JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 3, November 2026, pp. 21 - 35
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/175

Naïve Bayes Decision Support for Junior High Student Graduation Prediction: Kartika VI Jayapura Study

Daniel Bienfield Manahan Siahaan[1], Heru Sutejo[2]
[1,2,] Sistem Informasi, Universitas Sepuluh Nopember Papua
Jl. Ardipura Raya No. 22B, 99222, Ardipura, Distrik Jayapura Selatan, Kota Jayapura, Papua, Indonesia
e-mail: danielsiahaan2603@gmail.com [1],heru.sutejo01@gmail.com [2]

ARTICLE INFO
Article History:

Received: 23-04-2026
Revised: 29-05-2026
Accepted: 24-07-2026
Published: 27-10-2026

ABSTRACT
This study aims to implement the Naïve Bayes algorithm as a decision support tool for predicting junior high school student graduation at SMP Kartika VI Jayapura. Student graduation is an important indicator in assessing the success of the learning process at school. However, students' academic and non-academic data are often not optimally utilized to support decision-making. Therefore, this study develops a prediction model to help the school identify students who are likely to graduate and students who are at risk of not graduating at an earlier stage. The dataset used in this study consists of 289 student records, divided into 202 training records and 87 testing records. The predictor attributes include Mathematics Score, Indonesian Language Score, English Language Score, Science Score, Social Science Score, Attitude, Attendance Percentage, School Rule Violations, Extracurricular Activeness, and Participation in Tutoring, while the predicted label is Graduation Status. The research process consists of data collection, data cleaning, modelling, and evaluation using RapidMiner. The results show that the Naïve Bayes model achieves an accuracy of 82.76%, indicating that the model performs fairly well in predicting student graduation. However, the evaluation also shows that the model is stronger in identifying the Graduated category than the Not Graduated category. Therefore, this model is more appropriate to be used as an early warning and decision support tool, not as the sole basis for determining student graduation. The prediction results can help the school identify students at risk of not graduating so that guidance, mentoring, and academic support can be provided more accurately and effectively.

Keywords: *Naïve Bayes, Data Mining, Decision Support, Graduation Prediction, Junior High School Students*

Abstrak
Penelitian ini bertujuan untuk mengimplementasikan algoritma *Naïve Bayes* sebagai alat bantu pendukung keputusan dalam memprediksi kelulusan siswa SMP di SMP Kartika VI Jayapura. Kelulusan siswa merupakan salah satu indikator penting dalam menilai keberhasilan proses pembelajaran di sekolah. Namun, data akademik dan nonakademik siswa sering kali belum dimanfaatkan secara optimal untuk mendukung proses pengambilan keputusan. Oleh karena itu, penelitian ini dilakukan untuk membangun model prediksi yang dapat membantu sekolah mengidentifikasi siswa yang berpotensi lulus maupun siswa yang berisiko tidak lulus secara lebih dini. Data yang digunakan dalam penelitian ini berjumlah 289 data siswa, yang dibagi menjadi 202 data latih dan 87 data uji. Atribut prediktor yang digunakan meliputi Nilai Matematika, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Sikap, Kehadiran Persen, Pelanggaran Tata Tertib, Keaktifan Ekstrakurikuler, dan Mengikuti Bimbingan Belajar, sedangkan label yang diprediksi adalah Status Kelulusan. Proses penelitian dilakukan melalui tahapan pengumpulan data, pembersihan data,

DOI: https://doi.org/10.51903/fs4qx719

p-ISSN : 2803-1507 e-ISSN : 2803-1531

pemodelan, dan evaluasi menggunakan *RapidMiner*. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* memperoleh nilai akurasi sebesar 82,76%, sehingga dapat dikatakan cukup baik dalam memprediksi kelulusan siswa. Namun, hasil evaluasi juga menunjukkan bahwa model lebih kuat dalam mengenali kategori Lulus dibandingkan kategori Tidak Lulus. Dengan demikian, model ini lebih tepat digunakan sebagai alat bantu peringatan dini dan pendukung keputusan, bukan sebagai dasar keputusan tunggal dalam menentukan kelulusan siswa. Hasil prediksi dapat membantu sekolah dalam mengidentifikasi siswa yang berisiko tidak lulus sehingga pembinaan, pendampingan, dan dukungan akademik dapat diberikan secara lebih tepat dan efektif.

