



Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru

Sulaiman Sinaga^{1*}, Rahmat W. Sembiring², S. Sumarno¹

¹Program Studi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

²Program Studi Manajemen Informatika, Politeknik Negeri Medan, Medan, Indonesia

*sulaiman.sinaga.stb@gmail.com

Diterima: 31/01/2022; Disetujui: 01/02/2022; Diterbitkan: 01/02/2022

Abstrak—Calon siswa baru yang mendaftar pada beberapa sekolah-sekolah khususnya di SMK (Sekolah Menengah Kejuruan) sangatlah bervariasi jumlahnya, tergantung dari kualitas sekolah serta minat dari calon siswa itu sendiri, juga turut andil dalam mempengaruhi banyaknya jumlah siswa. Selain itu status swasta biasanya menjadikan SMK swasta sebagai alternatif kedua setelah negeri, sehingga jumlah siswa baru yang mendaftar di sekolah SMK swasta sulit untuk diprediksi. Begitu pula hal nya dengan SMK Swasta Anak Bangsa. Oleh karena itu tujuan dari penelitian ini adalah untuk melakukan prediksi siswa baru yang akan diterima berupa klasifikasi pada SMK swasta Anak Bangsa dengan menggunakan algoritma Naive Bayes yang merupakan salah satu metode dari pembelajaran mesin. Data penelitian ini bersumber dari sekolah SMK Swasta Anak Bangsa tahun 2018 sampai dengan tahun 2021 dengan jumlah sample data yang diambil sebanyak 110 data siswa, dari 162 data siswa yang mendaftar. Hasil dari penelitian ini berdasarkan pengujian sebanyak 30 data testing yang diolah menggunakan *Rapid Miner*, diperoleh tingkat akurasi sebesar 76,67% yaitu 26 siswa diterima dan sebanyak 4 tidak diterima. Kesimpulannya bahwa proses prediksi berupa klasifikasi dengan menggunakan Naive Bayes dapat lebih cepat dan akurat serta menghasilkan tingkat akurasi yang tinggi.

Kata Kunci: Naive Bayes; Klasifikasi; Prediksi; Pembelajaran Mesin; Siswa Baru

Abstract—The number of prospective new students who enroll in several private schools, especially in Vocational High Schools (SMK), varies wildly, depending on the quality of the school and the interests of the prospective students themselves, also contributing to influencing the number of students. In addition, the private status usually makes private vocational schools the second alternative after the state, so the number of new students who will enter private vocational schools is difficult to predict. The same is true for the Children of the Nation Private Vocational School. Therefore, the purpose of this study is to indicate the number of new students in the form of classification at the Anak Bangsa private vocational school using the Naive Bayes algorithm, which is one of the methods of machine learning. This research data is sourced from the Anak Bangsa Private Vocational School from 2018 to 2021 with a total sample of 110 student data, from 162 student data who registered. This study based on testing as many as 30 testing data processed using Rapid Miner obtained an accuracy rate of 86.6%, namely 26 students accepted and four were not taken. The conclusion is that the prediction process in the form of classification using Naive Bayes can be faster and more accurate and produces a high level of accuracy.

Keywords: Naive Bayes; Classification; Prediction; Machine Learning; New Student

1. PENDAHULUAN

Setiap sekolah memiliki tujuan menghasilkan lulusan yang berkualitas dan berdaya saing [1]. Namun dalam menghasilkan lulusan yang berkualitas tentunya tidak terlepas dari proses seleksi siswa baru pada saat mendaftar. Hal ini sesuai dengan pendapat untuk dapat menghasilkan produk yang baik maka harus menanam bibit – bibit yang baik [2]. Untuk mendapatkan bibit yang baik perlu adanya seleksi yang baik pula, dengan demikian untuk mendapatkan calon siswa yang berkualitas maka perlu adanya saringan yang baik [3]. Penerimaan siswa baru merupakan proses penyeleksian yang harus diputuskan dengan cepat dan tepat. Pada beberapa sekolah, penerimaan siswa baru memerlukan beberapa pertimbangan dan syarat-syarat yang harus dipenuhi misalnya nilai ujian nasional, nilai tes dari beberapa bidang studi dan kebijakan-kebijakan dari lembaga pendidikan tersebut [4].

Dalam proses penyeleksian siswa diharapkan pihak sekolah dapat melakukannya dengan seefektif mungkin sehingga menghasilkan calon siswa yang benar-benar sesuai dengan kualifikasi yang ditentukan. Namun pada kenyataannya, unsur manusia sering kali menimbulkan subyektifitas dalam pengambilan keputusan [5]. Oleh sebab itu diperlukan sebuah sistem yang mampu memprediksi siswa yang diterima untuk tahun ajaran yang akan datang berdasarkan calon siswa yang telah mendaftar. Pada penelitian ini sekolah yang dibahas adalah SMK swasta Anak Bangsa yang terletak di desa Bandar Siantar Kabupaten Simalungun.

