

# Implementasi Data Mining Menggunakan Algoritma *Naïve Bayes* Untuk Prediksi Kelulusan Siswa

Dimas Abisono Punkastyo<sup>\*1</sup>, Fajar Septian<sup>2</sup>, Ari Syaripudin<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang  
Email: <sup>1</sup>dosen00675@unpam.ac.id, <sup>2</sup>dosen00677@unpam.ac.id, <sup>3</sup>dosen00671@unpam.ac.id

## Abstrak

Proses kelulusan siswa melibatkan serangkaian tahapan dan persyaratan yang harus dipenuhi oleh setiap siswa. Kelulusan ini menjadi langkah penting dalam pendidikan, karena menentukan kemampuan siswa untuk melanjutkan ke jenjang berikutnya. Dalam penelitian ini, data kelulusan siswa diolah menggunakan teknik data mining, khususnya metode *Naïve Bayes*, untuk melakukan prediksi kelulusan siswa. Atribut-atribut seperti Nilai Praktik, Ujian Sekolah (US), Ujian Nasional (UN), dan Perilaku Siswa digunakan dalam proses prediksi ini. Hasil penelitian menunjukkan bahwa dari populasi sebanyak 60 siswa, model berhasil memprediksi bahwa 45 siswa akan lulus dan 15 siswa tidak akan lulus. Metode *Naïve Bayes* diimplementasikan menggunakan aplikasi Orange dan menghasilkan tingkat akurasi sebesar 98,33% dengan tingkat presisi sebesar 100,00%. Hasil ini menunjukkan bahwa metode *Naïve Bayes* sangat efektif dalam memprediksi kelulusan siswa dengan tingkat akurasi yang tinggi. Oleh karena itu, teknik data mining ini dapat menjadi alat yang berharga bagi institusi pendidikan dalam meningkatkan kualitas hasil kelulusan siswa.

**Kata Kunci:** *Prediksi Kelulusan Siswa, Data Mining, Metode Naïve Bayes*

## Abstract

*The student graduation process involves a series of stages and requirements that must be met by each student. Graduation is an important step in education because it determines the student's ability to continue to the next level. In this research, student graduation data is processed using data mining techniques, especially the Naïve Bayes method, to predict student graduation. Attributes such as Practice Scores, School Exams (US), National Exams (UN), and Student Behavior are used in this prediction process. The research results show that from a population of 60 students, the model successfully predicted that 45 students would pass, and 15 students would not pass. The Naïve Bayes method was implemented using the orange application and produced an accuracy level of 98.33% with a precision level of 100.00%. These results show that the Naïve Bayes method is very effective in predicting student graduation with a high level of accuracy. Therefore, this data mining technique can be a valuable tool for educational institutions in improving the quality of student graduation outcomes.*

**Keywords:** *Student Graduation Predictions, Data Mining, Naïve Bayes Methods*

## 1. Pendahuluan

Lulusan siswa memiliki signifikansi yang penting dalam kerangka pendidikan, termasuk pemenuhan persyaratan kelulusan dan penyelesaian rapat kelulusan. Tingkat kelulusan siswa sering menjadi landasan dalam merancang kebijakan baru dan strategi untuk meningkatkan mutu pendidikan (Etriyanti et al., 2020). Namun, proses

pengolahan data kelulusan masih sering mengalami kesalahan dan kurang akurat. Oleh karenanya, tujuan penelitian ini untuk mengatasi masalah tersebut dengan menggunakan metode *Naïve Bayes*. Data Mining, sebuah proses penambangan informasi dari data berukuran besar, memfokuskan pada identifikasi pola atau aturan tertentu dalam data. Tujuannya adalah memberikan solusi yang dapat mengatasi berbagai situasi atau masalah khusus. Data Mining dapat didefinisikan sebagai proses mengekstrak informasi menarik dari data tertentu menggunakan teknik atau metode tertentu (Muchlis et al., 2021). Salah satu teknik utama dalam data mining adalah klasifikasi. Klasifikasi adalah proses mengidentifikasi atribut- Teknik ini memungkinkan penggunaan algoritma untuk mengevaluasi dan memprediksi kelompok atau kategori data. Sebuah analisis literatur singkat mengungkapkan bahwa algoritma seperti C4.5, *Naïve Bayes*, dan k-Nearest Neighbor (k-NN) sering digunakan dalam klasifikasi data (Al Arif et al., 2022). Algoritma C4.5 efektif dalam mengelola dataset besar dan rumit tetapi memiliki kekurangan dalam menangani tumpang tindih dan merancang pohon keputusan yang optimal. Sementara k-NN digunakan untuk masalah kelas ganda tetapi memiliki kendala dalam menentukan tetangga terdekat pada titik kueri dari dataset.

