

# Implementasi *Data Mining* Untuk Mengklasifikasi Hasil Belajar Siswa/i Dengan Metode *Naïve Bayes*

Agus Riyanto<sup>a</sup>, Elvis Sastra Ompusunggu<sup>b</sup>

<sup>a,b</sup>*Fakultas Sains dan Teknologi, Universitas Prima Indonesia*

*Corresponding Author:*

<sup>a</sup>*Sastraelvis@gmail.com*

## ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasi hasil belajar siswa/i dan melihat tingkat keakurasian metode penelitian yang digunakan. Penelitian dilaksanakan di MAS Amaliyah Sunggal pada tahun 2024. Penelitian ini menerapkan *data mining* menggunakan metode algoritma *Naïve Bayes* dan pengujian tingkat keakurasian metode *Naïve Bayes* dengan RapidMiner. Data yang digunakan terdiri dari data sekunder yang diperoleh langsung dari pihak sekolah dan data dari penyebaran angket. Populasinya adalah seluruh siswa/i kelas III SMA T.A. 2023/2024 yang berjumlah 151 orang. Teknik pengambilan sampel yang digunakan adalah *simple random sampling* dengan besar sampel sebanyak 60 orang yang akan dijadikan sebagai *dataset*. Atribut yang digunakan berjumlah 10 dan 1 atribut *class* untuk klasifikasi. Analisis data dilakukan dengan persamaan teorema *bayes* dengan 60 data *training* dan 1 data *testing*. Hasil analisis menunjukkan nilai probabilitas tertinggi berada pada *class* P(H=Sangat Baik) sebesar 0,4061516 yang dapat disimpulkan bahwa klasifikasi hasil belajar pada nilai rapor T.A. 2023/2024 dikategorikan sangat baik. Berdasarkan hasil pengujian tingkat keakurasian metode *Naïve Bayes* dengan RapidMiner menggunakan ketentuan 70% data *training* dan 30% data *testing* menunjukkan nilai akurasi sebesar 94,44% yang dapat diartikan bahwa metode *Naïve Bayes* cukup baik digunakan untuk melakukan klasifikasi hasil belajar siswa/i.

**Kata Kunci:** *Data mining*, *Naïve Bayes*, Klasifikasi, RapidMiner

## ABSTRACT

This study aims to classify student learning outcomes and determine the accuracy of the research methods. The research was conducted at MAS Amaliyah Sunggal in 2024. This research applies data mining using the Naïve Bayes algorithm method and testing the accuracy of the Naïve Bayes method with RapidMiner. The data used consists of secondary data obtained directly from the school and data from distributing questionnaires. The population is all third-grade high school students T.A. 2023/2024, which amounted to 151 people. The sampling technique used is simple random sampling with a sample size of 60 people, which will be used as a dataset. The attributes used amounted to 10 and 1 class attribute for classification. Data analysis is done with the Bayes theorem equation with 60 training data and 1 testing data. The analysis results show that the highest probability value is in the P (H = Very Good) class of 0.4061516, which can be concluded by the classification of learning outcomes on report cards T.A. 2023/2024, categorized as very good. Based on the

results of testing the level of accuracy of the Naïve Bayes method with RapidMiner using the provisions of 70% training data and 30% testing data, it shows an accuracy value of 94.44%, which means that the Naïve Bayes method is good enough to be used to classify student learning outcomes.

**Keywords:** Data Mining, Naïve Bayes, Classification, RapidMiner

## PENDAHULUAN

Pendidikan merupakan pengalaman belajar terencana yang berlangsung seumur hidup yang mencakup pendidikan formal, non-formal, dan in-formal baik di dalam maupun diluar sekolah dengan tujuan untuk memaksimalkan kemampuan individu (Ahdar, 2021). Untuk dapat mengetahui pendidikan telah mencapai tujuannya atau tidak, dilakukannya kegiatan evaluasi yang menjadi salah satu kegiatan yang sangat penting dalam dunia pendidikan dengan melakukan penilaian terhadap hasil belajar siswa (Hidayat & Abdillah, 2019). Dimana hasil belajar siswa/i akan menjadi tolak ukur kemampuan peserta didik dalam menguasai satu atau lebih mata pelajaran (Firmansyah & Yulianto, 2023).

Salah satu masalah yang terdapat di sekolah MAS Amaliyah Sunggal adalah adanya perbedaan hasil belajar pada setiap siswa/i. Dimana setiap semesternya siswa/i akan memperoleh hasil belajar secara berulang-ulang terhadap mata pelajaran yang diajarkan. Akibatnya data yang ada di sekolah seiring berjalannya waktu terus bertambah sehingga terjadinya penumpukkan data. Maka sekolah perlu melakukan pengkajian lebih dini mengenai hasil belajar siswa/i agar dapat menentukan dengan tepat pembelajaran dan bimbingan yang sesuai dengan kemampuan siswa/i dengan harapan hal ini dapat menjadiantisipasi terhadap siswa/i yang berpotensi mengalami hambatan atau kurang prestasi dalam belajar.

Perbedaan hasil belajar dapat terjadi karena dipengaruhi beberapa faktor. Menurut (Basri et al., 2022) adanya perbedaan dari hasil belajar dapat terjadi akibat pengaruh dari beberapa faktor, baik faktor internal maupun eksternal. Dimana menurut (Gunawan et al., 2019), faktor internal adalah faktor yang berasal dari dalam diri peserta didik itu sendiri seperti kecerdasan, minat, bakat, perhatian, komitmen, dan lainnya. Sedangkan faktor eksternal adalah faktor yang berasal dari luar diri peserta didik seperti jenis kelamin, umur, pendidikan orang tua, ekonomi keluarga dan lainnya.

