

Prediksi Kelulusan Siswa SDN 016528 BP. Mandoge dengan Metode *Naïve Bayes*

Cetryn Ayu Diah Lestari¹, Juwita Sari², Sri Wulandari³

^{1,2,3} Mahasiswa Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal

¹ saragihcetrynuyudiahlestari@gmail.com*

* Email Koresponden

INFO ARTIKEL

Histori Artikel

Diterima: 26/Juli/2024

Ditinjau: 28/Juli/2024

Disetujui: 31/Juli/2024

ABSTRAK

Kelulusan menandai akhir dari suatu jenjang pendidikan di sekolah. Penelitian ini bertujuan untuk memprediksi kelulusan siswa di SDN 016528 BP Mandoge berdasarkan kemampuan mereka. Tujuan dari penelitian ini adalah mengurangi tingkat kegagalan siswa dalam mencapai kelulusan, dengan melakukan prediksi berdasarkan data nilai ujian yang dikumpulkan oleh lembaga. Metode yang digunakan dalam penelitian ini adalah Naive Bayes, sebuah teknik dalam Data Mining yang memanfaatkan probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan data sebelumnya. Teknik ini dipilih karena keunggulannya dalam memprediksi jumlah kelulusan berdasarkan data konkret, sehingga hasilnya dapat dipercaya dan digunakan untuk prediksi selanjutnya. Dataset penelitian ini mencakup data kelulusan siswa di SDN 016528 BP Mandoge tahun ajaran 2019/2020, dengan total 171 siswa, terdiri dari 120 siswa untuk data pelatihan dan 51 siswa untuk data pengujian, menghasilkan akurasi pemodelan sebesar 98%.

Kata Kunci: Kelulusan, Data Mining, Naïve Bayes



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright ©2024 by Author. Published by PT Beranda Teknologi Academia

ABSTRACT

Graduation marks the completion of a certain level of schooling. This study aims to predict the graduation of students at SDN 016528 BP Mandoge based on their abilities. The goal of this research is to reduce the rate of student failure to graduate by making predictions based on examination scores collected by the institution. The method used in this study is Naive Bayes, a technique in Data Mining that utilizes probability and statistics to predict future outcomes based on previous data. This method was chosen due to its advantage in predicting graduation rates from concrete data, ensuring the results are reliable and applicable for future predictions. The dataset used in this study includes graduation data for SDN 016528 BP Mandoge students for the 2019/2020 academic year, comprising 171 students, with 120 students used for training data and 51 students for testing data, achieving a model accuracy of 98%.

Keywords: Graduation, Data Mining, Naïve Bayes

PENDAHULUAN

Pendidikan adalah sebuah sistem yang terdiri dari berbagai komponen yang saling terhubung dan memiliki tujuan bersama, yaitu mencetak sumber daya yang berkualitas. Pendidikan memiliki andil besar dalam meningkatkan kualitas SDM, yang secara signifikan mempengaruhi masa depan bangsa [1].

Sekolah Dasar (SD) Negeri 016528 Bandar Pasir Mandoge merupakan sekolah SDN yang lokasinya berada dipedesaan yang jumlah penduduknya sedikit dan membuat kurangnya minat siswa

untuk masuk ke sekolah tersebut. Untuk menjadikan sekolah tersebut menjadi sekolah yang berkualitas, sekolah tersebut harus memiliki tingkat kelulusan siswa yang tinggi.

Saat ini, Sekolah Dasar (SD) Negeri 016528 Bandar Pasir Mandoge belum memiliki sistem untuk memprediksi kelulusan siswa. Oleh sebab itu, dibuatlah sebuah sistem yang dapat membantu sekolah dalam memprediksi kelulusan siswa, yaitu “Prediksi Kelulusan Siswa SDN 016528 BP Mandoge dengan Metode Naïve Bayes”. Sistem ini diinginkan dapat digunakan oleh pihak sekolah untuk memprediksi kelulusan siswa setiap tahunnya [2].