*Naïve Bayes* merupakan metode klasifikasi probabilitas sederhana yang berdasarkan pada teorema Bayes dengan asumsi bahwa semua atribut bersifat independen atau tidak berhubungan (A. Oktian Permana & Sudin Saepudin, 2023; Rifai et al., 2019; Rovidatul et al., 2023). Metode ini memerlukan data pelatihan yang relatif sedikit untuk menghitung parameter dan telah terbukti efektif dalam klasifikasi dalam berbagai penelitian sebelumnya. Meskipun demikian, belum ada pengetahuan yang cukup tentang bagaimana mengimplementasikan metode *Naïve Bayes* dengan efektif dalam konteks prediksi kelulusan siswa di sekolah. Penelitian ini bertujuan untuk mengisi celah ini dengan membandingkan praktik, ujian sekolah, ujian nasional, dan perilaku siswa sebagai faktor yang dominan dalam prediksi kelulusan siswa. Penelitian ini memberikan kontribusi baru dalam aplikasi metode *Naïve Bayes* dalam prediksi kelulusan siswa di sekolah, yang dapat membantu sekolah merumuskan strategi berdasarkan analisis data ini. Hasil dari penelitian ini diharapkan akan meningkatkan efisiensi pengambilan keputusan terkait kelulusan siswa dan meningkatkan daya saing sekolah.

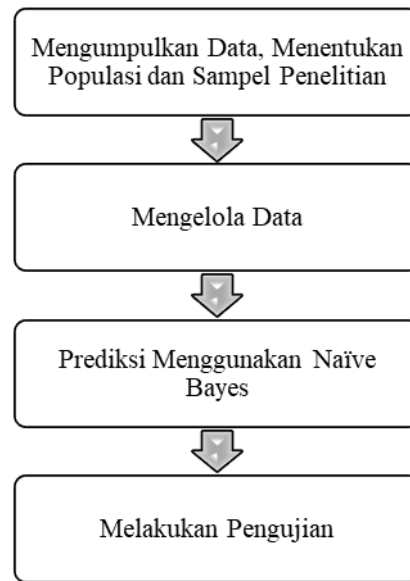
## 2. Metodologi Penelitian

Penelitian ini memanfaatkan teknik klasifikasi dari data mining, yaitu metode *Naïve Bayes*. Metode ini merupakan metode klasifikasi probabilitas sederhana yang berdasarkan teorema Bayes dengan asumsi bahwa semua atribut independen atau tidak saling berhubungan (Borman & Wati, 2020). Metode ini efektif dengan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter (Putro et al., 2021). Data yang digunakan dalam penelitian ini mencakup nilai-nilai Praktik, Ujian Sekolah (US), Ujian Nasional (UN), dan perilaku siswa. Data ini kemudian dianalisis menggunakan metode *Naïve Bayes* untuk menentukan faktor mana yang menjadi penunjang yang paling dominan dalam mempengaruhi tingkat kelulusan siswa.

### 2.1. Tahapan Penelitian

Tahapan penelitian merujuk pada serangkaian langkah yang harus diikuti untuk melaksanakan investigasi secara sistematis terhadap suatu permasalahan atau pertanyaan penelitian (Fernando et al., 2022). Dalam rangka menjelaskan pengolahan

tingkat kelulusan siswa, akan diuraikan langkah-langkah penelitian yang diilustrasikan dalam Gambar 1 berikut:



Gambar 1. Tahapan-Tahapan Penelitian

Pada Gambar 1 memperlihatkan proses langkah-langkah dalam melaksanakan penelitian. Secara lebih rinci setiap fasenya dijelaskan sebagai berikut:

1) Mengumpulkan Data, Menentukan Populasi dan Sampel Penelitian

Penelitian ini dimulai dengan melakukan wawancara bersama Ibu Neng Wida Novira, seorang staff guru di sekolah, yang bertujuan untuk mengumpulkan data yang diperlukan. Sejumlah data dikumpulkan melalui metode observasi, wawancara, dan studi pustaka. Pendekatan ini digunakan untuk mengumpulkan informasi yang dibutuhkan untuk penelitian. Sumber data terdiri dari data primer dan data sekunder. Data primer diperoleh langsung dari sumber yang relevan dengan penelitian, seperti data nilai praktik, nilai Ujian Sekolah (US), nilai Ujian Nasional (UN), dan perilaku siswa. Data primer diperoleh melalui observasi dan wawancara dengan salah satu staf guru di sekolah, yaitu Ibu Neng Wida Novira. Data sekunder adalah data tambahan yang diperoleh dari buku dan jurnal yang tersedia di internet (Sukamto & Fauziah, 2020). Populasi penelitian ini terdiri dari data tingkat kelulusan siswa, yang mencakup siswa sebanyak 60 siswa yang diambil sampelnya dari tahun akademik 2016 sampai dengan 2018. Sampel diambil dengan menggunakan teknik random sampling pada masing-masing tahun akademik. Data yang digunakan dari tahun 2016 sampai dengan 2018 karena data tiga tahun tersebut telah cukup representatif untuk digunakan sebagai prediksi. Tahun 2019 hingga 2021 tidak digunakan karena pada tahun itu sedang Pandemi Covid 19 sehingga data tidak dapat dijadikan acuan.