Tujuan sekolah ini adalah meningkatkan kecerdasan, pengetahuan, kepribadian, akhlak mulia, serta keterampilan untuk hidup mandiri dan mengikuti pendidikan lebih lanjut. SMK swasta Anak Bangsa ini memiliki peranan untuk memberikan pelayanan terbaik khususnya bagi masyarakat Simalungun yang memerlukan pendidikan di jenjang Menengah Keatas. Dengan adanya SMK Swasta Anak Bangsa diharapkan dapat memberikan warna di dunia pendidikan baik di wilayah kabupaten Simalungun dan sekitarnya [6].

Naive Bayes adalah algoritma yang diusulkan pada penelitian ini, karena merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas [7][8][9]. Algoritma Naive Bayes banyak digunakan untuk menyelesaikan masalah prediksi berupa klasifikasi [10][11][12]. Selain itu Naive Bayes memiliki kelebihan yaitu sederhana, cepat dan memiliki akurasi yang tinggi [13]. Proses klasifikasi pada naive bayes biasanya dibagi



menjadi dua fase yaitu *learning/training* dan *testing/classify*. Pada fase *learning*, sebagian data yang telah diketahui kelas data nya diumpulkan untuk membentuk model perkiraan. Kemudian pada fase *testing* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut [14].

Penelitian terkait yang telah dilakukan antara lain: Penelitian yang dilakukan untuk memprediksi potensi pendaftaran siswa pada SMK Taman Siswa Teluk Belitung menggunakan model Naive Bayes. Pada penelitian ini sistem yang dibangun berbasis web berupa sistem informasi prediksi potensi pendaftaran siswa dengan menggunakan model naive bayes yang memiliki kemampuan untuk memprediksi potensi banyaknya siswa yang mendaftar dan tidak mendaftar di SMK Taman Siswa.

Aplikasi ini diuji menggunakan perhitungan akurasi dimana hasilnya didapatkan tingkat akurasi sebesar 86% untuk prediksi potensi pendaftaran siswa, sehingga dapat disimpulkan bahwa algoritma Naive Bayes cukup baik digunakan untuk melakukan prediksi potensi pendaftaran siswa [15]. Berikutnya penelitian yang dilakukan untuk memprediksi jumlah siswa baru pada SD Negeri Sukosari No. 095126 menggunakan algoritma Naive Bayes dengan *tools Rapi Miner*. Penelitian ini menghasilkan nilai Akurasi sebesar 85,00 %, *class precision* pada prediksi Layak memiliki nilai 92,31%, sedangkan pada prediksi Tidak Layak memiliki nilai 55,56%. *Class recall* pada true Layak memiliki nilai 85,71%, sedangkan pada true Tidak Layak memiliki nilai 71,43 %.

Penerapan data mining dalam menentukan klasifikasi penerimaan siswa baru dengan Naive Bayes dapat digunakan untuk memprediksi layak atau tidak layak seorang siswa yang akan diterima [16]. Berikutnya penelitian dengan menerapkan algoritma Naive Bayes untuk klasifikasi dalam mendukung keputusan pada sekolah SMP Al Azhar 1 Bandar lampung dalam melakukan penentuan penerima beasiswa. Sistem pendukung keputusan penerima beasiswa dibangun dengan menggunakan bahasa pemrograman Java dan database MySQL.

Hasil dari penelitian ini, bahwa dengan menerapkan metode naive bayes pada sistem pendukung keputusan penerima beasiswa dapat membantu pihak sekolah dalam melakukan penentuan penerima beasiswa dengan lebih cepat dan tepat [17]. Berdasarkan permasalahan dan uraian tersebut maka dilakukan penelitian ini.

2. METODOLOGI PENELITIAN

2.1 Data Penelitian

Dataset yang digunakan pada penelitian ini dikumpulkan dengan menggunakan metode Observasi, melihat serta mempelajari permasalahan yang ada di lapangan yang berkaitan dengan objek yang diteliti, yaitu informasi mengenai Data siswa SMK Swasta Anak Bangsa angkatan 2020 yang berjumlah 162 dengan menggunakan 4 variabel yaitu nilai Matematika, nilai IPA, nilai Bahasa Inggris dan nilai Bahasa Indonesia yang digunakan sebagai subjek dalam penelitian. Serta melaksanakan wawancara kepada pihak SMK Swasta Anak Bangsa.

