berganda Dengan Pendekatan Bayesian Pada Data Aset Bank di Indonesia’, *Jurnal Keteknian Dan Sains*, I(1).
- [2] L. J. Bain dan M. Engelhardt. (1991) *introduction To Probability And Mathematical Statistics*. Second Edi. California: Duxbury Press.
- [3] P. Batarius dan F. Tedy. (2017) ‘Pendekatan Metode Bayes untuk Menentukan Jenis Penyakit pada Ternak Babi’, *Jurnal Ilmiah Widya Teknik*, pp. 1–8.
- [4] A. Firdaus, M. Walid dan A. Anwari. (2022) ‘Klasifikasi Kasus Covid-19 Menggunakan Model Naive Bayes Classifier’, *JATI (Jurnal Mahasiswa Teknik ...*, 6(2), pp. 583–588. Available at: <https://ejournal.itn.ac.id/index.php/jati/article/view/53333%0Ahttps://ejournal.itn.ac.id/index.php/jati/article/download/5333/3371>.
- [5] A. S. Huda, R. M. Awangga. dan R.N.S, Fathonah. (2020) *Prediksi Penerimaan*

- Pegawai Baru Dengan Metode Naive Bayes.*
Edited by R.M. Awangga. Bandung: Kreatif Industri Nusantara.
- [6] E. Karyadiputra dan N. Hijriana. (2018) ‘Analisis Penerapan Algoritma Naive Bayes Untuk Klasifikasi Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan’, *Technologia: Jurnal Ilmiah*, 9(2), p. 105. <https://doi.org/10.31602/tji.v9i2.1374>.
- [7] S. Kusumadewi. (2009) ‘Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification’, *CommIT (Communication and Information Technology) Journal*, 3(1), pp. 6–11. <https://doi.org/10.21512/commit.v3i1.506>.
- [8] A. Z. Mafakhir dan A. Solichin. (2020) ‘Penerapan Metode Naive Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta’, *Fountain of Informatics Journal*, 5(1), pp. 21–26.
- [9] A. Z. Oktaviana. (2016) *Penerapan Datamining Klasifikasi Pola Nasabah Menggunakan Algoritma C4.5 Pada Bank Bri Batang*. Semarang.
- [10] D. A. Pratiwi, R. M. Awangga dan Muhammad Yusril Helmi Setyawan (2020) *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika*. Bandung: Kreatif Industri Nusantara.
- [11] A. Saleh. (2015) ‘Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga’, *Citec Journal*, 2(3), pp. 207–217. <https://doi.org/10.20895/inista.v1i2.73>.
- [12] L. Setiyani. (2020) ‘Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes: Systematic Review’, *Faktor Exacta*, 13(1), pp. 35–43.
- [13] Y. N. S. Sidiq, R. N. S. Fathonah dan Novita Riza (2020) *Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur Dengan Algoritma Naive Bayes Pada Sistem Juragan Berbasis Komunitas*. Bandung: Kreatif Industri Nusantara.
- [14] H. Susana. (2022) ‘Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet’, *Jurnal Sistem Informasi dan Teknologi Informasi (JURSISTEKNI)*, 4(No.1).
- [15] S. Winanta, Y. Oslan dan G. Santoso. (2013) ‘Implementasi Metode Bayesian Dalam Penjurusan Di SMA Bruderan Purworejo’, *Jurnal Eksis*, 06(02), pp. 21–28.
-