2) Mengelola Data

Beberapa instrumen penting digunakan pada penelitian ini. Pertama, data riset yang akan dimasukkan dan diolah terdiri dari nilai praktik, nilai Ujian Sekolah (US), nilai Ujian Nasional (UN), dan perilaku siswa. Data ini adalah sampel yang mewakili populasi siswa dan digunakan dalam proses klasifikasi untuk memprediksi tingkat kelulusan siswa. Kedua, data tersebut sudah tersedia dalam format Microsoft Excel. Dataset ini mencakup 60 siswa, baik yang lulus maupun yang tidak, yang diambil secara random sampling dari tahun akademik 2016-2018. Ketiga, perangkat lunak

yang digunakan untuk pengolahan data adalah Microsoft Excel 2010. Pengolahan data sampel yang identik dengan populasi dilakukan dengan menggunakan perangkat Microsoft Excel. Langkah ini dimaksudkan untuk mendapatkan hasil pengujian data. Selain itu, aplikasi khusus digunakan untuk mengelola dan menguji data uji. Selanjutnya, data tersebut diolah dan dianalisis sesuai dengan prosedur dan teknik pengolahan data yang telah ditentukan. Proses pengolahan data meliputi beberapa langkah. Pertama, dilakukan pemilihan dan penyusunan klasifikasi data agar dapat dikategorikan dengan baik. Selanjutnya, setiap data diberikan kode-kode tertentu guna mempermudah analisis lebih lanjut. Terakhir, dilakukan analisis data sesuai dengan klasifikasinya yang telah ditetapkan sebelumnya.

### 3) Prediksi Menggunakan *Naïve Bayes*

Setelah data dikelola proses selanjutnya adalah memasukkan data tersebut untuk dilakukan prediksi. Untuk melakukan prediksi menggunakan pendekatan *Naïve Bayes*. Pendekatan ini diterapkan pada data mulai dari tahap pelatihan hingga pengujian untuk mendapatkan hasil prediksi. Pendekatan *Naïve Bayes* akan diimplementasikan pada aplikasi Orange, dimana ini merupakan toolkit *open-source* untuk visualisasi data, pembelajaran mesin, dan penambangan data.

### 4) Melakukan Pengujian

Hasil prediksi yang dilakukan oleh algoritma *Naïve Bayes* selanjutnya dilakukan pengujian untuk mengukur kinerja dari prediksi yang telah dihasilkan. Pengukuran kinerjanya melalui *accuracy*, *precision*, *recall*, dan *F1 Score*.

## 2.2. Data Mining

Data mining merupakan proses penelusuran pola atau aturan dalam suatu kumpulan data berukuran besar guna menemukan wawasan baru (Puspita et al., 2023). Proses ini juga dapat didefinisikan sebagai metode yang digunakan untuk menentukan struktur data (Gultom & Rikki, 2020). Karakteristik dari data mining (Yuniarti & Gunawan, 2021), mencakup: (1) berkaitan dengan penemuan pola data baru yang sebelumnya tidak diketahui; (2) penggunaan dataset besar untuk menghasilkan temuan signifikan; dan (3) aplikasinya dalam penetapan strategi dan keputusan kritis.

## 2.3. *Naïve Bayes*

Metode *Naïve Bayes* merupakan sebuah teknik klasifikasi atau pengelompokan data yang memanfaatkan perhitungan probabilitas berdasarkan dataset yang telah terkumpul (Alfa Saleh, 2019). Sesuai dengan penjelasan Bustami (Bustami, 2014), *Naïve Bayes* digunakan untuk meramalkan peluang masa depan dengan memanfaatkan metode probabilitas dan analisis statistik berdasarkan pengalaman sebelumnya (Wijaya & Dwiasnati, 2020).

Salah satu keunggulan utama dari algoritma *Naïve Bayes* adalah kebutuhan akan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter dalam proses klasifikasi. Selain itu, implementasi *Naïve Bayes* relatif mudah dan sering memberikan hasil yang baik dalam berbagai situasi (Chely Aulia Misrun et al., 2023). Namun, kekurangan dari metode ini adalah asumsi independensi antar fitur. Dalam kenyataannya, keterkaitan antar fitur seringkali ada dan tidak dapat dimodelkan oleh *Naïve Bayesian Classifier*. *Teorema Bayes* memiliki bentuk universal, untuk melakukan perhitungannya melalui persamaan (1).

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

Berikut adalah penjelasannya dari persamaan (1) tersebut.

X: Data dengan kelas yang belum diidentifikasi.

H: Hipotesis bahwa data termasuk dalam suatu kelas tertentu.

P(H|X): Probabilitas hipotesis H berdasarkan kondisi X, juga dikenal sebagai probabilitas posterior.

P(H): Probabilitas hipotesis H, juga dikenal sebagai probabilitas prior.

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

P(X): Probabilitas X.