Kata Kunci: *Naïve Bayes*, *Data Mining*, Pendukung Keputusan, Prediksi Kelulusan, Siswa SMP, *RapidMiner*

1. PENDAHULUAN

Kemajuan teknologi informasi saat ini telah memberikan dampak signifikan di berbagai aspek kehidupan, termasuk pada sektor pendidikan. Pendidikan merupakan proses pembelajaran yang bertujuan untuk mengembangkan potensi peserta didik agar memiliki pengetahuan, keterampilan, dan sikap yang dibutuhkan dalam kehidupan sehari-hari [1]. Sejalan dengan perkembangan zaman, sistem pendidikan terus mengalami perubahan dengan memanfaatkan teknologi untuk meningkatkan kualitas pembelajaran, pengelolaan data, serta proses pengambilan keputusan di sekolah [2]. Salah satu teknologi yang dapat dimanfaatkan dalam bidang pendidikan adalah *data mining*, yaitu proses pengolahan dan analisis data untuk menemukan pola atau informasi yang bernilai [3]. Dalam konteks sekolah, *data mining* dapat digunakan untuk membantu menganalisis data akademik dan nonakademik siswa sehingga menghasilkan informasi yang dapat mendukung pengambilan keputusan [1], [4]. Salah satu penerapannya adalah prediksi kelulusan siswa, yang dapat membantu sekolah dalam mengenali siswa yang berpotensi lulus maupun siswa yang berisiko tidak lulus sejak dini. Kelulusan siswa merupakan salah satu indikator penting dalam menilai keberhasilan proses pembelajaran di suatu institusi pendidikan [5]. Tingkat kelulusan yang tinggi dapat mencerminkan keberhasilan proses pendidikan, sedangkan tingkat kelulusan yang rendah dapat menunjukkan adanya permasalahan dalam pembelajaran, kedisiplinan, kehadiran, atau faktor pendukung lainnya. Namun, pada kenyataannya, data siswa yang dimiliki sekolah sering kali hanya digunakan sebagai arsip administrasi dan belum dimanfaatkan secara optimal sebagai dasar analisis untuk mendukung keputusan sekolah [6]. Akibatnya, pihak sekolah dapat mengalami kesulitan dalam mengidentifikasi siswa yang membutuhkan perhatian khusus sebelum keputusan kelulusan ditetapkan.

SMP Kartika VI Jayapura sebagai salah satu lembaga pendidikan di Papua memiliki data akademik dan nonakademik siswa yang dapat dimanfaatkan untuk mendukung prediksi kelulusan. Data tersebut meliputi nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar. Dalam penelitian ini, data yang digunakan berjumlah 289 data siswa yang dibagi menjadi 202 data latih dan 87 data uji. Atribut yang digunakan meliputi Nilai Matematika, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Sikap, Kehadiran Persen, Pelanggaran Tata Tertib, Keaktifan Ekstrakurikuler, dan Mengikuti Bimbingan Belajar, sedangkan label yang diprediksi adalah Status Kelulusan. Apabila data tersebut diolah dengan metode yang tepat, maka hasilnya dapat menjadi informasi pendukung bagi sekolah dalam mengambil langkah pembinaan dan pendampingan siswa secara lebih terarah. Salah satu metode klasifikasi dalam *data mining* yang dapat digunakan untuk melakukan prediksi adalah algoritma *Naïve Bayes*. Algoritma ini merupakan metode klasifikasi berbasis probabilitas yang bekerja dengan menghitung kemungkinan suatu data termasuk ke dalam kelas tertentu berdasarkan pola data sebelumnya [7]. *Naïve Bayes* memiliki kelebihan karena sederhana, mudah diterapkan, dan mampu menghasilkan performa yang cukup baik dalam proses klasifikasi [8]. Oleh karena itu, algoritma ini sesuai digunakan dalam penelitian prediksi kelulusan siswa berdasarkan data akademik dan nonakademik yang tersedia.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk mengimplementasikan algoritma *Naïve Bayes* sebagai alat bantu pendukung keputusan dalam memprediksi kelulusan siswa SMP di SMP Kartika VI Jayapura. Penelitian ini tidak hanya berfokus pada penerapan algoritma, tetapi juga diarahkan untuk menghasilkan informasi prediktif yang dapat membantu pihak sekolah dalam melakukan identifikasi dini terhadap siswa yang berisiko tidak lulus. Dengan adanya hasil prediksi tersebut, sekolah dapat memberikan pembinaan, pendampingan belajar, dan pengawasan secara lebih tepat sasaran sehingga proses peningkatan kelulusan siswa dapat dilakukan secara lebih efektif.

