

# Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar

<sup>1</sup>\*Firmansyah, <sup>2</sup>Agus Yulianto

<sup>1</sup>Universitas Bina Sarana Informatika, <sup>2</sup>Universitas Nusa Mandiri  
Jakarta, Indonesia

<sup>1</sup>[firmansyah.fmh@bsi.ac.id](mailto:firmansyah.fmh@bsi.ac.id), <sup>2</sup>[agus.aag@nusamandiri.ac.id](mailto:agus.aag@nusamandiri.ac.id)

\*Penulis Korespondensi

Diajukan : 19/04/2023

Diterima : 29/04/2023

Dipublikasi : 29/04/2023

## ABSTRAK

Sekolah dasar merupakan lembaga pendidikan yang menyelenggarakan program pendidikan untuk menyiapkan peserta didik ke jenjang selanjutnya. Untuk dapat melanjutkan ke jenjang berikutnya tentunya dibutuhkan hasil belajar yang baik sehingga siswa dapat lulus tepat waktu. Hasil belajar adalah indikator prestasi dari peserta didik sehingga dibutuhkan model algoritma yang dapat memprediksi hasil belajar, tujuannya adalah sebagai alat pendukung keputusan lembaga dalam mengevaluasi proses pembelajaran, Hasil belajar seorang siswa merupakan patokan kemampuan siswa menguasai pelajaran sekolah, Akan lebih baik hasil belajar selalu dievaluasi oleh pendidik dengan tujuan untuk melihat kemampuan tiap-tiap anak, dengan menggunakan sampling 2 fitur yaitu kehadiran dan peringkat dan 1 fitur kelas yaitu klasifikasi. Fitur kehadiran merupakan jumlah kehadiran siswa dimana dibagi menjadi 4 klasifikasi yaitu  $\geq 90$ ,  $\geq 70$ ,  $\geq 50$  dan  $< 50$ . Fitur peringkat untuk menggambarkan peringkat siswa di kelas dengan klasifikasi peringkat 1-3, 4-10 dan  $> 10$ , sedangkan kelas nilai adalah kelas yang sudah memiliki label  $\geq 90$  dan  $< 90$ . *Naive Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Dengan *Naive Bayes Classifier* dapat memprediksi belajar dari peserta didik sekolah dasar berdasarkan variabel kehadiran dan peringkat. Hasil dari pengujian menggunakan 10 *Fold-Cross Validation* pada *Rapid Miner* menunjukkan nilai akurasi mencapai 89,93%, nilai presisi mencapai 68,95 % dan nilai mencapai 65%.

**Kata Kunci:** *naïve bayes classifier*, klasifikasi *naïve bayes*, prediksi kelulusan siswa, prediksi

## I. PENDAHULUAN

Pendidikan merupakan bimbingan atau pertolongan yang diberikan oleh orang dewasa kepada perkembangan anak untuk mencapai kedewasaannya dengan tujuan agar anak cukup cakap melaksanakan tugas hidupnya sendiri tanpa bantuan orang lain (Kosilah & Septian, 2020). Jenjang pendidikan dimulai dari sekolah dasar, sekolah menengah pertama, sekolah menengah atas hingga perguruan tinggi. Jenjang dasar yang dilalui dalam proses pembelajaran adalah sekolah dasar dimana siswa harus menempuh 6 tahun belajar untuk mempelajari *sains* dasar, kemampuan *kognitif*, *afektif* dan *psikomotorik* (Melianti et al., 2023).

Hasil belajar merupakan tolak ukur kemampuan siswa dalam menguasai satu atau lebih mata pelajaran. Akan lebih baik hasil belajar selalu dievaluasi oleh pendidik dengan tujuan untuk melihat kemampuan tiap-tiap anak, namun evaluasi hanyalah kegiatan yang hanya bisa dilakukan

pada akhir semester. Dibutuhkan suatu alat yang dapat memprediksi hasil belajar siswa berdasarkan parameter yang sudah ditentukan. Prediksi prestasi belajar peserta didik dapat dijadikan sebagai dasar untuk melakukan intervensi dini terhadap potensi kegagalan peserta didik untuk mencapai tujuan pembelajaran; dan di saat yang sama dapat melakukan perubahan strategi pembelajaran agar dapat memfasilitasi keragaman peserta didik (Rajagukguk, 2021).