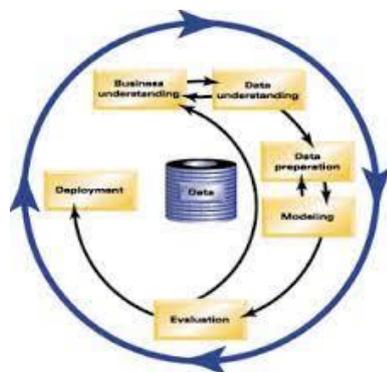
Kelulusan adalah faktor penting dalam proses akreditasi suatu lembaga pendidikan. Dengan demikian, jika banyak siswa yang lulus tepat waktu, hal ini dapat mempengaruhi nilai akreditasi lembaga tersebut [3]. Penurunan tingkat kelulusan siswa dapat berdampak pada akreditasi sekolah. Oleh sebab itu, penting untuk melakukan pemantauan dan evaluasi terhadap kecenderungan kelulusan siswa, baik yang tepat waktu maupun tidak [4]. Kelulusan siswa dipengaruhi oleh berbagai faktor, termasuk kondisi ekonomi, nilai akademik siswa, dan faktor lain yang terkait dengan sekolah. Untuk menangani masalah ini, diperlukan teknik yang dapat memprediksi kelulusan siswa. Dengan demikian dibutuhkan suatu model klasifikasi dengan Algoritma *Naive Bayes* yang dapat memprediksi dengan akurat tingkat kelulusan tepat waktu pada siswa, maka dibuat suatu system pendeteksi kelulusan pada SDN 016528 BP. Mandoge menggunakan metode *Naive Bayes* [5].

Penelitian yang dilaksanakan di SDN 016528 BP. Mandoge bertujuan untuk menerapkan Algoritma *Naive Bayes* untuk memprediksi kelulusan siswa. *Output* yang dihasilkan dapat menjadi solusi alternatif khususnya di SDN 016528 BP. Mandoge dalam menanggulangi permasalahan tingkat akurasi kelulusan tepat waktu pada siswa.

METODE

Data Mining adalah proses pengumpulan data yang selanjutnya diekstrak atau diolah untuk memperoleh knowledge baru. Data Mining menggabungkan berbagai bidang ilmu, seperti sistem basis data, gudang data, statistik, pembelajaran mesin, pengambilan informasi, dan komputasi tingkat tinggi, untuk mengelola data yang terus berkembang. Teknik yang digunakan dalam Data Mining bertujuan untuk menangani data dalam skala besar untuk menghasilkan kesimpulan dan keputusan yang paling tepat [6]. Data mining adalah proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi serta pengetahuan berharga dari berbagai database besar [7]. Data mining memiliki berbagai fungsi yang dapat diterapkan. Fungsi-fungsi ini bisa digabungkan dalam situasi tertentu untuk menyelesaikan masalah yang dihadapi [8].

Menurut Larose, Data Mining terdiri dari 6 fase dalam CRISP-DM (Cross Industry Standard Process for Data Mining).



Gambar 1. Fase CRISP-DM

Berdasarkan Gambar 1, terdapat enam fase dalam CRISP-DM sebagai berikut:

Bussines Understanding

Kelulusan menandakan selesainya suatu jenjang pendidikan di sekolah. Siswa di SDN 016528 BP Mandoge dinyatakan lulus dari sekolah setelah memenuhi beberapa syarat, yaitu menyelesaikan 12 semester dari program 6 tahun, yang ditunjukkan melalui rapor setiap semester, serta memperoleh nilai sikap atau perilaku minimal baik. Penelitian ini bertujuan untuk menganalisis cara yang efektif dalam mengklasifikasikan dan menentukan kelulusan siswa di SDN 016528 BP Mandoge.