Tabel 1. Dataset Penelitian

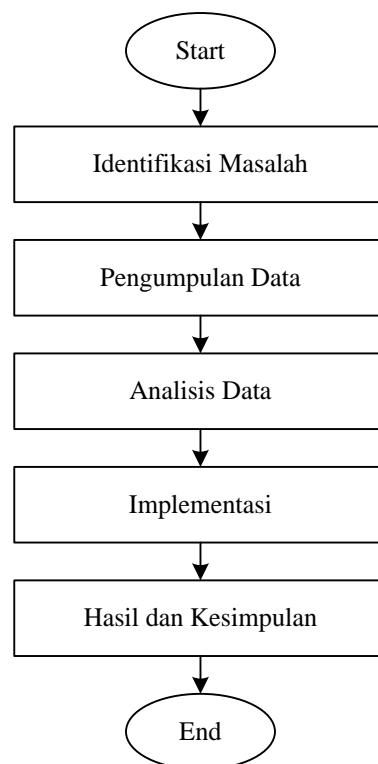
No	Nama Siswa	Nilai			
		Matematika	IPA	B. Inggris	B. Indonesia
1	Ade Rani Lestari	70	60	70	60
2	Aditya Budisantoso	75	65	80	65
3	Agung Syah Ramadhan	70	60	75	60
4	Aisyah Ramadhani	75	60	80	60
5	Aji	70	60	80	60
6	Alfi Ramadani	65	65	80	65
7	Aliya Firanti	70	70	80	70
8	Anisa Reva Amelia	70	60	80	60
9	Aqila Sakinah Pramatika	70	75	80	75
10	Azima Putri Sadiva	70	70	80	70
11	Dian Permana	70	80	85	80
12	Dimas Herianto	80	70	80	70
13	Egi Perdana	70	70	70	70
14	Eva Setia Nanda	80	60	80	60
15	Gadis Oktavia Syahfitri	85	80	80	80
16	Handi Wijaya	80	80	60	80
17	Jovanka Alkapari	70	75	85	75
18	Muhammad Ramadani	80	60	75	60
19	Nur Aisyah	80	65	70	65
20	Nur Fauzia	85	65	85	65
21	Retno Anjani	80	65	80	65
22	Rica May Anggel	70	70	80	70
23	Rico Ibrahim	80	70	80	70
24	Sri Amelia	70	70	85	70



No	Nama Siswa	Nilai			
		Matematika	IPA	B. Inggris	B. Indonesia
25	Suriono	80	60	80	60
26	Tuti Widia Ningsih	70	70	80	70
27	Waliansyah Pangestu	75	70	75	70
28	Yessy Olidia	80	70	80	70
...
161	Yopi Irawan	80	65	80	65
162	Zulham Afansyah	80	70	80	70

2.2 Tahapan Penelitian

Tahapan penelitian digunakan untuk menyelesaikan permasalahan dalam penelitian. Adapun tahapan-tahapan dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

- Identifikasi Masalah
Hal ini bertujuan untuk memahami permasalahan dengan jelas serta dapat mengatasi masalah tersebut dengan baik. Dengan demikian dapat ditentukan rencana kerja serta hal apa saja yang dibutuhkan dalam penelitian ini
- Pengumpulan Data
Dataset yang digunakan pada penelitian ini dikumpulkan dengan menggunakan metode Observasi, melihat serta mempelajari permasalahan yang ada di lapangan yang berkaitan dengan objek yang diteliti, yaitu informasi mengenai Data siswa SMK Swasta Anak Bangsa
- Analisis Data
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. Kemudian data diolah menggunakan *Rapid Miner* menggunakan Performance yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data. Berikut ini sampel data yang digunakan dalam penelitian
- Implementasi
Selanjutnya proses tahapan implementasi dengan menggunakan *Tools Rapid Miner*
- Hasil dan Kesimpulan
Setelah implementasi menggunakan *tools Rapid Miner* dilakukan, maka selanjutnya adalah melihat hasil dari penggunaan rapid miner, dilanjutkan dengan membuat kesimpulan berdasarkan hasil yang telah diperoleh.



3. HASIL DAN PEMBAHASAN

Dalam menggunakan Algoritma Naive Bayes hal yang perlu dilakukan yaitu menentukan *input training* serta target *testing* yang ingin dihasilkan. Dalam penelitian ini data yang digunakan adalah data penerimaan siswa dengan total 110. Kemudian data dibagi menjadi 2 data set yaitu 80 data training dan 30 sebagai data testing. Data tersebut akan digunakan untuk melakukan proses perhitungan data probabilitas baik itu training dan testing serta proses menghitung probabilitas akhir. Setelah mendapatkan probabilitas akhir proses terakhir untuk menentukan prediksi penerimaan siswa baru yaitu membandingkan antara diterima dan tidak diterima. Data penelitian yang digunakan terlebih dahulu akan dikonversi menjadi huruf agar lebih mudah dalam proses perhitungan di *tools Rapid miner* yang dapat di lihat pada tabel 2.

Tabel 2. Dataset Penelitian

No	Nilai	Konversi
1	80-100	Sangat Baik
2	70-79	Baik
3	60-69	Cukup
4	55-59	Kurang
5	50-54	Sangat Kurang

Tujuan pembagian data ini untuk menentukan data yang akan digunakan sebagai data *training* dan data *testing* dalam proses menentukan hasil menggunakan Algoritma Naive Bayes, dapat dilihat pada tabel 3 dan 4.