## II. STUDI LITERATUR

### 1. Penelitian Terkait

Prediksi hasil belajar peserta didik menggunakan *model multiple linier regression* menggunakan 2 variabel independen yaitu kehadiran dan peringkat dan 1 variabel dependen yaitu hasil belajar (Firmansyah & Yulianto, 2022). Penelitian juga dilakukan untuk memprediksi nilai siswa menggunakan metode *radial basis function* dan menggunakan aplikasi *web* sebagai antarmukanya (Afriani et al., 2019). Penelitian juga dilakukan untuk membandingkan algoritma untuk memprediksi nilai dan waktu kelulusan, algoritma yang dibandingkan yaitu algoritma C4.5, *Naïve Bayes*, KNN dan *Support Vector Machine*. Hasil menunjukkan di antara 4 algoritma tersebut *naïve bayes* mendapatkan nilai akurasi tertinggi (Widaningsih, 2019). *Naïve bayes* digunakan untuk memprediksi ketercapaian nilai kriteria ketuntasan minimal (KKM) siswa (Dwi Hartanti et al., 2018).

### 2. Penilaian Hasil Belajar

Hasil belajar adalah perubahan tingkah laku sebagai hasil belajar dalam definisi yang lebih luas mencakup bidang *kognitif*, *efektif*, dan *psikomotorik* (Nurdin Mansur, 2015). Secara sederhana, hasil belajar adalah kemampuan yang diperoleh peserta didik setelah melalui kegiatan belajar. Secara praktis, hasil belajar juga dimaksudkan untuk mengungkapkan kemampuan siswa dalam bentuk angka-angka (Achdiyat & Utomo, 2017) dimana angka tersebut adalah hasil penilaian terhadap kemampuan peserta didik setelah menjalani proses pembelajaran. Dan dimaksudkan untuk mengetahui kemampuan peserta didik (Azizah & Widjajanti, 2019).

### 3. Naïve Bayes

*Naive Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan *teorema Bayes* dan mengasumsikan semua *atribut* independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil & S. S. Shrekar, 2013). Persamaan dari *teorema bayes* adalah :

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

Dimana :

$X$  : Data dengan *class* yang belum diketahui

$H$  : Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$  : Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$  : Probabilitas X

## III. METODE

### A. CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) digunakan sebagai kerangka kerja dalam menerapkan proses *data mining* (Larose, 2006) adapun prosesnya yaitu :

#### 1. *Business Understanding Phase*

Memahami visi dan misi dari organisasi, bisnis dan proses bisnis dari organisasi serta mendefinisikan masalah dan menetapkan tujuan dari pemodelan yang akan dibuat. Tahap ini dilakukan untuk mendapatkan pemahaman tujuan berdasarkan perspektif bisnis untuk kemudian diubah menjadi definisi masalah data mining. Setelah itu ditentukan bagaimana solusi yang tepat dalam menangani permasalahan tersebut (Wiratama & Pradnya, 2022).

## **2. Data Understanding Phase**

Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, serta mengevaluasi kualitas data (Hasanah et al., 2021). Memahami dari data yang akan digunakan sebagai data training dan data testing yang akan digunakan untuk pemodelan. Fitur-fitur yang ada di dalam data juga harus dapat dimengerti diambil dari mana dan menggambarkan apa.

## **3. Data Preparation Phase**

Menyiapkan data dan mengolah data dari data sumber. Data sumber tentunya dapat dari berbagai macam seperti *spreadsheet*, *csv* bahkan *database*. Data yang diolah harus menyesuaikan dengan model algoritma yang digunakan, fitur-fiturnya harus disesuaikan juga dengan kebutuhan.