Data Understanding

Data yang digunakan berasal dari SDN 016528 BP Mandoge di Kabupaten Asahan, khususnya dari Desa BP Mandoge, Kec. BP Mandoge. Data ini mencakup atribut seperti: Nomor Urut, NIS, NISN, Nama Peserta, Nilai Pendidikan Agama dan Budi Pekerti, Matematika, IPA, IPS, Seni Budaya, PKN, Bahasa Indonesia, dan Prakarya, Pendidikan Jasmani Olahraga dan Kesehatan, Jumlah Nilai, Rata-rata, serta Keterangan.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171 entries, 0 to 170
Data columns (total 15 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   NO URUT                                   171 non-null    int64
1   NIS                                       171 non-null    int64
2   NISN                                      171 non-null    int64
3   NAMA PESERTA                             171 non-null    object
4   PENDIDIKAN AGAMA DAN BUDI PEKERTI        171 non-null    float64
5   PPKN                                      171 non-null    float64
6   BHS INDONESIA                            171 non-null    float64
7   MATEMATIKA                               171 non-null    float64
8   IPA                                       171 non-null    float64
9   IPS                                       171 non-null    float64
10  SENI BUDAYA DAN PRAKARYA                 171 non-null    float64
11  PENDIDIKAN, JASMANI, OLAHRAGA DAN KESEHATAN 171 non-null    float64
12  JUMLAH                                    171 non-null    float64
13  RATA-RATA                                 171 non-null    float64
14  KET                                       171 non-null    object
dtypes: float64(10), int64(3), object(2)
memory usage: 20.2+ KB
```

Gambar 2. Deskripsi Data

Data Preparation

Dalam persiapan data, terdapat beberapa tahapan yang harus dilakukan, antara lain: pemilihan data, pembangunan data, integrasi data, dan pembersihan data.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171 entries, 0 to 170
Data columns (total 9 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   PENDIDIKAN AGAMA DAN BUDI PEKERTI        171 non-null    int32
1   PPKN                                      171 non-null    int32
2   BHS INDONESIA                            171 non-null    int32
3   MATEMATIKA                               171 non-null    int32
4   IPA                                       171 non-null    int32
5   IPS                                       171 non-null    int32
6   SENI BUDAYA DAN PRAKARYA                 171 non-null    int32
7   PENDIDIKAN, JASMANI, OLAHRAGA DAN KESEHATAN 171 non-null    int32
8   KET                                       171 non-null    object
dtypes: int32(8), object(1)
memory usage: 6.8+ KB
```

Gambar 3. Detail Data

Modelling

Naïve Bayes Classifier merupakan algoritma yang digunakan dalam metode klasifikasi dalam Data Mining. Algoritma ini mengandalkan prinsip statistik dengan memanfaatkan teori probabilitas untuk menyelesaikan masalah dalam Supervised Learning, di mana dataset memiliki label, kelas, atau target sebagai acuan atau panduan [9]. Naive Bayes adalah teknik klasifikasi probabilistik yang

sederhana, menghitung probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dalam dataset. Algoritma ini memanfaatkan Teorema Bayes dan menganggap bahwa semua atribut bersifat independen satu sama lain, berdasarkan nilai variabel kelas [10].

Naive Bayes adalah teknik prediksi sederhana yang berbasis probabilitas, mengacu pada teorema Bayes (aturan-aturan Bayes), dengan asumsi bahwa atribut-atributnya bersifat independen (tidak saling bergantung). Model yang diterapkan dalam Naive Bayes adalah "model fitur independen" [11].