Tabel 3. Data Training

No	Nama	Kriteria				Hasil
		A1	A2	A3	A4	
1	Siswa 1	B	B	SB	B	Diterima
2	Siswa 2	SB	B	SB	B	Diterima
3	Siswa 3	SB	SB	SB	SB	Diterima
4	Siswa 4	SB	SB	SB	B	Diterima
5	Siswa 5	SB	B	SB	B	Diterima
6	Siswa 6	SB	K	B	SB	Diterima
7	Siswa 7	C	C	C	K	Tidak
8	Siswa 8	K	K	K	K	Tidak
9	Siswa 9	B	C	C	K	Diterima
10	Siswa 10	B	B	C	K	Tidak
11	Siswa 11	B	C	B	B	Tidak
12	Siswa 12	B	B	SB	B	Diterima
13	Siswa 13	C	C	C	C	Diterima
14	Siswa 14	SB	SB	SB	B	Diterima
15	Siswa 15	B	B	B	B	Diterima
16	Siswa 16	B	C	C	K	Tidak
17	Siswa 17	B	B	B	B	Diterima
18	Siswa 18	SB	B	SB	B	Diterima
19	Siswa 19	C	C	K	K	Diterima
..
80	Siswa 78	B	SB	B	B	Diterima

Keterangan :

SB = Sangat Baik, B = Baik, C = Cukup, K = Kurang, SK = Sangat Kurang

Tabel 4. Data Testing

No	Nama	Kriteria				Hasil
		A1	A2	A3	A4	
1	Siswa 1	B	C	C	C	?
2	Siswa 2	B	B	B	B	?
3	Siswa 3	B	B	B	B	?
4	Siswa 4	SB	B	B	C	?
5	Siswa 5	B	B	B	B	?
6	Siswa 6	B	B	B	B	?
7	Siswa 7	B	B	B	C	?
8	Siswa 8	B	B	B	B	?



No	Nama	Kriteria				Hasil
		A1	A2	A3	A4	
9	Siswa 9	SB	B	B	B	?
10	Siswa 10	SB	B	B	B	?
11	Siswa 11	B	B	SB	B	?
12	Siswa 12	B	C	B	B	?
13	Siswa 13	B	B	B	B	?
14	Siswa 14	B	B	B	B	?
15	Siswa 15	B	B	B	B	?
16	Siswa 16	SB	B	B	C	?
17	Siswa 17	B	C	C	K	?
18	Siswa 18	B	C	B	K	?
19	Siswa 19	B	C	C	SK	?
20	Siswa 20	B	B	B	C	?
21	Siswa 21	B	B	B	C	?
22	Siswa 22	SB	B	K	K	?
23	Siswa 23	B	B	B	C	?
24	Siswa 24	B	B	B	C	?
25	Siswa 25	B	B	B	K	?
26	Siswa 26	C	C	B	B	?
27	Siswa 27	B	B	B	B	?
28	Siswa 28	B	C	C	B	?
29	Siswa 29	B	B	B	B	?
30	Siswa 30	B	C	C	B	?

Kemudian data diolah dengan menggunakan Microsoft Excel lalu diproses secara perhitungan manual dengan Algoritma Naive Bayes dan proses pengujian dilakukan dengan menggunakan *Tools Rapid Miner 5.3*.

3.1 Penerapan Algoritma Naive Bayes

Setelah data ditentukan, langkah awal adalah melakukan perhitungan diterima dan tidak. Dari 80 data set yang digunakan, diketahui diterima sebanyak 62 data, dan tidak sebanyak 18 data. Adapun tahapan proses perhitungan menggunakan algoritma Naive Bayes untuk prediksi penerimaan siswa sebagai berikut:

3.1.1 Baca Data Training

Seperti yang tertera pada tabel 3 dataset yang digunakan bukanlah data numerik, maka proses selanjutnya menghitung jumlah dan probabilitas dari setiap kategori.

3.1.2 Hitung Jumlah Probabilitas

a. Perhitungan probabilitas prior ($P(C_i)$)

Berdasarkan dari data set jumlah data yang digunakan sebanyak 80 data dimana terdapat 62 data yang diterima dan 18 data tidak diterima. Proses perhitungan probabilitas adalah sebagai berikut :

$$P(C_{Diterima}) = \frac{80}{62} = 0,775$$

$$P(C_{Tidak}) = \frac{80}{18} = 0,225$$

b. Selanjutnya menghitung probabilitas pada masing-masing kategori.

1. Parameter Nilai Matematika (A1).

$$P(\text{Sangat Baik} | \text{Diterima}) = \frac{17}{62} = 0,2742$$

$$P(\text{Sangat Baik} | \text{Tidak}) = \frac{1}{18} = 0,0556$$

$$P(\text{Baik} | \text{Puas}) = \frac{37}{62} = 0,5968$$

$$P(\text{Baik} | \text{Tidak}) = \frac{8}{18} = 0,4444$$

$$P(\text{Cukup} | \text{Diterima}) = \frac{8}{62} = 0,1290$$

$$P(\text{Cukup} | \text{Tidak}) = \frac{4}{18} = 0,2222$$

$$P(\text{Kurang} | \text{Diterima}) = \frac{0}{62} = 0$$

$$P(\text{Kurang} | \text{Tidak}) = \frac{3}{18} = 0,1667$$

$$P(\text{Sangat Kurang} | \text{Diterima}) = \frac{0}{62} = 0$$

$$P(\text{Sangat Kurang} | \text{Tidak}) = \frac{2}{18} = 0,1111$$



Hasil perhitungan probabilitas pada parameter Nilai Matematika dengan kategori A1 dapat dilihat pada tabel 5 berikut.