## **4. Modelling Phase**

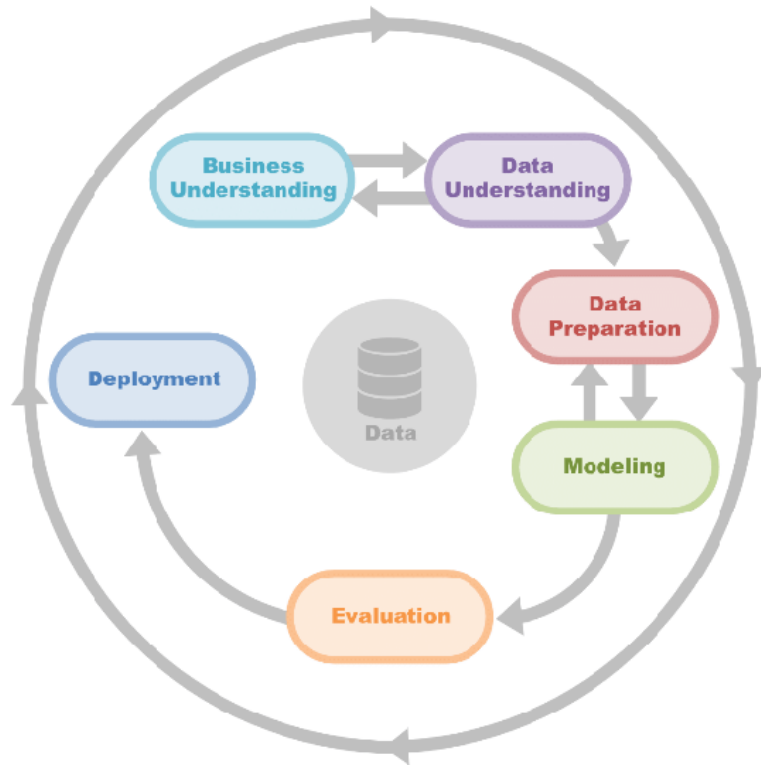
Menyiapkan model atau algoritma yang akan digunakan, bergantung dari kebutuhan seperti *klasifikasi*, *prediksi*, *clustering* atau *asosiasi*.

## **5. Evaluation Phase**

Data yang sudah dibuat pemodelan dengan *algoritma* tertentu, kemudian dilakukan evaluasi untuk menghitung seberapa akurat dan presisi dari sebuah *algoritma*. Nilai ini nantinya dapat dijadikan acuan *algoritma* yang akan digunakan.

## **6. Deployment Phase**

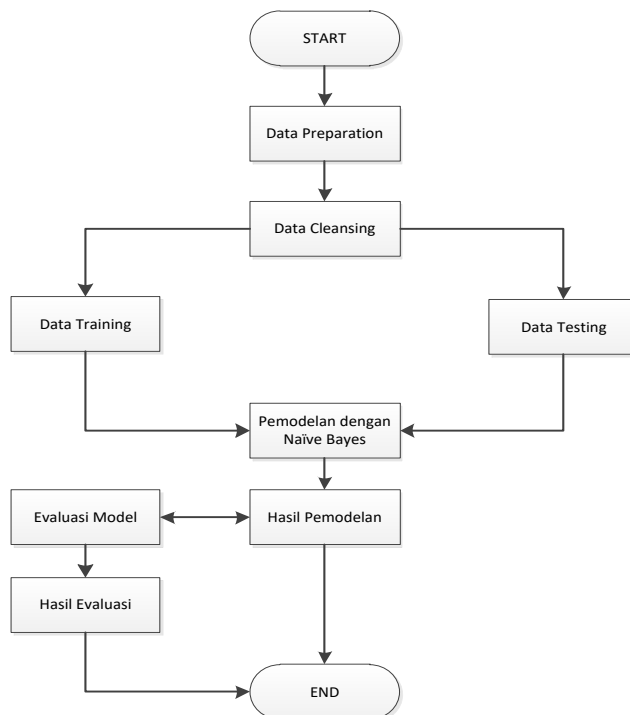
Setelah dibuatkan pemodelan dan sudah menentukan *model algoritma* yang akan diterapkan, tahap berikutnya yaitu menerapkan model menggunakan berbagai macam alat seperti *rapidminer*, *weka*, *python* dan *tools* lainnya.