Salah satu cara yang dapat dilakukan untuk mengetahui adanya perbedaan hasil belajar siswa/i adalah dengan mengklasifikasi hasil belajar siswa/i menggunakan algoritma *Naïve Bayes*. Klasifikasi merupakan sebuah model *data mining* yang digunakan untuk memprediksi kategori label dari suatu objek (Wanto et al., 2020). *Data mining* merupakan alat bantu yang digunakan untuk menggali data dalam ukuran besar dan mengubahnya menjadi sebuah informasi yang bermanfaat agar dapat mempermudah pengambilan keputusan dan membantu dalam pemecahan masalah (Triwidianti et al., 2021). Algoritma *Naïve Bayes* adalah salah satu klasifikasi *data mining* yang digunakan dalam pemecahan masalah yang merupakan algoritma sederhana dimana tiap atributnya bersifat bebas dan memungkinan tiap atribut dapat berkontribusi terhadap hasil akhir (Marpaung et al., 2021). Kemudian untuk dapat melihat tingkat keakurasian metode *Naïve Bayes* dalam mengklasifikasi hasil belajar siswa/i

akan dilakukan pengujian dengan RapidMiner. Dimana RapidMiner merupakan perangkat lunak yang digunakan untuk mengolah suatu data dengan berbagai teknik dan metode dalam *data mining* sehingga data dapat menjadi informasi yang berguna (Wanto et al., 2020). Maka dengan itu, penelitian ini diharapkan dapat memberikan kontribusi dalam menggali informasi yang akurat untuk membantu pengambilan keputusan terkait hasil belajar siswa agar dapat menentukan dengan tepat pembelajaran dan bimbingan yang sesuai.

## METODE

### Jenis Penelitian

Penelitian ini dilakukan dengan mengimplementasi *data mining* untuk mengklasifikasi hasil belajar siswa/i menggunakan metode algoritma *Naïve Bayes* dan pengujian tingkat keakurasian metode *Naïve Bayes* dalam mengklasifikasi hasil belajar menggunakan RapidMiner.

### Tempat dan waktu Penelitian

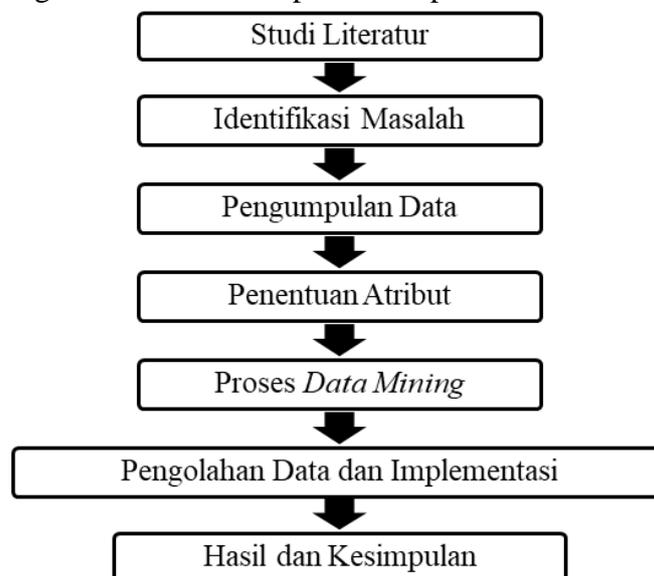
Penelitian akan dilaksanakan di MAS Amaliyah Sunggal dan waktu penelitian dilaksanakan pada bulan Januari – Februari tahun 2024.

### Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh siswa/i kelas III SMA Tahun Ajaran (T.A.) 2023/2024 yang berjumlah 151 orang. Besar sampel dihitung menggunakan rumus Slovin, 1991 sehingga diperoleh hasil perhitungan sebanyak 60 orang yang akan dijadikan sebagai *dataset*. Teknik pengambilan sampel dalam penelitian ini adalah *simple random sampling*.

### Tahap Penelitian

Tahapan penelitian yang akan dilakukan dapat dilihat pada skema 1.

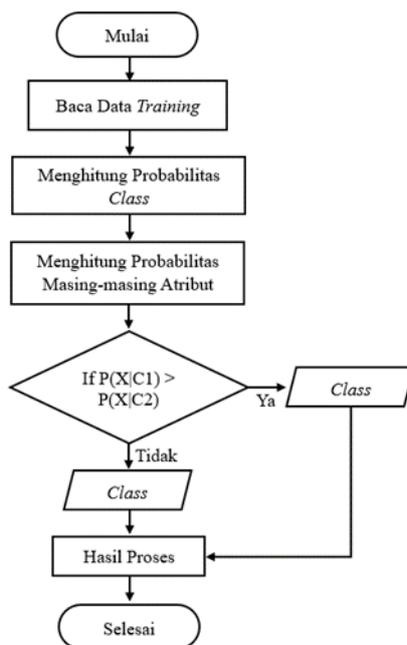


**Gambar 1. Skema Tahap Penelitian**

1. Studi literatur, tahap pertama yang dilakukan peneliti yaitu melakukan studi literatur dengan mengumpulkan beberapa referensi yang terkait dengan *data mining*, algoritma