Tabel 5. Hasil Perhitungan Probabilitas Kategori A1

Kategori : A1	Diterima	Tidak	P.Diterima	P.Tidak
Sangat Baik	17	1	0,2742	0,0556
Baik	37	8	0,5968	0,4444
Cukup	8	4	0,1290	0,2222
Kurang	0	3	0	0,1667
Sangat Kurang	0	2	0	0,1111
Jumlah	62	18	1	1
Total Data		80		

2. Parameter Nilai IPA (A2).

$$P(\text{Sangat Baik}|\text{Diterima}) = \frac{14}{62} = 0,2258$$

$$P(\text{Sangat Baik}|\text{Tidak}) = \frac{1}{18} = 0,0556$$

$$P(\text{Baik}|\text{Diterima}) = \frac{39}{62} = 0,6290$$

$$P(\text{Baik}|\text{Tidak}) = \frac{7}{18} = 0,3889$$

$$P(\text{Cukup}|\text{Diterima}) = \frac{8}{62} = 0,1290$$

$$P(\text{Cukup}|\text{Tidak}) = \frac{5}{18} = 0,2778$$

$$P(\text{Kurang}|\text{Diterima}) = \frac{1}{62} = 0,0161$$

$$P(\text{Kurang}|\text{Tidak}) = \frac{3}{18} = 0,1667$$

$$P(\text{Sangat Kurang}|\text{Diterima}) = \frac{0}{62} = 0$$

$$P(\text{Sangat Kurang}|\text{Tidak}) = \frac{2}{18} = 0,1111$$

Hasil perhitungan probabilitas pada parameter Nilai IPA dengan kategori A2 dapat dilihat pada tabel 6. berikut:

Tabel 6. Hasil Perhitungan Probabilitas Kategori A2

Kategori : A2	Diterima	Tidak	P. Diterima	P.Tidak
Sangat Baik	14	1	0,2258	0,0556
Baik	39	7	0,6290	0,3889
Cukup	8	5	0,1290	0,2778
Kurang	1	3	0,0161	0,1667
Sangat Kurang	0	2	0,0000	0,1111
Jumlah	62	18	1	1
Total Data		80		

3. Parameter Nilai Bahasa Inggris (A3).

$$P(\text{Sangat Baik}|\text{Diterima}) = \frac{17}{62} = 0,2742$$

$$P(\text{Sangat Baik}|\text{Tidak}) = \frac{1}{18} = 0,0556$$

$$P(\text{Baik}|\text{Diterima}) = \frac{28}{62} = 0,4516$$

$$P(\text{Baik}|\text{Tidak}) = \frac{2}{18} = 0,1111$$

$$P(\text{Cukup}|\text{Diterima}) = \frac{16}{62} = 0,2581$$

$$P(\text{Cukup}|\text{Tidak}) = \frac{8}{18} = 0,4444$$

$$P(\text{Kurang}|\text{Diterima}) = \frac{1}{62} = 0,0161$$

$$P(\text{Kurang}|\text{Tidak}) = \frac{5}{18} = 0,2778$$

$$P(\text{Sangat Kurang}|\text{Diterima}) = \frac{0}{62} = 0$$

$$P(\text{Sangat Kurang}|\text{Tidak}) = \frac{2}{18} = 0,1111$$

Hasil perhitungan probabilitas pada parameter Bahasa Inggris dengan kategori A3 dapat dilihat pada tabel 7 berikut:

**Tabel 7.** Hasil Perhitungan Probabilitas Kategori A3

Kategori : A3	Diterima	Tidak	P. Diterima	P.Tidak
Sangat Baik	17	1	0,2742	0,0556
Baik	28	2	0,4516	0,1111
Cukup	16	8	0,2581	0,4444
Kurang	1	5	0,0161	0,2778
Sangat Kurang	0	2	0,0000	0,1111
Jumlah	62	18	1	1
Total Data		80		