2. TINJAUAN PUSTAKA

p-ISSN : 2803-1507 e-ISSN : 2803-1531

2.1. Pendidikan

Penelitian terbaru dalam bidang pendidikan menunjukkan bahwa pemanfaatan data siswa memiliki peran penting dalam mendukung proses evaluasi, pemantauan, dan pengambilan keputusan di sekolah. Le Quy et al. [9] menjelaskan bahwa prediksi performa siswa menjadi salah satu fokus dalam *educational data mining* karena dapat membantu institusi pendidikan melakukan identifikasi lebih awal terhadap peserta didik yang membutuhkan perhatian khusus.menunjukkan bahwa perkembangan teknologi pendidikan menghasilkan data akademik dalam jumlah besar, sehingga diperlukan pendekatan analisis berbasis *machine learning* untuk memahami pola belajar, perkembangan akademik, dan faktor-faktor yang memengaruhi keberhasilan siswa. Hal ini menunjukkan bahwa data pendidikan tidak hanya berfungsi sebagai arsip administrasi, tetapi juga dapat diolah menjadi informasi yang mendukung peningkatan mutu pendidikan. Selain itu, Al-Giffary et al. [10] menjelaskan bahwa sistem pengelolaan data pendidikan dapat membantu mengungkap atribut-atribut penting yang berpengaruh terhadap performa akademik siswa. yang menunjukkan bahwa data aktivitas belajar dapat digunakan untuk memprediksi performa siswa serta membantu pendidik mengidentifikasi siswa yang mengalami kesulitan belajar. Berdasarkan penelitian-penelitian tersebut, penelitian ini memanfaatkan data akademik dan nonakademik siswa SMP Kartika VI Jayapura, seperti nilai mata pelajaran, kehadiran, sikap, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar, untuk membangun model prediksi kelulusan menggunakan algoritma *Naïve Bayes*. Dengan demikian, penelitian ini sejalan dengan arah penelitian pendidikan saat ini yang menekankan pemanfaatan data siswa sebagai dasar pendukung keputusan sekolah secara lebih objektif dan terarah.

2.2. *Knowledge Discovery in Database (KDD)*

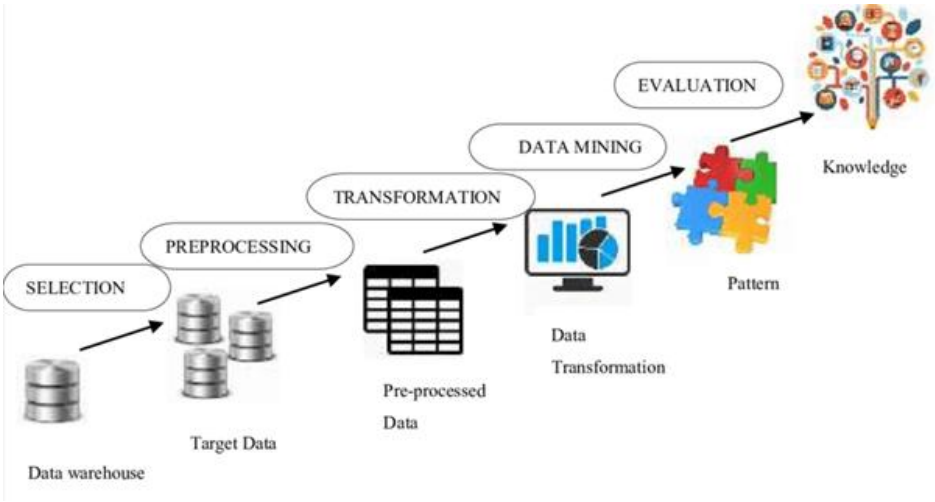
Penelitian yang dilakukan oleh Al-Giffary dan Martanto. [10] membahas penerapan *Knowledge Discovery in Database (KDD)* dalam proses klasifikasi kelulusan siswa menggunakan metode *Decision Tree.* Penelitian tersebut menggunakan tahapan KDD untuk mengolah data siswa menjadi informasi prediktif mengenai status kelulusan. Tahapan yang digunakan meliputi pemilihan data, pembersihan data, transformasi data, penerapan *algoritma data mining,* dan evaluasi hasil. Penelitian ini relevan dengan penelitian yang dilakukan di SMP Kartika VI Jayapura karena sama-sama menggunakan data siswa sebagai dasar untuk menghasilkan informasi yang dapat mendukung prediksi kelulusan. Selain itu, Nugroho et al. [11] menerapkan KDD dalam penelitian klasterisasi data menggunakan *algoritma K-Means.* Penelitian tersebut menunjukkan bahwa KDD dapat digunakan sebagai alur sistematis dalam proses *data mining,* mulai dari pengolahan data awal sampai menghasilkan pola yang dapat dianalisis. Penelitian Fitriana et al [12] juga menggunakan KDD sebagai metode analisis data dengan tahapan yang terstruktur sebelum proses pemodelan dilakukan. Berdasarkan beberapa penelitian tersebut, pendekatan KDD dalam penelitian ini digunakan sebagai alur kerja untuk mengolah data akademik dan nonakademik siswa SMP Kartika VI Jayapura melalui tahapan pengumpulan data, seleksi atribut, pembersihan data, transformasi, pemodelan menggunakan *Naïve Bayes,* dan evaluasi hasil. Dengan demikian, KDD tidak hanya digunakan sebagai konsep, tetapi sebagai proses sistematis untuk menghasilkan informasi prediksi kelulusan siswa yang lebih terarah. Tahapan dalam KDD meliputi:

1. *Selection* (Pemilihan Data)
   Tahap ini merupakan proses memilih data yang relevan dari sumber data, seperti data *warehouse,* agar analisis lebih terfokus.
2. *Preprocessing* (Pembersihan Data)
   Pada tahap ini dilakukan pembersihan data dari kesalahan, data ganda, atau data yang tidak lengkap sehingga kualitas data menjadi lebih baik.
3. *Transformation* (Transformasi Data)
   Data yang telah dibersihkan kemudian diubah ke dalam format yang sesuai untuk proses analisis, misalnya melalui normalisasi atau pengelompokan.
4. *Data Mining*
   Tahap inti dalam KDD, yaitu proses penerapan algoritma atau teknik tertentu untuk menemukan pola atau informasi penting dari data.
5. *Evaluation* (Evaluasi)
   Tahap ini bertujuan untuk menilai hasil pola yang ditemukan apakah memiliki nilai guna dan relevan untuk digunakan.
6. *Knowledge* (Pengetahuan)
   Tahap ini merupakan hasil akhir dari seluruh proses KDD, yaitu berupa informasi atau pengetahuan baru yang diperoleh dari data. Pengetahuan ini dapat digunakan sebagai dasar dalam pengambilan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

keputusan, misalnya dalam menentukan strategi pembelajaran atau mengidentifikasi siswa yang berpotensi tidak lulus.

Melalui tahapan KDD tersebut, data yang awalnya tidak terstruktur dapat diolah secara sistematis hingga menghasilkan informasi yang bernilai. Informasi ini kemudian dapat dimanfaatkan dalam berbagai bidang, termasuk pendidikan, khususnya dalam membantu memprediksi kelulusan siswa.



Gambar 1. Alur penelitian proses KDD

Tahapan proses *Knowledge Discovery in Database* yang digunakan sebagai dasar alur penelitian ditampilkan pada Gambar 1. Gambar tersebut menunjukkan proses pengolahan data mulai dari pemilihan data, pembersihan, transformasi, proses data *mining*, evaluasi, hingga menghasilkan pengetahuan.

2.3. *Data Mining*

Penelitian terbaru menunjukkan bahwa data *mining* banyak digunakan dalam bidang pendidikan untuk mengolah data akademik menjadi informasi yang mendukung pengambilan keputusan. Rachardian dan Sediyono [13] menerapkan metode data mining untuk memprediksi kelulusan tepat waktu mahasiswa sebagai alat pemantauan program studi. Penelitian tersebut menunjukkan bahwa data historis pendidikan dapat dianalisis untuk menghasilkan informasi prediktif yang membantu institusi dalam melakukan evaluasi dan pengawasan akademik. Selain itu, Rachardian dan Sediyono [13] menjelaskan bahwa penerapan *data mining* dalam bidang pendidikan dapat menemukan pola dan pengetahuan tersembunyi dari data akademik, sehingga hasilnya dapat membantu manajemen pendidikan dalam meningkatkan sistem pembelajaran dan pengambilan keputusan.