### 3. Hasil dan Pembahasan

Pada bagian ini, kami akan mengulas hasil dari penggunaan metode *Naïve Bayes*, sebuah teknik klasifikasi dalam Data Mining, pada data penelitian yang melibatkan nilai-nilai Praktik, Ujian Sekolah (US), Ujian Nasional (UN), dan Perilaku Siswa. Metode ini memungkinkan untuk mengidentifikasi faktor-faktor dominan yang mempengaruhi tingkat kelulusan siswa. Melalui analisis ini, dapat digunakan untuk memahami bagaimana masing-masing variabel berkontribusi terhadap hasil akhir dan memberikan wawasan penting untuk meningkatkan tingkat kelulusan di masa mendatang. Penyelesaian kasus ini menggunakan aplikasi Orange, dimana ini merupakan toolkit *open-source* untuk visualisasi data, pembelajaran mesin, dan penambangan data. Orange juga mencakup seperangkat alat untuk pra-pemrosesan data yang efektif, evaluasi model, dan lainnya (Hozairi et al., 2021). *Toolkit* ini menyediakan antarmuka pemrograman visual, sehingga pengguna dapat dengan mudah melakukan analisis, eksplorasi dan analisis data dengan mudah.

#### 3.1. Persiapan Dataset

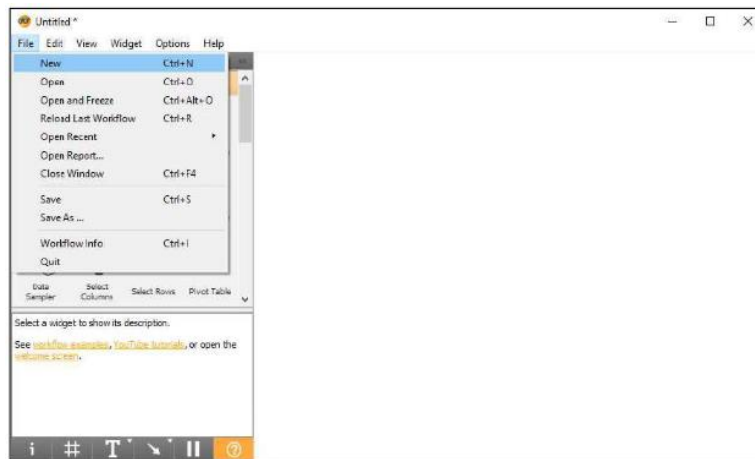
Data yang digunakan disebut sebagai `data_training_1.xlsx`. Dataset ini mencakup 60 siswa, baik yang lulus maupun yang tidak, yang diambil dari tahun akademik 2016 hingga 2018. Populasi penelitian ini terdiri dari data tingkat kelulusan siswa, yang mencakup siswa sebanyak 60 siswa yang diambil sampelnya dari tahun akademik 2016 sampai dengan 2018. Sampel diambil dengan menggunakan teknik random sampling pada masing-masing tahun akademik. Data yang digunakan dari tahun 2016 sampai dengan 2018 karena data tiga tahun tersebut telah cukup representatif untuk digunakan sebagai prediksi. Tahun 2019 hingga 2021 tidak digunakan karena pada tahun itu sedang Pandemi Covid 19 sehingga data tidak dapat dijadikan acuan. Atribut yang digunakan antara lain nilai praktik, nilai Ujian Sekolah (US), nilai Ujian Nasional (UN), dan perilaku siswa. Data dikumpulkan melalui observasi berdasarkan data kelulusan sekolah dan wawancara dengan salah satu staf guru di sekolah, yaitu Ibu Neng Wida Novira. Data yang didapatkan kemudian dipindah ke Microsoft Excel dan kemudian diperiksa kembali terkait kelengkapan data. Data sebanyak 60 siswa yang digunakan sebagai dataset merupakan hasil seleski data yang seluruh atribut yang terdiri dari nilai praktik, nilai Ujian Sekolah (US), nilai Ujian Nasional (UN), dan perilaku siswa lengkap dan terisi. Seluruh atribut terdokumentasi datanya dan digunakan untuk melakukan prediksi. Prediksi kelulusan berdasarkan pada label Status Kelulusan, yang

terdapat dua label yaitu Lulus dan Tidak Lulus. Dari 60 data yang digunakan berikut merupakan 10 data sampel dari dataset yang digunakan yang ditampilkan pada Tabel 1.

Tabel 1. Data Nilai Siswa

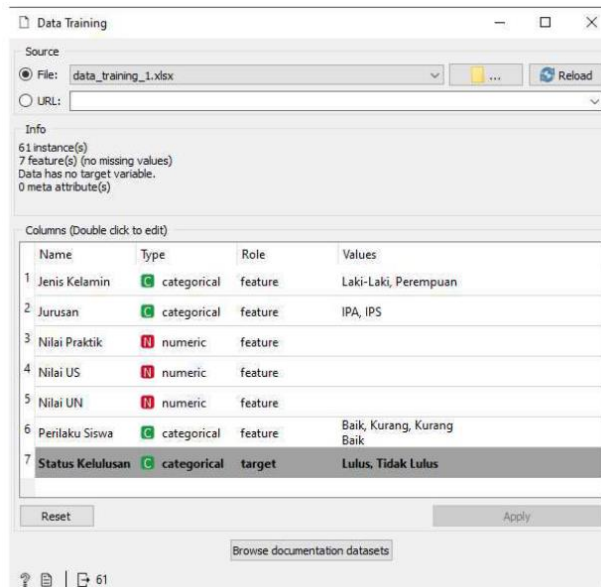
No	Nama	Jenis	Nilai Praktik	Nilai US	Nilai UN	Perilaku Siswa	Status Kelulusan
1	Ahmad Taji Saputra	L	95	88	33.35	Baik	Lulus
2	Agung	L	83	91	30	Baik	Lulus
3	Ahmad Gunadi	L	80	90	31.23	Baik	Lulus
4	Ahmad Resal	L	84	77	30.55	Baik	Lulus
5	Ahmad Ripai	L	83	82	31.38	Baik	Lulus
6	Ajeng Siti Khodijah	P	65	66	20.01	Kurang Baik	Tidak Lulus
7	Aldo	L	80	97	33.21	Baik	Lulus
8	Dani	P	70	70	20.1	Kurang Baik	Tidak Lulus
9	Derince	L	90	82	32.57	Baik	Lulus
10	Diana Lestari	P	95	77	32.5	Baik	Lulus