Bentuk dasar dari *teorema bayes* :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan :

X : Kelas yang belum diketahui

H : Hipotesis data x

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (postreioriprobability)

P(H) : Probabilitas H (priorprobability)

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

P(X) : Probabilitas X

Untuk mendapatkan dataset dengan jumlah atribut yang lebih sedikit namun tetap informatif [12]. 171 Data dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, dengan proporsi 70:30.

```
x_train.shape
(119, 8)

x_test.shape
(52, 8)
```

Gambar 4. *Split Data*

Evaluation

Evaluasi bertujuan untuk memverifikasi dan memastikan bahwa data berfungsi dengan baik. Dalam penelitian ini, digunakan Confusion Matrix yang umumnya menampilkan metrik akurasi, precision, dan recall. Confusion Matrix ini menghasilkan dua kategori output, yaitu kelas negatif dan kelas positif [13].

Deployment

Hasil dari Data Mining dipresentasikan melalui visualisasi dengan bahasa pemrograman sehingga siswa dapat paham dan membaca informasi tersebut [14].

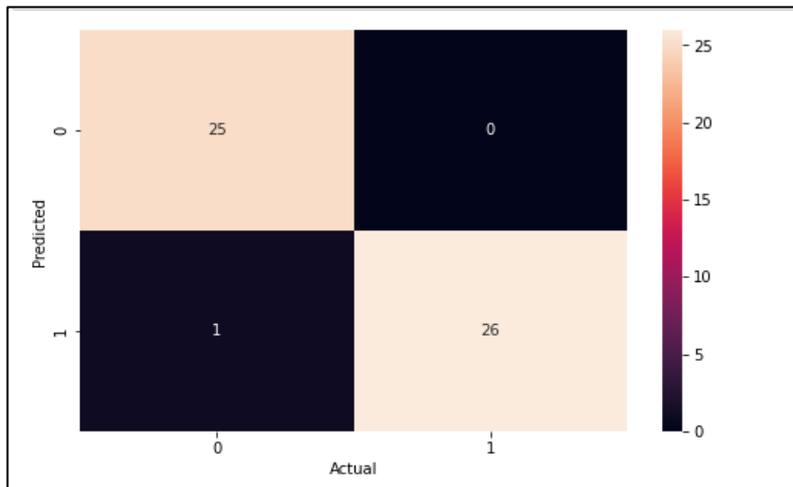
HASIL DAN PEMBAHASAN

Langkah pertama dalam analisis adalah mengimpor modul yang diperlukan dan dataset. Daftar Kolektif Nilai Dan Kelulusan Siswa SDN 016528 BP. Mandoge.

PENDIDIKAN AGAMA DAN BUDI PEKERTI	PPKN	BHS INDONESIA	MATEMATIKA	IPA	IPS	SENI BUDAYA DAN PRAKARYA	PENDIDIKAN, JASMANI, OLAHRAGA DAN KESEHATAN	KET
90	87	90	82	87	85	95	95	LULUS
70	50	65	70	78	65	78	80	TIDAK LULUS
66	50	63	72	65	55	70	88	TIDAK LULUS
97	95	92	90	95	95	93	95	LULUS
90	85	85	85	85	85	87	85	LULUS

Gambar 5. Dataset Kolektif Nilai dan Kelulusan Siswa

Setelah membangun model menggunakan data pelatihan, lakukan pengujian dengan menggunakan data uji. Hasil dari Confusion Matrix:



Gambar 6. Confusion Matrix

	precision	recall	f1-score	support
LULUS	0.96	1.00	0.98	25
TIDAK LULUS	1.00	0.96	0.98	27
accuracy			0.98	52
macro avg	0.98	0.98	0.98	52
weighted avg	0.98	0.98	0.98	52

Gambar 7. Nilai Akurasi

Berdasarkan Gambar 7, algoritma menghasilkan nilai precision sebesar 96% untuk kelas positif dan 100% untuk kelas negatif. Nilai recall mencapai 100% untuk kelas positif dan 96% untuk

kelas negatif. Sementara itu, nilai f1-score adalah 98% untuk kedua kelas, positif dan negatif. Akurasi keseluruhan metode Naïve Bayes adalah 98%.