4. Parameter Nilai Bahasa Indonesia (A4).

$$P(\text{Sangat Baik} | \text{Diterima}) = \frac{7}{62} = 0,1129$$

$$P(\text{Sangat Baik} | \text{Tidak}) = \frac{1}{18} = 0,0556$$

$$P(\text{Baik} | \text{Diterima}) = \frac{31}{62} = 0,5$$

$$P(\text{Baik} | \text{Tidak Diterima}) = \frac{5}{18} = 0,2778$$

$$P(\text{Cukup} | \text{Diterima}) = \frac{14}{62} = 0,2258$$

$$P(\text{Cukup} | \text{Tidak}) = \frac{0}{18} = 0$$

$$P(\text{Kurang} | \text{Diterima}) = \frac{10}{62} = 0,1613$$

$$P(\text{Kurang} | \text{Tidak}) = \frac{8}{18} = 0,4444$$

$$P(\text{Sangat Kurang} | \text{Diterima}) = \frac{0}{62} = 0$$

$$P(\text{Sangat Kurang} | \text{Tidak}) = \frac{4}{18} = 0,2222$$

Hasil perhitungan probabilitas pada parameter Bahasa Indonesia dengan kategori A4 dapat dilihat pada tabel 8.berikut:

Tabel 8. Hasil Perhitungan Probabilitas Kategori A4

Kategori : A4	Diterima	Tidak	P. Diterima	P.Tidak
Sangat Baik	7	1	0,1129	0,0556
Baik	31	5	0,5000	0,2778
Cukup	14	0	0,2258	0,0000
Kurang	10	8	0,1613	0,4444
Sangat Kurang	0	4	0,0000	0,2222
Jumlah	62	18	1	1
Total Data		80		

3.1.3 Baca Data Testing

Data uji yang digunakan dalam penelitian ini terdiri dari 30 data responden. Berikut adalah data uji yang digunakan dapat dilihat pada tabel 9.

Tabel 9. Data Testing

No	Nama	Kriteria				
		A1	A2	A3	A4	Hasil
1	Siswa 1	B	C	C	C	?
2	Siswa 2	B	B	B	B	?
3	Siswa 3	B	B	B	B	?
4	Siswa 4	SB	B	B	C	?
5	Siswa 5	B	B	B	B	?
6	Siswa 6	B	B	B	B	?
7	Siswa 7	B	B	B	C	?
8	Siswa 8	B	B	B	B	?
9	Siswa 9	SB	B	B	B	?
10	Siswa 10	SB	B	B	B	?
...
27	Siswa 27	B	B	B	B	?
28	Siswa 28	B	C	C	B	?
29	Siswa 29	B	B	B	B	?
30	Siswa 30	B	C	C	B	?



3.1.4 Menghitung Probabilitas Data Testing

Berdasarkan data *training* berikut merupakan perhitungan nilai Diterima dari data 81 sampai 110 dengan jumlah 30 mahasiswa. Sehingga untuk mendapatkan nilai dilakukan proses perhitungan sebagai berikut.

$$\begin{aligned} P(M1|Puas) &= P(XA1=Baik|Diterima) \times P(XA2=Cukup|Diterima) \times P(XA3=Cukup|Diterima) \times P(XA4=Cukup|Diterima) \\ &= 0,5968 \times 0,1290 \times 0,2581 \times 0,2258 = 0,00448673 \end{aligned}$$

$$\begin{aligned} P(M2|P_{Diterima}) &= P(XA1=Baik|Diterima) \times P(XA2=Baik|Diterima) \times P(XA3=Baik|Diterima) \times P(XA4=Baik|Diterima) \\ &= 0,5968 \times 0,6290 \times 0,4516 \times 0,5 = 0,08476 \end{aligned}$$

Lakukan Perhitungan samapai dengan M30.

Sedangkan untuk menghitung nilai Tidak diterima data 80 sampai ke 110. Rumus yang digunakan sama dengan rumus untuk menentukan nilai tidak diterima. Sehingga untuk mendapatkan nilai dilakukan sebagai berikut.

$$\begin{aligned} P(R1|Tidak Puas) &= P(XA1=Baik|Tidak) \times P(XA2=Cukup|Tidak) \times P(XA3=Cukup|Tidak) \times P(XA4=Cukup|Tidak) \\ &= 0,4444 \times 0,0000 \times 0,3889 \times 0,2778 = 0 \end{aligned}$$

$$\begin{aligned} P(R2|Tidak Puas) &= P(XA1=Baik|Tidak) \times P(XA2=Baik|Tidak) \times P(XA3=Baik|Tidak) \times P(XA4=Baik|Tidak) \\ &= 0,4444 \times 0,3889 \times 0,1111 \times 0,2778 = 0,005334 \end{aligned}$$

Lakukan Perhitungan samapai dengan M30

3.1.5 Menghitung Probabilitas Akhir

Probabilitas akhir merupakan pemaksimalan klasifikasi Diterima dan tidak:

$$\begin{aligned} P(Diterima|X) &= P(X|Diterima) * P(C|Diterima) \\ &= 0,00448673 \times 0,775 \end{aligned}$$

$$= 0,00347721575$$

$$\begin{aligned} P(Tidak|X) &= P(X|Tidak) * P(C|Tidak) \\ &= 0 \times 0,225 \\ &= 0 \end{aligned}$$

3.2 Implementasi

Setelah mengetahui jumlah responden prediksi puas dan tidak puas berikutnya adalah proses pembuktian nilai probabilitas di atas akan diuji dengan data sebanyak 30 data dan diselesaikan dengan menggunakan *tools Rapid Miner 5.3* sehingga diperoleh hasil pengujian probabilitas *confidence klasifikasi* seperti yang terlihat pada gambar 2.