Dari dataset tersebut kemudian dimasukkan pada aplikasi Orange sebagai data pelatihan. Sebelum melakukan pengolahan data untuk memulai analisis pada aplikasi Orange dimulai dari membuat *project* baru dengan membuka menu 'New'. Menu ini merupakan titik awal bagi pengguna untuk memulai analisis, dengan memungkinkan untuk membuat proyek data mining baru dan menentukan parameter dan variabel yang relevan. Dengan demikian, menu 'New' ini berfungsi sebagai alat penting dalam proses penambangan data, memfasilitasi persiapan dan pengolahan data sebelum analisis lebih lanjut dilakukan. Gambar 2 menampilkan menu 'New' yang digunakan untuk memulai proses perhitungan data mining baru.



Gambar 2. Memulai Proses Perhitungan Data Mining

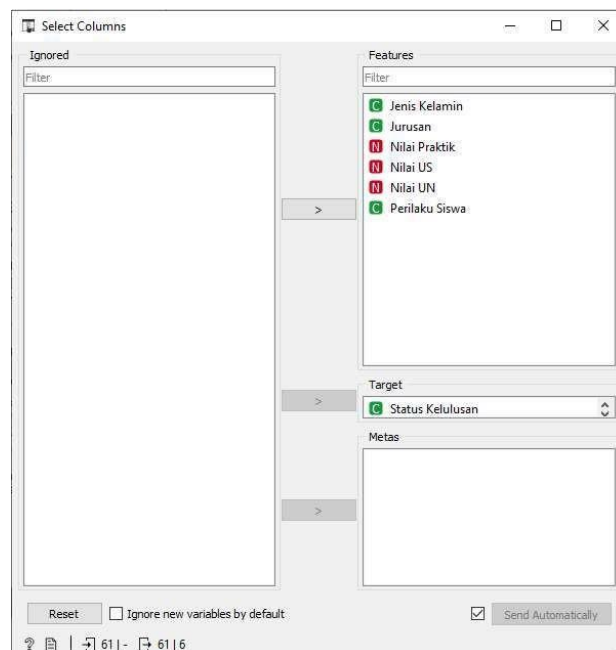
Dalam contoh ini, pengguna memilih file dengan nama 'data\_training\_1.xlsx'. File ini adalah kumpulan data pelatihan yang akan dipergunakan dalam proses data mining. Dataset ini biasanya berisi sejumlah besar entri data yang relevan untuk masalah atau pertanyaan penelitian tertentu. Dengan memilih file 'data\_training\_1.xlsx', pengguna mengatur dasar untuk analisis lebih lanjut, dan langkah selanjutnya akan melibatkan penerapan teknik dan algoritma data mining pada dataset ini untuk menghasilkan wawasan yang berharga. Gambar 3 menunjukkan proses pemilihan dataset untuk analisis.



Gambar 3. Memilih dataset dengan nama file data\_training\_1.xlsx

### 3.2 Select Columns

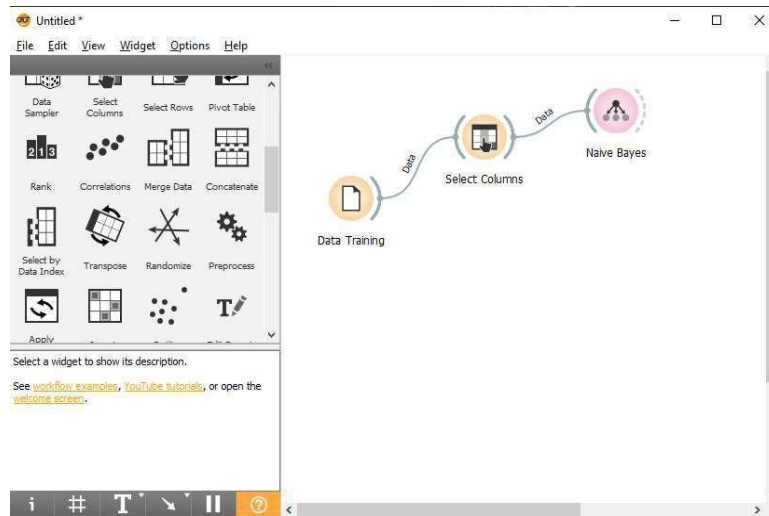
*Select Columns* menerima input dari data dan fitur, kemudian menghasilkan *output* berupa data dan fitur. Pada tahap ini, pengguna dapat memilih kolom spesifik dari dataset yang akan digunakan dalam analisis. Pemilihan kolom ini penting karena memungkinkan pengguna untuk fokus pada variabel atau fitur tertentu dari data, dan mengabaikan informasi lain yang mungkin tidak relevan untuk tujuan analisis data. Tahapan memilih fitur atribut dan target dari dataset tersaji pada Gambar 4.