- Naïve Bayes* dan referensi lainnya terkait penelitian yang kemudian referensi tersebut dibaca dan dipelajari oleh peneliti.
2. Identifikasi masalah, selanjutnya peneliti melakukan identifikasi masalah untuk dapat mengenal dan menentukan batasan masalah yang akan diteliti sehingga dapat membantu peneliti dalam menentukan solusi untuk pemecahan masalah tersebut. Dimana masalah dalam penelitian ini adalah untuk mengklasifikasi hasil belajar siswa/i dengan metode *Naïve Bayes*.
  3. Pengumpulan data, pengumpulan sejumlah data dilakukan peneliti dengan menyebarkan angket yang kemudian langsung diisi oleh siswa/i di MAS Amaliyah Sunggal. Data yang diperoleh berupa pendidikan orang tua, pekerjaan orang tua dan penghasilan orang tua. Selain itu, peneliti juga menggunakan data sekunder yang diperoleh langsung dari pihak sekolah berupa data diri siswa/i kelas III SMA Tahun Ajaran (T.A.) 2023/2024 dan hasil belajar siswa/i dari T.A. 2021/2022 atau tahun pertama, T.A. 2022/2023 atau tahun kedua dan T.A. 2023/2024 atau tahun ketiga.
  4. Penentuan atribut, Dimana atribut yang akan digunakan dalam mengklasifikasi hasil belajar siswa/i terdiri dari 10 atribut jenis kelamin, tahun lahir, pendidikan terakhir ayah, pekerjaan ayah, penghasilan ayah, pendidikan terakhir ibu, pekerjaan ibu, penghasilan ibu, rata-rata nilai rapor tahun pertama, rata-rata nilai rapor tahun kedua dan 1 atribut class untuk klasifikasi yaitu rata-rata nilai rapor tahun ketiga yang diklasifikasikan. Dimana *class* klasifikasi dikategorikan menjadi 4, yaitu:
    - a. Sangat Baik (SB) jika nilai rata-rata rapor 90 – 100.
    - b. Baik (B) jika nilai rata-rata rapor 80 – 89.
    - c. Cukup (C) jika nilai rata-rata rapor 70 – 79.
    - d. Kurang (D) jika nilai rata-rata rapor <69.
  5. Proses *data mining*, dimana proses ini dibagi menjadi beberapa tahapan proses yang sesuai dengan proses *Knowledge Discovery in Database* (KDD) diantaranya sebagai berikut (Muslim et al., 2019).
    - a. Pembersihan data, tahap awal proses KDD adalah pembersihan data yang dilakukan terhadap *noise* yang ditemukan berupa *missing value*, *inkonsisten data* dan *redundant data*. Seluruh atribut akan dibersihkan terlebih dahulu untuk mendapatkan atribut-atribut yang relevan, tidak *missing value* dan *redundant*. Dimana untuk *missing value*, akan dilakukan teknik penghapusan baris terhadap atribut data yang tidak tersedia atau kosong. Hal ini dilakukan karena total data yang tersedia cukup besar dibandingkan dengan jumlah data yang kosong.
    - b. Penyeleksian data, tahap selanjutnya berupa penyeleksian data dengan tujuan untuk memperoleh data yang sebenar-benarnya dengan melihat atribut mana saja yang diperlukan dalam penelitian dan membuang atribut yang tidak dibutuhkan. Sehingga dalam seleksi data, akan dilakukan teknik pemilihan data yang sesuai dengan atribut yang telah ditentukan untuk dijadikan *dataset* dan selanjutnya dianalisis.
    - c. Integritas data, selanjutnya dilakukannya integritas data berupa penggabungan data terhadap data yang telah bersih dari *missing value*, *inkonsisten data* dan *redundant data* menjadi satu yang relevan kemudian akan digunakan sebagai data akhir siswa/i yang sesuai dengan kebutuhan.

- d. Transformasi data, tahap dimana data yang telah bersih dan relevan, kemudian akan ditransformasikan menjadi 2 data yaitu data *training* dan data *testing* agar dapat diolah menggunakan *data mining*.
6. Pengolahan data dan implementasi, dimana dalam penelitian ini data akan diolah dengan *data mining* menggunakan metode *Naïve Bayes* untuk mengklasifikasi hasil belajar siswa/i. Alur metode *Naïve Bayes* yang dilakukan dalam penelitian dapat dilihat pada skema 2.



**Gambar 2. Skema Alur Metode Naïve Bayes**

Rumus persamaan teorema *bayes* yang digunakan sebagai berikut (Muslim et al., 2019).

$$P(E) = \frac{P(H)P(H)}{P(E)}$$

Keterangan :

H = Hipotesis data E merupakan suatu *class spesifik*

E = Data dengan *class* yang belum diketahui

$P(E)$  = Probabilitas hipotesis H berdasarkan kondisi E (*posteriori probability*)

$P(H)$  = Probabilitas E berdasarkan kondisi hipotesis H (*likelihood*)

$P(H)$  = Probabilitas dari hipotesis H (*prior probability*)

$P(E)$  = Probabilitas dari E (*predictor prior probability*)

Setelah dilakukannya perbandingan hasil *class* dan dapat ditentukan nilai tertinggi, kemudian dilakukan implementasi untuk pengujian tingkat keakurasian dari metode *Naïve Bayes* dalam mengklasifikasi hasil belajar menggunakan *software* RapidMiner.