Row No.	PROSES	A1	A2	A3	A4
1	Diterima	Baik	Cukup	Cukup	Cukup
2	Diterima	Baik	Baik	Baik	Baik
3	Diterima	Baik	Baik	Baik	Baik
4	Diterima	Sangat Baik	Baik	Baik	Cukup
5	Tidak	Baik	Baik	Baik	Baik
6	Tidak	Baik	Baik	Baik	Baik
7	Diterima	Baik	Baik	Baik	Cukup
8	Tidak	Baik	Baik	Baik	Baik
9	Diterima	Sangat Baik	Baik	Baik	Baik
10	Diterima	Sangat Baik	Baik	Baik	Baik
11	Diterima	Baik	Baik	Sangat Baik	Baik
12	Tidak	Baik	Cukup	Baik	Baik
13	Diterima	Baik	Baik	Baik	Baik

Gambar 2. Hasil Pengujian Nilai Probabilitas Data Testing dengan *Rapid Miner*

Berdasarkan gambar 2 terdapat empat siswa dengan prediksi tidak diterima dan 26 siswa dinyatakan diterima terhadap proses yang telah diuji menggunakan *tools*. Setelah melihat bukti pengujian hasil nilai probabilitas selanjutnya melihat seberapa besar akurasi yang didapatkan pada proses prediksi yang telah dilakukan. Akurasi Hasil pengujian Model Algoritma Naive Bayes *Classifier* dapat dilihat pada gambar 3.

accuracy: 76.67%	
pred. Diterima	true Diterima
pred. Diterima	23
pred. Tidak	4
pred. Tidak	3
pred. Tidak	0
class recall	88.46%
class precision	85.19%
0.00%	0.00%
0.00%	0.00%

Gambar 3. Nilai Accuracy Performance

Keterangan :



- a. Pada gambar diatas Nilai *Accuracy* sebesar 76,6 %. Dengan Demikian Algoritma Naive bayes dapat diterapkan pada prediksi penerimaan murid baru dengan *Accuracy* mendekati 80 %.
- b. Data pengujian sebanyak 30 data *testing* yang diolah di dalam *RapidMiner 5.3*. mendapatkan hasil pengujian dengan akurasi sebesar 76,6% yaitu 26 siswa diterima dan sebanyak 4 tidak diterima. Sehingga proses dalam melakukan prediksi lebih cepat dan akurat serta dapat menghasilkan tingkat akurasi dengan nilai yang didapatkan.

3.3 Hasil Akhir

Hasil akhir dari proses prediksi berupa klasifikasi secara keseluruhan berdasarkan data yang telah valid dan diuji menggunakan pemodelan *tools Rapid Miner 5.3* dapat dilihat pada tabel 10.

Tabel 10. Hasil Akhir

Nama Siswa	A1	A2	A3	A4	Hasil
Riza	Baik	Cukup	Cukup	Cukup	Diterima
Anastasia	Baik	Baik	Baik	Baik	Diterima
victoria	Baik	Baik	Baik	Baik	Diterima
Sri Bela	Sangat Baik	Baik	Baik	Cukup	Diterima
Ira	Baik	Baik	Baik	Baik	Diterima
Atikah	Baik	Baik	Baik	Baik	Diterima
Putri	Baik	Baik	Baik	Cukup	Diterima
Chrisna	Baik	Baik	Baik	Baik	Diterima
Gaudentius	Sangat Baik	Baik	Baik	Baik	Diterima
Ayu	Sangat Baik	Baik	Baik	Baik	Diterima
Noni	Baik	Baik	Sangat Baik	Baik	Diterima
Nurul	Baik	Cukup	Baik	Baik	Diterima
Gandhi Pradana	Baik	Baik	Baik	Baik	Diterima
Irfan	Baik	Baik	Baik	Baik	Diterima
Zulpadly Harahap	Baik	Baik	Baik	Baik	Diterima
Kurniawan	Sangat Baik	Baik	Baik	Cukup	Diterima
Adelia	Baik	Cukup	Cukup	Kurang	Tidak
Timbul	Baik	Cukup	Baik	Kurang	Tidak
Agung	Baik	Cukup	Cukup	Sangat Kurang	Tidak
siti	Baik	Baik	Baik	Cukup	Diterima
Wulansari	Baik	Baik	Baik	Cukup	Diterima
Tidki	Sangat Baik	Baik	Kurang	Kurang	Tidak
Abdul Aziz	Baik	Baik	Baik	Cukup	Diterima
Karina	Baik	Baik	Baik	Cukup	Diterima
Gideon	Baik	Baik	Baik	Kurang	Diterima
Lio	Cukup	Cukup	Baik	Baik	Diterima
ningsih	Baik	Baik	Baik	Baik	Diterima
Bobby	Baik	Cukup	Cukup	Baik	Diterima
Novauli	Baik	Baik	Baik	Baik	Diterima
Susi	Baik	Cukup	Cukup	Baik	Diterima