Gambar 1 CRISP-DM Process

**B. Flowchart Model Naïve Bayes**

Flowchart digunakan untuk menggambarkan alur proses mulai dari data preparation hingga pemodelan. Flowchart untuk menggambarkan process pemodelan menggunakan metode regresi linier berganda ini seperti di bawah ini :



Gambar 2 Proses Pemodelan

## IV. HASIL DAN PEMBAHASAN

### 1. *Business Understanding Phase*

Sekolah dasar merupakan lembaga pendidikan yang menyelenggarakan program pendidikan untuk menyiapkan peserta didik ke jenjang selanjutnya. Untuk dapat melanjutkan ke jenjang berikutnya tentunya dibutuhkan hasil belajar yang baik sehingga siswa dapat lulus tepat waktu. Hasil belajar adalah indikator prestasi dari peserta didik sehingga dibutuhkan model *algoritma* yang dapat memprediksi hasil belajar. Tujuannya adalah sebagai alat pendukung keputusan lembaga dalam mengevaluasi proses pembelajaran.

### 2. *Data Understanding Phase*

Sumber data yang diambil merupakan data primer yaitu hasil belajar peserta didik yang diambil secara *random sampling* dari sekolah dasar di wilayah Jakarta Barat untuk periode akademik tahun 2019-2020. Terdapat 3 fitur pada data ini yaitu kehadiran, peringkat dan 1 kelas yaitu nilai rapor.

### 3. *Data Preparation Phase*

Jumlah data yang diambil berjumlah 300 baris dengan 2 fitur yaitu kehadiran dan peringkat dan 1 fitur kelas yaitu klasifikasi. Fitur kehadiran merupakan jumlah kehadiran siswa dimana dibagi menjadi 4 klasifikasi yaitu  $\geq 90$ ,  $\geq 70$ ,  $\geq 50$  dan  $< 50$ . Fitur peringkat untuk menggambarkan peringkat siswa di kelas dengan klasifikasi peringkat 1-3, 4-10 dan  $> 10$ , sedangkan kelas nilai adalah kelas yang sudah memiliki label  $\geq 90$  dan  $< 90$ . Detailnya dibawah ini :

Tabel 1 Fitur dan Kelas

Nama fitur/variabel	Tipe data	Klasifikasi
Kehadiran	Numerik (persentase)	$\geq 90$ $\geq 70$ $\geq 50$ $< 50$
Peringkat	Numerik (persentase)	1-3 4-10 $> 10$
Nilai	Numerik	$\geq 90$ $< 90$

Dari total data sebanyak 300, dilakukan *proses cleaning* seperti *missing values*, redundansi dan data inkonsisten terhadap data sehingga data training berjumlah 248 baris dengan 3 fitur yaitu kehadiran, peringkat dan klasifikasi nilai. Hasil yang akan diprediksi dari data adalah klasifikasi nilai siswa masuk ke kategori nilai lebih besar atau sama dengan 90 atau lebih kecil dari 90. Tujuannya untuk melihat pola siswa yang memiliki nilai sangat baik agar menjadi evaluasi.

### 4. *Modelling Phase*

Data training dibentuk menjadi spreadsheet dimana ada 2 variabel ( $X_1$  dan  $X_2$ ) dan class label ( $Y$ ) yang dimulai dari periode tahun 2019 sampai dengan 2020. Kemudian dilakukan simulasi menggunakan data uji sebagai berikut :

Tabel 2 Data Uji

Nama	Kehadiran	Peringkat
Tono	$\geq 90$	4-10

Dari data uji di atas, kemudian akan diproses menggunakan *algoritma naïve bayes classifier* berdasarkan data latih yaitu :

Probabilitas kelas Nilai <90 :  
 $P(<90) = 213/248 = 0.86$   
 Probabilitas kelas Nilai >=90 :  
 $P(>=90) = 35/248 = 0.14$

Menghitung probabilitas kelas Nilai <90:

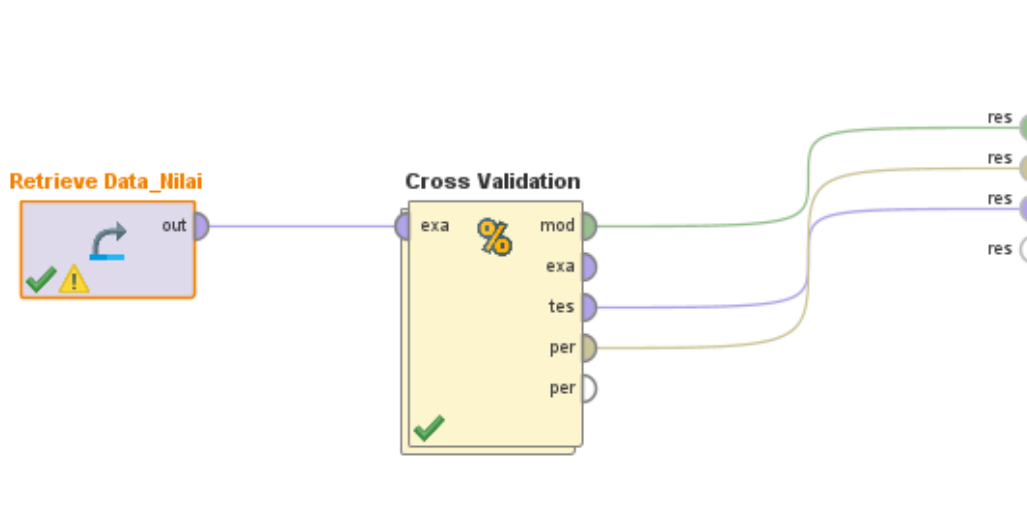
$P(Class.<90 | Kehadiran.>=90) = 208/213 = 0.98$   
 $P(Class.<90 | Peringkat 4-10) = 11/213 = 0.35$   
 Probabilitas termasuk kategori Nilai <90 :  
 $Class.<90 = 0.86 \times 0.98 \times 0.35 = 0.29$

Menghitung probabilitas kelas Nilai >=90:

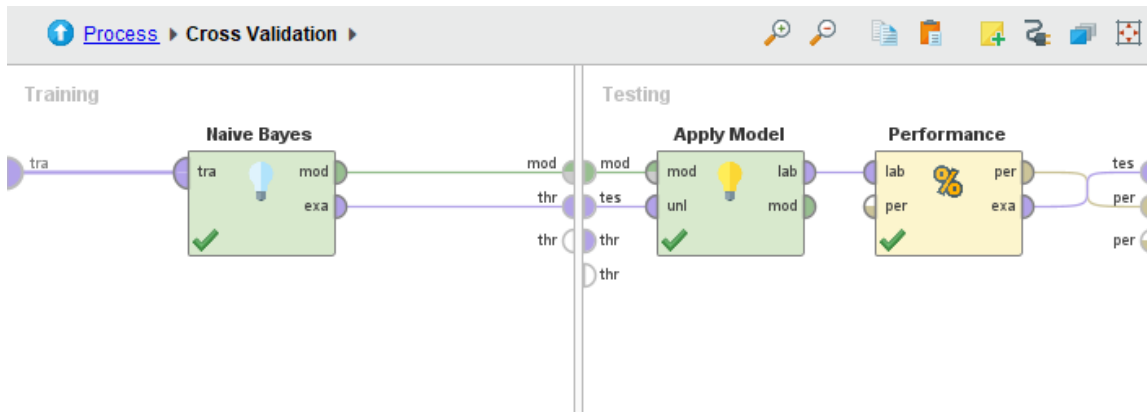
$P(Class.>=90 | Kehadiran.>=90) = 35/35 = 1$   
 $P(Class.>=90 | Peringkat 4-10) = 85/35 = 0.51$   
 Probabilitas termasuk kategori Nilai >=90 :  
 $Class.>=90 = 0.14 \times 1 \times 0.51 = 0.0714$

Dari kesimpulan hasil algoritma di atas maka disimpulkan bahwa Tono diprediksi masuk kategori Nilai <90 karena nilai probabilitasnya paling tinggi dibanding probabilitas kelas Nilai >=90.