7. Hasil dan kesimpulan, dimana tahap ini dilakukan untuk menampilkan hasil yang telah digali peneliti dari penggunaan metode *Naïve Bayes* dan pengujian menggunakan RapidMiner. Setelah itu peneliti mengambil kesimpulan apakah algoritma *Naïve Bayes* efektif atau tidak dengan melihat nilai akurasi yang diperoleh dalam mengklasifikasi hasil belajar siswa/i.

## HASIL

### Pengolahan Data Dengan *Naïve Bayes*

Dalam klasifikasi hasil belajar, dalam penelitian ini akan dilakukannya pengolahan dan analisis data dengan metode *Naïve Bayes*. Yang mana data yang digunakan terbagi menjadi dua bagian yaitu sebanyak 60 data *training* dan 1 data *testing*. Pengolahan data dilakukan dengan perhitungan secara manual dengan persamaan teorema *bayes*. Alur metode *Naïve Bayes* sebagai berikut (lihat pada skema 2).

a. Baca data *training* pada tabel 1.

b. Menghitung probabilitas *class*

Untuk klasifikasi hasil belajar akan ditemukan 3 *class* yaitu “Sangat Baik (SB)”, “Baik (B)” dan “Cukup (C)”. Maka untuk dapat menghitung probabilitas *class*, hitunglah jumlah data yang sesuai dengan *class* lalu dibagi dengan total keseluruhan data *training* yang digunakan. Jumlah data yang digunakan dalam penelitian ini sebanyak 60 data. Maka didapatkan perhitungan sebagai berikut.

$$P(H = \text{Sangat Baik (SB)}) = 23/60 = 0,38$$

$$P(H = \text{Baik (B)}) = 36/60 = 0,60$$

$$P(H = \text{Cukup (C)}) = 1/60 = 0,02$$

c. Menghitung probabilitas masing-masing atribut

Untuk mencari probabilitas masing-masing atribut, cara yang dilakukan adalah dengan membandingkan atribut dari data *testing* (lihat pada tabel 2) dengan atribut dari data *training*. Yang mana jumlah atribut yang ada pada data *training* akan dibagi dengan jumlah data pada *class*. Maka didapatkan perhitungan sebagai berikut.

1) Atribut Jenis Kelamin (JK)

Pada data *testing*, jenis kelamin adalah Perempuan (P).

$$P(JK = P | H = \text{SB}) = 22/23 = 0,96$$

$$P(JK = P | H = \text{B}) = 21/36 = 0,58$$

$$P(JK = P | H = \text{C}) = 0/1 = 0$$

2) Atribut Tahun Lahir (TL)

Pada data *testing*, tahun lahir adalah 2006.

$$P(TL = 2006 | H = \text{SB}) = 16/23 = 0,70$$

$$P(TL = 2006 | H = \text{B}) = 25/36 = 0,70$$

$$P(TL = 2006 | H = \text{C}) = 1/1 = 1$$

3) Atribut Pendidikan Terakhir Ayah (PTA)

Pada data *testing*, pendidikan terakhir ayah adalah Sekolah Menengah Atas (SMA).

$$P(PTA = \text{SMA} | H = \text{SB}) = 11/23 = 0,48$$

$$P(PTA = \text{SMA} | H = \text{B}) = 26/36 = 0,72$$

$$P(PTA = \text{SMA} | H = \text{C}) = 1/1 = 1$$

4) Atribut Pekerjaan Ayah (PKA)

Pada data *testing*, pekerjaan ayah adalah wiraswasta (WS).

$$P(PKA = \text{WS} | H = \text{SB}) = 15/23 = 0,65$$

$$P(PKA = \text{WS} | H = \text{B}) = 20/36 = 0,50$$

$$P(PKA = \text{WS} | H = \text{C}) = 1/1 = 1$$

5) Atribut Penghasilan Ayah (PHA)

Pada data testing, penghasilan ayah adalah Rp 1.000.000 – 2.000.000.

$$P(\text{PHA} = \text{Rp } 1.000.000 - 2.000.000 \mid H = \text{SB}) = 7/23 = 0,31$$

$$P(\text{PHA} = \text{Rp } 1.000.000 - 2.000.000 \mid H = \text{B}) = 8/36 = 0,20$$

$$P(\text{PHA} = \text{Rp } 1.000.000 - 2.000.000 \mid H = \text{C}) = 0/1 = 0$$

6) Atribut Pendidikan Terakhir Ibu (PTI)

Pada data testing, pendidikan terakhir ibu adalah Sekolah Menengah Atas (SMA).

$$P(\text{PTI} = \text{SMA} \mid H = \text{SB}) = 15/23 = 0,65$$

$$P(\text{PTI} = \text{SMA} \mid H = \text{B}) = 20/36 = 0,50$$

$$P(\text{PTI} = \text{SMA} \mid H = \text{C}) = 1/1 = 1$$

7) Atribut Pekerjaan Ibu (PKI)

Pada data testing, pekerjaan ibu adalah Ibu Rumah Tangga (IRT).

$$P(\text{PKI} = \text{IRT} \mid H = \text{SB}) = 17/23 = 0,74$$

$$P(\text{PKI} = \text{IRT} \mid H = \text{B}) = 24/36 = 0,50$$

$$P(\text{PKI} = \text{IRT} \mid H = \text{C}) = 1/1 = 1$$

8) Atribut Penghasilan Ibu (PHI)