Penelitian lain yang dilakukan oleh Wijaya et al. [14] menerapkan data mining untuk memprediksi kelulusan siswa sekolah dasar menggunakan algoritma *Naïve Bayes Classifier*. Penelitian tersebut menunjukkan bahwa data siswa dapat diolah menggunakan teknik data mining untuk menghasilkan prediksi kelulusan, sekaligus menegaskan bahwa evaluasi model tidak hanya dapat dilihat dari akurasi, tetapi juga perlu memperhatikan nilai *precision* dan *recall* pada setiap kelas prediksi. Lastari dan Jasmir [15] juga menunjukkan bahwa data mining dapat digunakan untuk memprediksi prestasi siswa SMA dengan memanfaatkan data pendidikan sebagai dasar analisis. Penelitian tersebut memperlihatkan bahwa penerapan data mining pada lingkungan sekolah dapat membantu menghasilkan informasi prediktif yang berguna dalam mendukung evaluasi akademik. Sementara itu, Imanuel et al. [16] menerapkan algoritma *Naïve Bayes* untuk memprediksi tingkat kelulusan mahasiswa dan menunjukkan bahwa data historis pendidikan dapat digunakan sebagai dasar untuk membangun model prediksi kelulusan.

Berdasarkan beberapa penelitian tersebut, data mining dalam penelitian ini digunakan sebagai pendekatan untuk mengolah data siswa SMP Kartika VI Jayapura menjadi informasi yang prediktif. Data yang dianalisis mencakup atribut akademik dan nonakademik, seperti nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar. Dengan demikian, penerapan data mining pada penelitian ini relevan karena bertujuan untuk menemukan pola dari data siswa dan menghasilkan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

informasi prediksi kelulusan yang dapat membantu sekolah dalam melakukan pembinaan serta pengambilan keputusan secara lebih tepat.

2.4. *Algoritma Naïve Bayes*

Perkembangan penggunaan algoritma *Naïve Bayes* dalam bidang pendidikan menunjukkan bahwa metode ini banyak diterapkan untuk membantu proses klasifikasi dan prediksi berbasis data akademik. Pusparini et al. [17] menerapkan algoritma *Naïve Bayes* untuk memprediksi waktu kelulusan mahasiswa pada Program Studi Pendidikan Teknik Informatika dan Komputer. Penelitian tersebut menunjukkan bahwa *Naïve Bayes* dapat digunakan untuk mengolah data pendidikan menjadi informasi prediktif yang membantu institusi memperkirakan kemungkinan kelulusan peserta didik. Hal ini menunjukkan bahwa *Naïve Bayes* relevan digunakan dalam lingkungan pendidikan karena mampu mengklasifikasikan data berdasarkan atribut-atribut yang berkaitan dengan hasil belajar atau kelulusan. Pada konteks siswa, Hudzaifah et al. [18] menerapkan algoritma *Naïve Bayes* untuk memprediksi tingkat kelulusan siswa pada sertifikasi *Mikrotik Certified Network Associate*. Penelitian tersebut menunjukkan bahwa *Naïve Bayes* dapat digunakan untuk memprediksi kelulusan siswa berdasarkan faktor-faktor yang berkaitan dengan proses pembelajaran dan kesiapan peserta didik. Berdasarkan penelitian tersebut, penggunaan algoritma *Naïve Bayes* dalam penelitian ini dinilai sesuai karena penelitian ini bertujuan untuk memprediksi status kelulusan siswa SMP Kartika VI Jayapura ke dalam kategori Lulus dan Tidak Lulus. Atribut yang digunakan meliputi nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar, sehingga *Naïve Bayes* digunakan sebagai metode klasifikasi untuk menghasilkan informasi prediksi kelulusan yang dapat mendukung pengambilan keputusan sekolah.