Pemodelan menggunakan *Rapid Miner* seperti ini :



Gambar 3 Pemodelan *Rapid Miner*



Gambar 4 Cross Validation di Rapid Miner

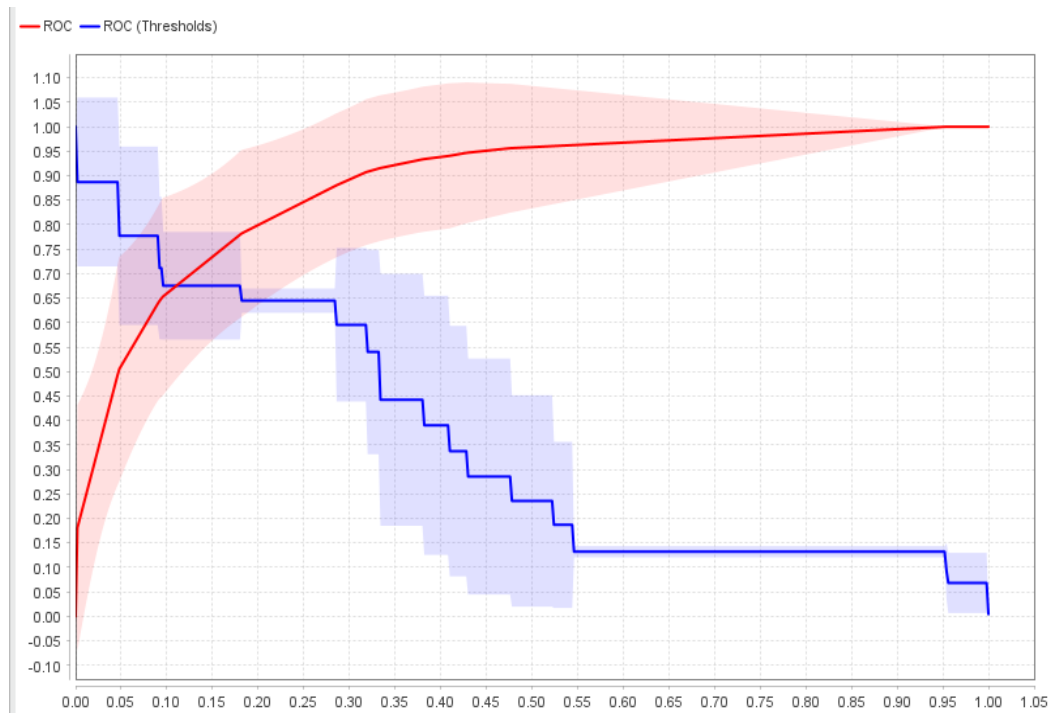
5. Evaluation Phase

Pengujian model menggunakan Confusion Matrix dan grafik AUC seperti di bawah ini :

accuracy: 89.93% +/- 4.32% (micro average: 89.92%)

	true <90	true >=90	class precision
pred. <90	200	12	94.34%
pred. >=90	13	23	63.89%
class recall	93.90%	65.71%	

Gambar 5 Confusion Matrix



Gambar 6 Grafik AUC



## 6. Deployment Phase

Implementasi menggunakan *Rapid Miner* versi 10.1 dibuat pemodelan menggunakan *naïve bayes* dengan hasilnya seperti ini :

Row No.	NOMOR	CLASS NILAI	prediction(C...	confidence(...	confidence(...	KEHADIRAN	PERINGKAT
1	4	>=90	>=90	0.376	0.624	>=90	1-3
2	33	<90	<90	0.874	0.126	>=90	4-10
3	37	<90	<90	0.991	0.009	>=90	>10
4	57	>=90	>=90	0.376	0.624	>=90	1-3
5	59	<90	<90	0.991	0.009	>=90	>10
6	71	<90	<90	0.874	0.126	>=90	4-10
7	86	<90	<90	0.991	0.009	>=90	>10
8	90	<90	<90	0.874	0.126	>=90	4-10
9	92	<90	<90	0.991	0.009	>=90	>10
10	103	<90	<90	0.991	0.009	>=90	>10
11	104	<90	<90	0.874	0.126	>=90	4-10
12	107	<90	<90	0.991	0.009	>=90	>10
13	123	<90	<90	0.874	0.126	>=90	4-10
14	131	<90	<90	0.991	0.009	>=90	>10
15	135	<90	<90	0.874	0.126	>=90	4-10