Pada data testing, penghasilan ibu adalah Tidak Berpenghasilan (TB).

$$P(\text{PHI} = \text{TB} \mid H = \text{SB}) = 16/23 = 0,70$$

$$P(\text{PHI} = \text{TB} \mid H = \text{B}) = 16/36 = 0,40$$

$$P(\text{PHI} = \text{TB} \mid H = \text{C}) = 0/1 = 0$$

9) Atribut rata-rata nilai rapor tahun pertama (T1) dan tahun kedua (T2)

Perhitungan mean pada data numerik:

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Ket :  $\mu$  = rata-rata hitung (mean)

$x_n$  = nilai sampel ke - n

n = jumlah sampel

a) Rata-rata T1

Sangat Baik = 85,80

Baik = 81,67

Cukup = 77,41

b) Rata-rata T2

Sangat Baik = 88,16

Baik = 84,06

Cukup = 80,89

Perhitungan standar deviasi data numerik:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

Ket :  $\sigma$  = standar deviasi

$x_i$  = nilai sampel ke - i

$\mu$  = rata-rata hitung (mean)

n = jumlah sampel

- a) Rata-rata T1  
Sangat Baik = 2,26  
Baik = 2,15  
Cukup = 0
- b) Rata-rata T2  
Sangat Baik = 1,57  
Baik = 1,74  
Cukup = 0

Perhitungan probabilitas data numerik:

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

Ket : P = probabilitas (peluang)

$X_i$  = atribut ke i

$x_i$  = nilai atribut ke i

$Y_j$  = class yang dicari

$y_j$  = subclass y yang dicari

$\mu$  = rata-rata hitung (mean)

$\sigma$  = standar deviasi

$\pi = 3,14$

$e = 2,718281$

- a) Rata-rata T1

$$P(T1 = 80 | H = \text{Sangat Baik}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 2,26}} x e^{-\frac{(80-85,80)^2}{2 \cdot (2,26)^2}} = 0,011$$

$$P(T1 = 80 | H = \text{Baik}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 2,15}} x e^{-\frac{(80-81,67)^2}{2 \cdot (2,15)^2}} = 0,1860$$

$$P(T1 = 80 | H = \text{Cukup}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 0}} x e^{-\frac{(80-77,41)^2}{2 \cdot (0)^2}} = 0$$

- b) Rata-rata T2

$$P(T2 = 85 | H = \text{Sangat Baik}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 1,57}} x e^{-\frac{(85-88,16)^2}{2 \cdot (1,57)^2}} = 0,0420$$

$$P(T2 = 85 | H = \text{Baik}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 1,74}} x e^{-\frac{(85-84,06)^2}{2 \cdot (1,74)^2}} = 0,26142$$

$$P(T2 = 85 | H = \text{Cukup}) = \frac{1}{\sqrt{2 \cdot 3,14 \cdot 0}} x e^{-\frac{(85-80,89)^2}{2 \cdot (0)^2}} = 0$$

- d. Mengalikan semua probabilitas akhir dari setiap class

$$\begin{aligned} 1) P(E | H = SB) &= P(P|SB) \times P(2006|SB) \times P(SMA|SB) \times P(WS|SB) \times P(Rp\ 1.000.000 \\ &\quad -\ 2.000.000|SB) \times P(SMA|SB) \times P(IRT|SB) \times P(TB|SB) \times P(80|SB) \\ &\quad \times P(85|SB) \\ &= 0,96 \times 0,70 \times 0,48 \times 0,65 \times 0,31 \times 0,65 \times 0,74 \times 0,74 \times 0,011 \times 0,042 \\ &= 1,06882 \end{aligned}$$

Maka,  $P(H = SB | E) = P(E | H = SB) P(H = SB) = 1,06882 \times 0,38 = 0,4061516$

$$2) P(E | H = B) = P(P|B) \times P(2006|B) \times P(SMA|B) \times P(WS|B) \times P(Rp\ 1.000.000$$

$$\begin{aligned}
 & - 2.000.000|B) \times P(SMA|B) \times P(IRT|B) \times P(TB|B) \times P(80|B) \times \\
 & P(85|B) \\
 & = 0,58 \times 0,70 \times 0,72 \times 0,50 \times 0,20 \times 0,50 \times 0,60 \times 0,40 \times 0,186 \times 0,261 \\
 & = 0,01703
 \end{aligned}$$

Maka,  $P(H = B | E) = P(E | H = B) P(H = B) = 0,01703 \times 0,60 = 0,010218$

$$\begin{aligned}
 3) P(E | H = C) & = P(P|C) \times P(2006|C) \times P(SMA|C) \times P(WS|C) \times P(Rp 1.000.000 \\
 & - 2.000.000|C) \times P(SMA|C) \times P(IRT|C) \times P(TB|C) \times P(80|C) \times \\
 & P(85|C) \\
 & = 0 \times 1 \times 1 \times 1 \times 0 \times 1 \times 1 \times 0 \times 0 \times 0 = 0
 \end{aligned}$$