Gambar 4. Memilih Fitur Atribut dan Target

### 3.3 Naïve Bayes

Langkah selanjutnya yaitu mengimplementasikan algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* telah tersedia di aplikasi Orange. Tampilan menambahkan algoritma *Naïve Bayes* tersaji pada Gambar 5.

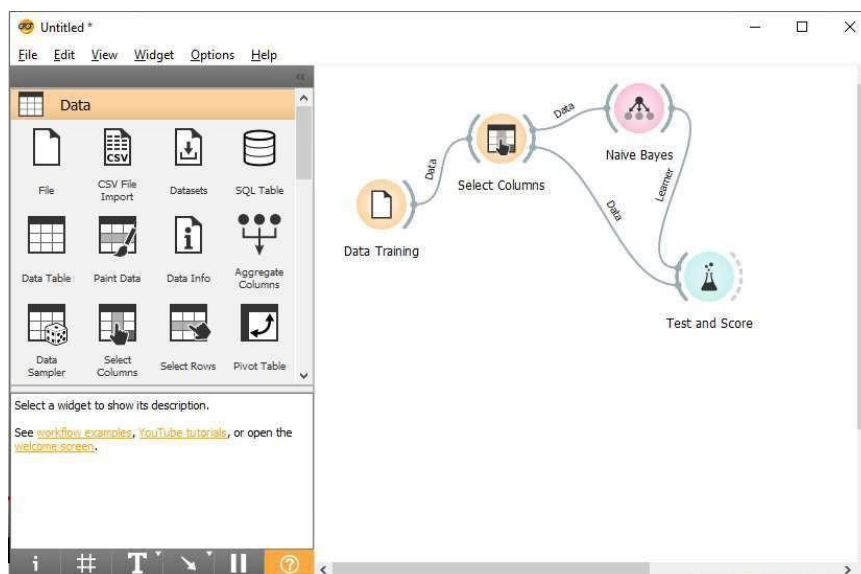


Gambar 5. Menambahkan komponen *Naïve Bayes*

Gambar 5 menunjukkan tahap penambahan algoritma *Naïve Bayes* ke dalam analisis data, yang digunakan untuk membantu dalam prediksi atau klasifikasi berdasarkan fitur dan target yang telah dipilih.

### 3.4 *Test dan Score*

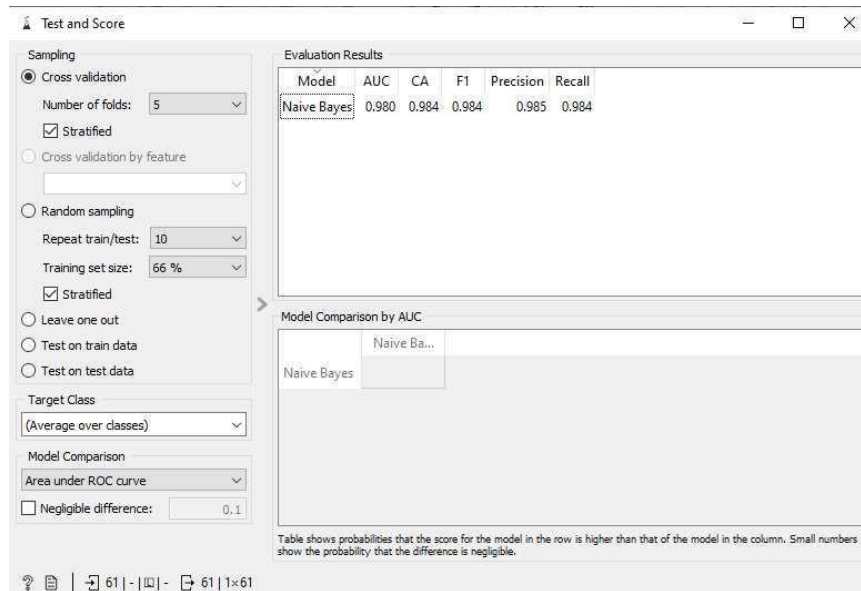
Input terdiri dari Data, Test Data, Preprocessor, dan Learner. *Output* yang dihasilkan adalah Prediksi dan Evaluasi Hasil. Di tahap ini, model yang telah dibuat dan dilatih sebelumnya diuji menggunakan set data tes untuk melihat sejauh mana kinerja dan akurasi model dalam membuat prediksi atau klasifikasi. Gambar 6 menggambarkan tahap *Test dan Score* dalam proses analisis data.



Gambar 6. *Test dan Score*



*Output* ini biasanya berupa laporan atau visualisasi yang merangkum hasil pengujian model dan metrik evaluasinya. Laporan ini dapat mencakup berbagai elemen, seperti matriks kebingungan, skor akurasi, skor presisi, skor recall, kurva ROC, dan lainnya. Dengan memeriksa output ini, pengguna dapat memahami kinerja model pada data tes dan menentukan apakah model tersebut sudah cukup baik atau perlu penyesuaian lebih lanjut. Gambar 7 menampilkan output dari tahap 'Test dan Score' dalam proses analisis data.