ExampleSet (248 examples, 5 special attributes, 2 regular attributes)

Gambar 7 Deployment dengan Rapidminer

## V. KESIMPULAN

*Naïve bayes classifier* terbukti dapat memprediksi hasil belajar dari peserta didik sekolah dasar berdasarkan variabel kehadiran dan peringkat. Hasil dari pengujian menggunakan 10 *Fold-Cross Validation* pada *Rapid Miner* menunjukkan nilai akurasi mencapai 89,93%, nilai presisi mencapai 68,95 % dan nilai mencapai 65%. Dengan melihat hasil pengujian, dapat disimpulkan bahwa *algoritma naïve bayes classifier* cukup baik dalam memprediksi hasil belajar siswa dan dapat digunakan untuk institusi menentukan kebijakan dan keputusan dalam rangka memperbaiki hasil belajar.

## VI. REFERENSI

- Achdiyat, M., & Utomo, R. (2017). Kecerdasan Visual-Spasial, Kemampuan Numerik dan Prestasi Belajar Matematika. *Jurnal Ilmiah Pendidikan MIPA*, 1(1), 234–245.
- Afriani, R., Hotlan Sitorus, S., Ristian, U., Rekeyasa Sistem Komputer, J., & Hadari Nawawi, J. H. (2019). Aplikasi Prediksi Nilai Siswa Sekolah Dasar Menggunakan Metode Radial Basis Function Berbasis Web (Studi Kasus : SDN 19 Sungai Raya). *Coding: Jurnal Komputer Dan Aplikasi*, 07(03), 85–96.
- Azizah, I. N., & Widjajanti, D. B. (2019). Keefektifan Pembelajaran Berbasis Proyek Ditinjau Dari Prestasi Belajar, Kemampuan Berpikir Kritis, dan Kepercayaan Diri Siswa. *Jurnal Riset Pendidikan Matematika*, 6(2), 233–243. <https://doi.org/10.21831/jrpm.v6i2.15927>
- Dwi Hartanti, Kusriani, & Taufiq, E. L. (2018). Penerapan Naïve Bayes Dalam Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa. *Jusikom Prima*, 2(1), 15–22.
- Firmansyah, & Yulianto, A. (2022). Prediksi Hasil Belajar Peserta Didik Menggunakan Model



---

Multiple Linier Regression. *REMIK : Riset Dan E-Jurnal Manajemen Informatika Dan Komputer*, 6(4), 654–663. <https://doi.org/10.33395/remik.v6i4.11763>

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>

Kosilah, & Septian. (2020). Penerapan Model Pembelajaran Kooperatif Tipe Assure Dalam Meningkatkan Hasil Belajar Siswa. *Jurnal Inovasi Pendidikan*, 1(6), 1139–1148.

Larose, D. T. (2006). *Data Mining Methods and Models*. Johns Wiley & Sons.

Melianti, E., Handayani, D., Novianti, F., Syahputri, S., & Hasibuan, S. A. (2023). Pentingnya Pendidikan Yang Ada di Sekolah Dasar. *Jurnal Pendidikan Dan Konseling*, 5(1), 3549–3554.

Nurdin Mansur. (2015). Pencapaian Hasil Belajar Ditinjau Dari Sikap Belajar Mahasiswa. *Lantanida Journal*, 3(2), 107–115.

Patil, T. R., & S. S. Sherekar. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications*, 6(2), 256–261. <https://doi.org/10.18201/ijisae.2019252786>

Rajagukguk, S. A. (2021). Tinjauan Pustaka Sistematis: Prediksi Prestasi Belajar Peserta Didik Dengan Algoritma Pembelajaran Mesin. *Jurnal SNATi*, 1, 22–30.

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, KNN Dan SVM. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>

Wiratama, M. A., & Pradnya, W. M. (2022). Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(1). <https://doi.org/10.23887/janapati.v11i1.45282>