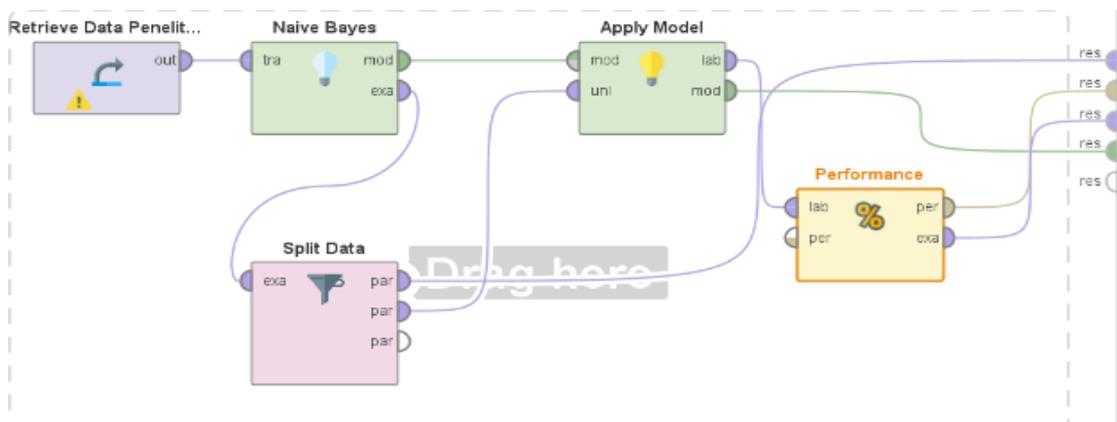
Maka,  $P(H = C | E) = P(E | H = C) P(H = C) = 0 \times 0,02 = 0$

e. Perbandingan hasil *class*

Dari hasil perhitungan akhir dari setiap class, maka nilai probabilitas tertinggi berada pada class  $P(H = \text{Sangat Baik})$  yaitu 0,4061516 dibandingkan class  $P(H = \text{Baik})$  yaitu 0,010218 dan class  $P(H = \text{Cukup})$  yaitu 0. Maka dapat disimpulkan bahwa klasifikasi hasil belajar pada nilai rapor tahun ketiga (T3) dari data testing dikategorikan “Sangat Baik”.

**Implementasi Pengujian Tingkat Keakurasian *Naïve Bayes* Dengan RapidMiner**

Dalam pengujian tingkat keakurasian dilakukan dengan bantuan *software* RapidMiner menggunakan ketentuan ratio 70% data *training* dan ratio 30% data *testing* pada operator *split data*. Dimana terlebih dahulu *dataset* yang akan digunakan di import dan disimpan dalam format .csv.



**Gambar 3. Proses Pengujian tingkat keakurasian**

Pada gambar 1 dapat lihat bahwa data diolah dengan algoritma klasifikasi *Naïve Bayes*. Lalu operator split data digunakan dalam membagi *dataset* menjadi dua data yaitu data *training* dan data *testing*. Selain itu juga, digunakannya operator *apply model* untuk menghubungkan model yang telah dilatih pada *dataset* baru yang ingin diprediksi. Operator *performance* juga digunakan untuk menghitung metrik akurasi dan memberikan daftar nilai secara otomatis dari data *training* terhadap data *testing* yang diprediksi. Dari operator yang digunakan akan berpengaruh terhadap hasil pengujian.

accuracy: 94.44%

	true B	true SB	true C	class precision
pred. B	10	0	0	100.00%
pred. SB	1	6	0	85.71%
pred. C	0	0	1	100.00%
class recall	90.91%	100.00%	100.00%	

**Gambar 4. Performance Vector Naïve Bayes**

Dari gambar 2 dapat dilihat tingkat akurasi yang diperoleh sebesar 94,44% dari data *training* sebanyak 42 data dan data *testing* sebanyak 18 data. Dimana sebanyak 10 data yang dinyatakan nilai B diprediksikan benar, 1 data dinyatakan nilai B diprediksi salah dan diprediksikan bernilai SB, sebanyak 6 data yang dinyatakan nilai SB diprediksikan benar dan 1 data yang dinyatakan nilai C diprediksi benar.

Dari nilai akurasi yang diperoleh mencapai 94,44% dimana nilai akurasi tersebut tinggi yang dapat diartikan bahwa sebagian besar data *testing* yang digunakan berhasil diklasifikasikan dengan sangat baik oleh *Naïve Bayes*.