5. KESIMPULAN

Berdasarkan pengujian sebanyak 30 data testing yang diolah menggunakan Rapid Miner, diperoleh tingkat akurasi sebesar 86,6% yaitu 26 siswa diterima dan sebanyak 4 tidak diterima. Sehingga dapat disimpulkan bahwa proses prediksi berupa klasifikasi dengan menggunakan Naive Bayes dapat lebih cepat dan akurat serta menghasilkan tingkat akurasi yang tinggi bila diterapkan untuk mengatasi masalah klasifikasi jumlah siswa baru pada SMK Anak Bangsa.

REFERENCES

- [1] K. H. Asri, A. Komariah, D. Meirawan, and D. A. Kurniady, "Kepemimpinan Kepala Sekolah Dalam Penyerapan Lulusan Berbasis Industri," *Research and Development Journal of Education*, vol. 7, no. 1, pp. 1–10, 2021.
- [2] A. Akrim, *Menjadi Generasi Pemimpin Apa Yang Dilakukan Sekolah?* 2019.
- [3] J. E. Falabitsa and L. Lelah, "Pengembangan Sistem Seleksi Penerimaan Calon Siswa Baru Dengan Metode Simple



- [4] Additive Weighting,” *Jurnal Ilmiah Informatika (JIF)*, vol. 8, no. 2, pp. 113–119, 2020.
- [4] A. Yusnita, S. Lailiyah, and K. Saumahudi, “Penerapan Algoritma Naïve Bayes Untuk Penerimaan Peserta Didik Baru,” *Jurnal Informatika Wicida*, vol. 10, no. 1, pp. 11–16, 2021.
- [5] S. R. D. Wibowo, D. M. Midyanti, and R. Hidayati, “Penerapan Metode Grey Relational Analysis Pada Penerimaan Pengajar Yayasan Pendidikan Sekolah Bruder Kota Pontianak,” *Coding : Jurnal Komputer dan Aplikasi*, vol. 08, no. 01, pp. 102–111, 2020.
- [6] A. Wanto, D. Suhendro, and A. P. Windarto, “Pelatihan dan Bimbingan dalam Pemanfaatan Internet yang Baik dan Aman bagi Pelajar SMK Anak Bangsa Desa Bandar Siantar Kabupaten Simalungun,” *E-Dimas: Jurnal Pengabdian kepada Masyarakat*, vol. 9, no. 2, pp. 149–157, 2018.
- [7] T. H. Apandi and C. A. Sugianto, “Analisis Komparasi Machine Learning Pada Data Spam Sms,” *Jurnal TEDC*, vol. 12, no. 1, pp. 58–62, 2018.
- [8] D. Y. Utami, E. Nurlelah, and N. Hikmah, “Attribute Selection in Naive Bayes Algorithm Using Genetic Algorithms and Bagging for Prediction of Liver Disease,” *Journal of Informatics and Telecommunication Engineering*, vol. 4, no. 1, pp. 76–85, 2020.
- [9] A. R. Damanik *et al.*, “Prediksi Tingkat Kematian di Indonesia Akibat Covid-19 Menggunakan Algoritma Naïve Bayes,” *SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 1, pp. 73–83, 2022.
- [10] T. H. Apandi and C. A. Sugianto, “Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP (Naive Bayes Algorithm for Satisfaction Prediction of e-ID),” *JUITA (Jurnal Informatika) UMP*, vol. 7, no. 2, pp. 125–128, 2019.
- [11] M. F. Rifai, H. Jatnika, and B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 12, no. 2, pp. 131–144, 2019.
- [12] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creative Information Technology Journal*, vol. 6, no. 1, pp. 1–11, 2020.
- [13] E. Indrayuni, “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *Jurnal Khatulistiwa Informatika*, vol. 7, no. 1, pp. 29–36, 2019.
- [14] R. Bahtiar, M. D. S. T, A. Setiawan, and P. Rosyani, “Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm),” *Jurnal Kreativitas Mahasiswa Informatika*, vol. 1, no. 2, pp. 99–103, 2020.
- [15] S. E. Y. Putri, Saniati, and A. Surahman, “Penerapan Model Naive Bayes Untuk Memprediksi Potensi Pendaftaran Siswa di SMK Taman Siswa Teluk Belitung Berbasis Web,” *Jurnal JECSET*, vol. 1, no. 1, pp. 82–87, 2021.
- [16] amik tunas bangsa and S. Utara, “Teknik Data Mining Dalam Prediksi Jumlah Siswa Baru Dengan Algoritma Naive Bayes,” *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 2, no. 2, pp. 108–117, 2021.
- [17] D. Alita, I. Sari, and A. Rahman Isnain, “Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa,” *Jurnal Data Mining dan Sistem Informasi (JDMSI)*, vol. 2, no. 1, pp. 17–23, 2021.