KESIMPULAN

Sesuai analisis dengan model Naïve Bayes dan Confusion Matrix, diperoleh nilai precision 98%, recall 98%, f1-score 98%, dan akurasi 98%, yang tergolong kategori Klasifikasi Baik. Dengan demikian, metode Naïve Bayes dapat dikategorikan sebagai model yang efektif dan dapat diandalkan untuk memprediksi peluang berdasarkan data historis, serta mempermudah proses klasifikasi kelulusan siswa di SDN 016528 BP Mandoge.

DAFTAR PUSTAKA

- [1] T. Lesmana and M. Silalahi, "Jurnal Comasie," *Comasie*, vol. 3, no. 3, pp. 21–30, 2020.
- [2] Tumini and L. Damayanti, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes Studi Kasus Smp Negeri 11 Kotabumi Utara," *Inform. SIMANTIK*, vol. 3, no. 2, pp. 23–30, 2018.
- [3] S. I. Gain, "PREDIKSI KELULUSAN MAHASISWA DENGAN NAÏVE BAYES DAN FEATURE SELECTION INFORMATION GAIN STUDENT GRADUATION PREDICTION WITH NAÏVE BAYES AND FEATURE," vol. 4, no. 02, pp. 223–235, 2022.
- [4] N. Abdillah, S. Defit, and Sumijan, "Penggunaan metode klasifikasi Data Metode Naïve Bayes merupakan metode klasifikasi Data Mining dengan membandingkan algoritma Decision Tree dan Naive Bayes . Penelitian ini bertujuan," *Semin. Nas. Syedza Saintika*, pp. 281–287, 2020.
- [5] N. Yustira, D. Witarsyah, and ..., "Implementasi Algoritma Naïve Bayes Classification Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu," *eProceedings ...*, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16721%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16721/16429>
- [6] R. A. Iswanto, J. Sahertian, and M. A. D. Widyadara, "Pengembangan Sistem Prediksi Kelulusan Mahasiswa Berdasarkan Nilai Disiplin, Hasil Belajar, Aktivitas Sosial Ekonomi, dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Teknol. Dan Sains*, vol. 1, pp. 349–358, 2022.
- [7] E. Rilvani, A. B. Trisnawan, and P. P. Santoso, "Penentuan Kelulusan Siswa Yayasan Cerdas Bakti Pertiwi Dengan Menggunakan Algoritma Naive Bayes Dan Cross Validation," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 2, pp. 145–153, 2019.
- [8] I. G. I. Suardika, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi Dan Bisnis Universitas Pendidikan Nasional," *J. Ilmu Komput. Indones.*, vol. 4, no. 2, pp. 37–44, 2019, doi: 10.23887/jik.v4i2.2775.
- [9] Suparyanto dan Rosad (2015, "濟無No Title No Title No Title," *Suparyanto dan Rosad (2015*, vol. 5, no. 3, pp. 248–253, 2020.
- [10] R. Marbun, "Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier Studi Kasus: Poltekkes Kemenkes RI Medan," *JURIKOM (Jurnal Ris. Komputer)*, vol. 6, no. 6, pp. 627–633, 2020, [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/1887>
- [11] J. Sumpena and N. Kurnia H., "Analisis Prediksi Kelulusan Siswa PKBM Paket C Dengan Metoda Algoritma Naive Bayes," *Nina Kurnia H. TEDC*, vol. 13, no. 2, pp. 127–133, 2019.
- [12] A. Suwarno *et al.*, "Jurnal Teknologi Pelita Bangsa," *J. Teknol. Pelita Bangsa*, vol. 12, no. 4, pp. 33–40, 2021.
- [13] Y. Septiani and P. F. Ariyani, "Penerapan Algoritma Naive Bayes Menentukan Klasifikasi Tingkat Kelulusan Siswa SMK Media Informatika Jakarta Application of The Naive Bayes



- Algorithm Determining Classification of Students ' Graduation Level of Jakarta Media Informatika Vocational School," no. September, pp. 607–613, 2022.
- [14] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.