**Tabel 1. Data Training**

NO	JK	TL	PTA	PKA	PHA	PTI	PKI	PHI	T1	T2	T3	Kategori
1.	L	2006	SMA	Wirusaha	Rp1.000.000 - Rp2.000.000	SMA	Ibu Rumah Tangga	Rp0 - Rp500.000	79.69	82.68	85.42	B
2.	P	2006	SMA	Wiraswasta	Rp1.000.000 - Rp2.000.000	SD	Wiraswasta	Rp2.000.000 - Rp3.000.000	87.08	89.45	93.05	SB
3.	P	2006	SMP	Wiraswasta	Rp1.000.000 - Rp2.000.000	SMA	Ibu Rumah Tangga	Tidak Berpenghasilan	85.51	87.92	91.79	SB
4.	P	2006	SMP	Wiraswasta	Rp2.000.000 - Rp3.000.000	SMA	Ibu Rumah Tangga	Tidak Berpenghasilan	87.37	89.74	92.95	SB
5.	P	2007	SMP	Supir	Rp1.000.000 - Rp2.000.000	SMP	Ibu Rumah Tangga	Tidak Berpenghasilan	81.55	84.26	87.26	B
6.	P	2006	SMA	Wiraswasta	Rp2.000.000 - Rp3.000.000	SMA	Wiraswasta	Rp500.000 - Rp1.000.000	81.18	84.26	87.16	B
7.	P	2006	SMA	Wirusaha	Rp500.000 - Rp1.000.000	SMA	Wiraswasta	Rp500.000 - Rp1.000.000	79.21	80.61	84.58	B
8.	P	2006	SMP	Wiraswasta	Rp2.000.000 - Rp3.000.000	SMA	Ibu Rumah Tangga	Tidak Berpenghasilan	90.19	92.05	93.79	SB
9.	P	2006	SMA	Tidak Bekerja	Tidak Berpenghasilan	Tamat Perguruan Tinggi	Guru	Rp500.000 - Rp1.000.000	88.56	88.95	91.16	SB
10.	P	2006	SMA	Wiraswasta	Rp1.000.000 - Rp2.000.000	SMP	Ibu Rumah Tangga	Tidak Berpenghasilan	85.61	87.58	89.16	B
11.	P	2006	SMP	Wiraswasta	>Rp3.000.000	SMP	Ibu Rumah Tangga	Tidak Berpenghasilan	88.17	89.76	92.11	SB
12.	L	2006	SMA	Wirusaha	Rp500.000 - Rp1.000.000	Tamat Perguruan Tinggi	PNS	Rp1.000.000 - Rp2.000.000	79.74	82.62	85.95	B
13.	P	2006	SMA	Pensiunan	Rp2.000.000 - Rp3.000.000	SMA	Ibu Rumah Tangga	Tidak Berpenghasilan	83.40	87.43	90.16	SB
14.	P	2006	SMA	Wiraswasta	Rp500.000 - Rp1.000.000	SMP	Ibu Rumah Tangga	Tidak Berpenghasilan	82.42	85.25	87.26	B
15.	P	2006	SMA	Petani	>Rp3.000.000	Tamat Perguruan Tinggi	PNS	>Rp3.000.000	84.81	85.25	86.84	B
60.	L	2007	SMP	Wiraswasta	Rp500.000 - Rp1.000.000	SMP	Ibu Rumah Tangga	Rp0 - Rp500.000	84.76	83.99	88.09	B

**Tabel 2. Data Testing**

NO	JK	TL	PTA	PKA	PHA	PTI	PKI	PHI	T1	T2	T3
1.	P	2006	SMA	Wiraswasta	Rp1.000.000 - Rp2.000.000	SMA	Ibu Rumah Tangga	Tidak Berpenghasilan	80	85	?

## PEMBAHASAN

### Pengolahan Data Dengan *Naïve Bayes*

Klasifikasi adalah sebuah model *data mining* yang digunakan dalam menilai objek data untuk dimasukkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia (Mustika et al., 2021). Sedangkan algoritma *Naïve Bayes* merupakan salah satu klasifikasi *data mining* yang digunakan dalam memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema *Bayes* (Rifai et al., 2019).

Berdasarkan pada pengolahan dan analisis data yang dilakukan dengan persamaan teorema *bayes* dengan 60 data *training* dan 1 data *testing* diperoleh hasil bahwa nilai probabilitas tertinggi berada pada *class* P(H = Sangat Baik) yaitu 0,4061516 dibandingkan *class* P(H = Baik) yaitu 0,010218 dan *class* P(H = Cukup) yaitu 0. Sehingga disimpulkan bahwa klasifikasi hasil belajar pada nilai rapor tahun ketiga (T3) dari data *testing* dikategorikan “Sangat Baik”. Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh (Firdaus, 2019) yang berjudul “Penerapan Metode *Naïve Bayes Classifier* Untuk Mengklasifikasi Tingkat Prestasi Akademik Santri Pondok Pesantren Mahasiswa (PPM) Baitul Jannah Malang” memperoleh hasil dimana dari perhitungan manual yang dilakukan menunjukkan nilai probabilitas tertinggi berada pada P(A = Sangat Baik) yaitu 0,01252 sehingga klasifikasi tingkat prestasi akademik dari data *testing* dikategorikan “Sangat Baik”. Hasil penelitian ini juga sejalan dengan penelitian (Gunawan et al., 2019) yang berjudul “Klasifikasi Hasil Belajar Peserta Didik Pada Jurusan Teknik Komputer dan Jaringan (TKJ) di SMK Negeri 3 Malang Menggunakan Algoritma *Naïve Bayes*” memperoleh hasil dimana dari perhitungan manual yang dilakukan menunjukkan nilai probabilitas tertinggi berada pada P(A = Sangat Baik) dengan nilai  $9,12 \times 10^{-6}$  sehingga klasifikasi pada rata-rata nilai tahun ketiga dari data *testing* dikategorikan “Sangat Baik”.

### Implementasi Pengujian Tingkat Keakurasian *Naïve Bayes* Dengan RapidMiner

Berdasarkan pada implementasi pengujian tingkat keakurasian metode *Naïve Bayes* dalam mengklasifikasi hasil belajar menggunakan RapidMiner dimana sebanyak 60 data dibagi menjadi dua data menggunakan rasio yang pertama 70% sebanyak 42 data *training* dan rasio kedua 30% sebanyak 18 data *testing* pada operator *split data*. Klasifikasi hasil belajar yang dilakukan memperoleh hasil sebanyak 10 siswa/i diprediksi hasil belajar bernilai B(Baik), sebanyak 6 siswa/i diprediksi hasil belajar bernilai SB(Sangat Baik) dan 1 siswa/i diprediksi hasil belajar bernilai C(Cukup). Berikut perhitungan untuk mendapatkan nilai akurasi dengan persamaan berikut.