Gambar 7. Output Test dan Score

Dari perhitungan yang telah dilakukan menghasilkan hasil berikut:

- AUC = 0,980
- CA = 0,984
- F1 = 0,984
- Precision = 0,985
- Recall = 0,984

### 3.5 Uji coba prediksi kelulusan dengan data Testing

Dataset yang digunakan bernama data\_testing\_1.xlsx. yang telah disiapkan pada aplikasi Orange sebagai data *testing*. Data pengujian tersebut ditampilkan pada Tabel 2.

Tabel 2. Atribut Data *Testing*

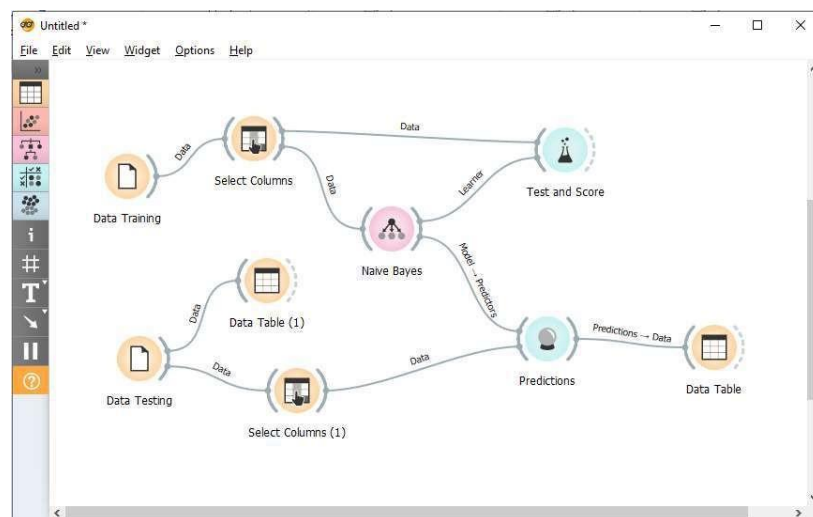
No	Nama	Jenis	Nilai Praktik	Nilai US	Nilai UN	Perilaku Siswa
1	a	P	93	78	30.22	Kurang Baik
2	b	L	91	83	30.17	Baik
3	c	P	87	78	32.19	Baik
4	d	P	65	70	20.12	Baik
5	e	P	82	99	30	Baik
6	f	L	65	65	20.17	Kurang Baik
7	g	L	83	83	30.2	Baik

Berdasarkan Pada Tabel 2 tersebut, kemudian dijadikan sebagai data uji. Gambar 8 menunjukkan tahap 'Data Testing', di mana model yang telah dilatih diuji dengan data baru untuk memverifikasi kinerja dan tingkat generalisasinya.

nama	Jenis Kelamin	Jurusan	Nilai Praktik	Nilai US	Nilai UN	Perilaku Siswa
1 a	Perempuan	IPA	93	78	30.22	Kurang Baik
2 b	Laki-Laki	IPA	91	83	30.17	Baik
3 c	Perempuan	IPA	87	78	32.19	Baik
4 d	Perempuan	IPS	65	70	20.12	Baik
5 e	Perempuan	IPS	82	99	30.00	Baik
6 f	Laki-Laki	IPS	65	65	20.17	Kurang Baik
7 g	Laki-Laki	IPS	83	83	30.20	Baik

Gambar 8. Data Testing

Selanjutnya dibuat susunan elemen dalam aplikasi Orange, mencakup alat dan modul untuk analisis data, dari pengolahan awal hingga evaluasi model. Susunan ini membantu memahami alur kerja analisis data. Susunan elemen tersebut tersaji pada Gambar 9.



Gambar 9. Susunan Elemen dalam aplikasi Orange

### 3.6 Hasil Prediksi

Hasil prediksi yang didapatkan dari aplikasi orange menggunakan algoritma *Naive Bayes*. Hasil ini memungkinkan evaluasi kemampuan generalisasi model. Hasil prediksi berdasarkan data uji yang dilakukan disajikan pada Gambar 10.

nama	Naive Bayes	Naive Bayes (Lulus)	Naive Bayes (Tidak Lulus)	Jenis Kelamin	Jurusan	Nilai Praktik	Nilai US	Nilai UN	Perilaku Siswa
1 a	Lulus	0.974165	0.0258352	Perempuan	IPA	93	78	30.22	Kurang Baik
2 b	Lulus	0.998836	0.00116437	Laki-Laki	IPA	91	83	30.17	Baik
3 c	Lulus	0.997302	0.00269842	Perempuan	IPA	87	78	32.19	Baik
4 d	Tidak Lulus	0.000513386	0.999487	Perempuan	IPS	65	70	20.12	Baik
5 e	Lulus	0.99921	0.000790007	Perempuan	IPS	82	99	30.00	Baik
6 f	Tidak Lulus	0.000112003	0.999888	Laki-Laki	IPS	65	65	20.17	Kurang Baik
7 g	Lulus	0.999398	0.000602024	Laki-Laki	IPS	83	83	30.20	Baik

Gambar 10. Hasil Prediksi

Rincian hasil prediksi menggunakan algoritma *Naïve Bayes* dapat dilihat pada Gambar 10. Dari prediksi terhadap tujuh data training yang diuji, ditemukan hasil prediksi lulus untuk lima siswa dan prediksi tidak lulus untuk dua siswa.