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi benar}}{\text{jumlah data}} \times 100\%$$
$$\text{Akurasi} = \frac{17}{18} \times 100\% = 94,44\%$$

Nilai 94,44% membuktikan bahwa metode *Naïve Bayes* cukup baik digunakan untuk melakukan klasifikasi hasil belajar siswa/i. Namun nilai 94,44% perlu ditinjau kembali dari jumlah *dataset* dan kompleksitas data. Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Rahman, (2023) yang berjudul “Klasifikasi Performa Akademik Siswa Menggunakan Metode *Decision Tree* dan *Naïve Bayes*” dimana data yang digunakan sebanyak 50 data yang kemudian dibagi menjadi data *training* dan data *testing* dengan

menggunakan operator *split data* pada RapidMiner, dengan ketentuan data *training* 70% atau sebanyak 35 data dan data *testing* 30% atau sebanyak 15 data. Berdasarkan pada metode *Decision Tree* menunjukkan hasil akurasi data performa akademik siswa sebesar 83.89%. Sedangkan pada metode *Naïve Bayes* menunjukkan hasil akurasi sebesar 85,97%. Dari kedua metode tersebut, data akurasi *Naïve Bayes* mendapatkan nilai yang lebih baik dibandingkan dengan *Decision Tree*. Hal ini berarti metode *Naïve Bayes* maksimal digunakan dalam klasifikasi Performa akademik siswa.

## SIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Klasifikasi hasil belajar dengan metode *Naïve Bayes* memperoleh hasil bahwa klasifikasi hasil belajar pada nilai rapor tahun ketiga (T3) dari data *testing* dikategorikan “Sangat Baik” dengan nilai probabilitas *class*  $P(H = \text{Sangat Baik})$  yaitu 0,4061516
2. Hasil pengujian tingkat keakurasian metode *Naïve Bayes* dalam mengklasifikasi hasil belajar memperoleh hasil akurasi sebesar 94,44% dimana nilai akurasi tersebut tinggi yang dapat diartikan bahwa sebagian besar data *testing* yang digunakan berhasil diklasifikasikan dengan sangat baik oleh *Naïve Bayes* dan dapat membuktikan bahwa metode *Naïve Bayes* cukup baik digunakan untuk melakukan klasifikasi hasil belajar siswa/i.

## DAFTAR PUSTAKA

- Ahdar. (2021). *Ilmu Pendidikan* (Musyarif (ed.)). IAIN Parepare Nusantara Press.
- Basri, H., Azis, M. S., Malau, Y., Fridayanthie, E. W., Rizal, K., & Rianto, H. (2022). Penerapan Particle Swarm Optimization Pada Algoritma Naïve Bayes Untuk Klasifikasi Hasil Belajar. *Information System for Educators and Professionals : Journal of Information System*, 6(2), 97. <https://doi.org/10.51211/isbi.v6i2.1752>
- Firdaus, Y. M. (2019). Penerapan Metode *Naïve Bayes Classifier* Untuk Mengklasifikasi Tingkat Prestasi Akademik Santri Pondok Pesantren Mahasiswa ( PPM ) Baitul Jannah Malang. *Jurnal Mahasiswa Teknik Informatika*, 3(1), 327–336. <https://doi.org/10.36040/jati.v3i1.1398>
- Firmansyah, & Yulianto, A. (2023). Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(2), 1174–1182. <https://doi.org/10.33395/remik.v7i2.12375>
- Gunawan, R., Wijoyo, S. H., & Wicaksono, S. A. (2019). Klasifikasi Hasil Belajar Peserta Didik Pada Jurusan Teknik Komputer dan Jaringan (TKJ) di SMK Negeri 3 Malang Menggunakan Algoritma Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(8), 7475–7482.
- Hidayat, R., & Abdillah. (2019). *Ilmu Pendidikan : Konsep, Teori dan Aplikasinya* (C. Wijaya & Amiruddin (eds.)). LPPPI.
- Marpaung, S., Solikhun, & Irawan. (2021). Penerapan Metode Naïve Bayes Dalam Memprediksi Prestasi Siswa Di SMA Negeri 1 Panombeian Panei. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 4(2), 8–13. <https://doi.org/10.34012/journalsisteminformasidanilmukomputer.v4i2.1522>

- Muslim, M. A., Prasetyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program komputer* (E. Listiana & N. Cahyani (eds.)). Unnes Repository.
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, Manuhutu, M. A., Ridwan, M., Hozairi, Wardhani, A. K., Alim, S., Romli, I., Religia, Y., Octafian, D. T., Sufandi, U. U., & Ernawati, I. (2021). *Data Mining dan Aplikasinya*. Widina Bhakti Persada Bandung.
- Rahman, A. (2023). Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes. *Jurnal SAINTEKOM*, 13(1), 22–31. <https://doi.org/10.33020/saintekom.v13i1.349>
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *PETIR : Jurnal Pengkajian Dan Penerapan Teknik Informatika*, 12(2), 131–144. <https://doi.org/10.33322/petir.v12i2.471>
- Triwidianti, J., Alfian, F. Y., & Prasojo, M. (2021). Perbandingan Metode *Data Mining* Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung. *Seminar Nasional Hasil Penelitian Dan Pengabdian Masyarakat*, 1, 126–133.
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). *Data Mining : Algoritma & Implementasi* (T. Limbong (ed.); Issue 1). Yayasan Kita Menulis.