#### 4. Kesimpulan

Dari hasil penelitian yang telah dikerjakan, maka dapat ditarik kesimpulan bahwa metode *Naïve Bayes* efektif dalam melakukan prediksi tingkat kelulusan siswa. Metode ini tidak hanya memudahkan dalam menentukan hasil kelulusan siswa, tetapi juga dapat memprediksi jumlah siswa yang akan lulus. Tingkat akurasi prediksi dari data tingkat kelulusan siswa mencapai 98,33%, menunjukkan bahwa metode *Naïve Bayes* sangat akurat dalam pengolahan tingkat kelulusan siswa. Hasil dari penelitian ini dapat dimanfaatkan sebagai bahan perbandingan atau referensi untuk penelitian lain yang terkait dengan prediksi tingkat kelulusan menggunakan metode *Naïve Bayes* atau algoritma lain. Selain itu, penggabungan atau perbandingan dengan algoritma lain dapat dilakukan untuk mendapatkan hasil prediksi yang lebih akurat.

#### Ucapan Terimakasih

Peneliti mengucapkan terima kasih kepada Universitas Pamulang dan LPPM Universitas Pamulang yang telah memberikan dukungan terhadap penelitian ini melalui Hibah Internal, sehingga penelitian ini dapat terlaksana.

#### Daftar Pustaka

- A. Oktian Permana, & Sudin Saepudin. (2023). Perbandingan algoritma k-nearest neighbor dan naïve bayes pada aplikasi shopee. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 25–32. <https://doi.org/10.37859/coscitech.v4i1.4474>
- Al Arif, A., Firdaus, M., Rahmaddeni, & Maruhawa, Y. (2022). Perbandingan Metode Data Mining untuk Prediksi Curah Hujan dengan Algoritma C4.5, Naïve Bayes, dan KNN. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 187–197.
- Alfa Saleh. (2019). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 73–81. <https://doi.org/10.20895/inista.v1i2.73>
- Borman, R. I., & Wati, M. (2020). Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Fakultas Ilmu Komputer*, 9(1), 25–34.
- Bustami. (2014). Penerapan Algoritma Naïve Bayes. *Jurnal Informatika*, 8(1), 884–898.
- Chely Aulia Misrun, Haerani, E., Fikry, M., & Budianita, E. (2023). Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naïve bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 207–215. <https://doi.org/10.37859/coscitech.v4i1.4790>
- Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining

- Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, 13(1), 56–67. <https://doi.org/10.35671/telematika.v13i1.881>
- Fernando, Y., Napianto, R., & Borman, R. I. (2022). Implementasi Algoritma Dempster-Shafer Theory Pada Sistem Pakar Diagnosa Penyakit Psikologis Gangguan Kontrol Impuls. *Insearch (Information System Research) Journal*, 2(2), 46–54.
- Gultom, J. P., & Rikki, A. (2020). Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin. *Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer (KAKIFIKOM)*, 02(01), 11–19.
- Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133. <https://doi.org/10.21107/nero.v6i2.237>
- Muchlis, M. M., Fitri, I., & Nuraini, R. (2021). Rancang Bangun Aplikasi Data Mining pada Penjualan Distro Bloods Berbasis Web menggunakan Algoritma Apriori. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 5(1), 26–33.
- Puspita, D. S., Shofa, S. H., & Agustia, H. (2023). Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama. *SMARTICS Journal*, 9(1), 14–19. <https://doi.org/10.21067/smartics.v9i1.8088>
- Putro, M. F., Prayitno, E., Siregar, J., & Muharrom, M. (2021). Penerapan Data Mining dengan Naive Bayes untuk Klasifikasi Siswa Sekolah Menengah Dalam Penentuan Perguruan Tinggi. *Jurnal Akrab Juara*, 6(2), 307–309.
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naive Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *Petir*, 12(2), 131–144. <https://doi.org/10.33322/petir.v12i2.471>
- Rovidatul, Yunus, Y., & Nurcahyo, G. W. (2023). Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), 193–199. <https://doi.org/10.37859/coscitech.v4i1.4755>
- Sukanto, R. N., & Fauziah, P. (2020). Identifikasi Pola Asuh di Kota Pontianak. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 5(1), 923–930. <https://doi.org/10.31004/obsesi.v5i1.638>
- Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>
- Yuniarti, D., & Gunawan, B. (2021). Penerapan Data Mining untuk Memprediksi Jumlah Siswa DIKTUKPA TNI AL Dengan Metode Artificial Neural Network dan Algoritma Backpropagation. *JISAMAR (Journal of ...)*, 5(4), 938–954. <https://doi.org/10.52362/jisamar.v5i4.574>