2.5. *RapidMiner*

Penggunaan *RapidMiner* dalam penelitian pendidikan banyak dimanfaatkan untuk membantu proses pemodelan data mining, mulai dari pembacaan dataset, penentuan atribut, pembentukan model, hingga evaluasi hasil klasifikasi. Fauziah et al. [19] menggunakan *RapidMiner* untuk memprediksi kelulusan mahasiswa dengan algoritma *Naïve Bayes*. Penelitian tersebut menunjukkan bahwa *RapidMiner* dapat digunakan sebagai alat bantu dalam menjalankan proses klasifikasi dan menampilkan hasil evaluasi model, sehingga memudahkan peneliti dalam mengetahui tingkat akurasi prediksi kelulusan yang dihasilkan. Penelitian Hakim et al. [5] juga menerapkan *RapidMiner* dalam proses prediksi prestasi siswa menggunakan metode *Naïve Bayes*. Dalam penelitian tersebut, *RapidMiner* digunakan untuk mengimpor data nilai siswa, menentukan label, membangun model klasifikasi, dan mengevaluasi hasil prediksi. Berdasarkan penelitian-penelitian tersebut, penggunaan *RapidMiner* dalam penelitian ini dinilai relevan karena penelitian ini juga melakukan tahapan pengolahan data siswa SMP Kartika VI Jayapura, seleksi atribut, penerapan algoritma *Naïve Bayes*, serta evaluasi model melalui nilai *accuracy, precision, recall*, dan *confusion matrix*. Dengan demikian, *RapidMiner* berperan sebagai perangkat pendukung dalam membangun model prediksi kelulusan siswa secara lebih sistematis.

2.6. Prediksi Kelulusan Siswa

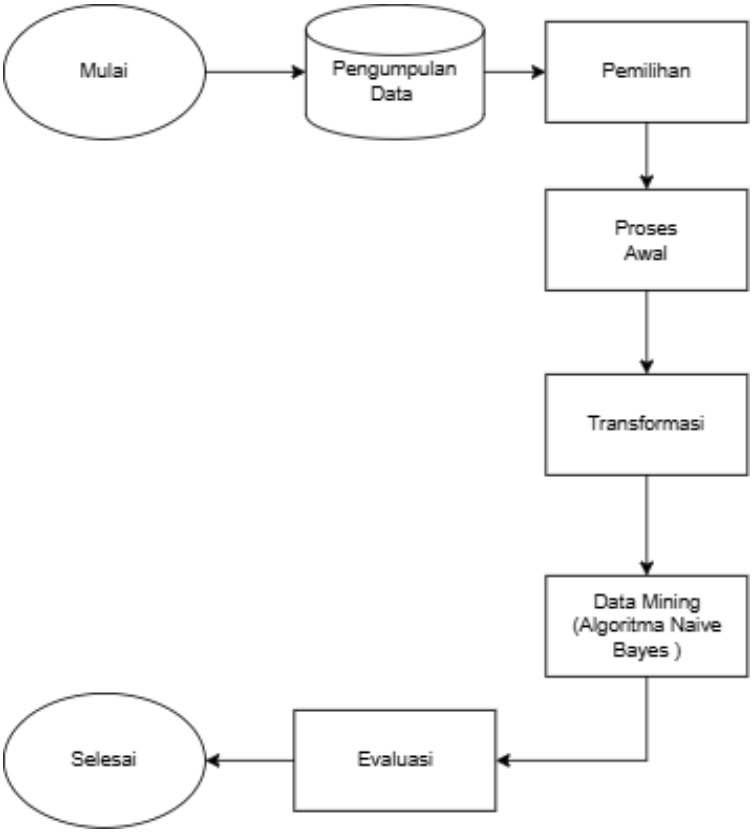
Penelitian mengenai prediksi kelulusan siswa menunjukkan bahwa data pendidikan dapat dimanfaatkan untuk membantu sekolah dalam mengidentifikasi peserta didik yang berpotensi lulus maupun berisiko tidak lulus. Peneliti Al-Giffary et al. [10] melakukan penelitian klasifikasi kelulusan siswa menggunakan metode *Decision Tree* pada SMA Islam Alazhar 5 Cirebon. Penelitian tersebut menggunakan data siswa sebagai dasar klasifikasi dan menghasilkan akurasi sebesar 91,43%. Hasil tersebut menunjukkan bahwa prediksi kelulusan siswa dapat digunakan sebagai alat bantu dalam proses evaluasi pendidikan karena mampu memberikan gambaran awal mengenai status kelulusan peserta didik berdasarkan data yang tersedia. Peneliti Riyanto et al. [20], Latifah et al. [21] juga menerapkan data mining dengan metode *Naïve Bayes* untuk mengklasifikasi hasil belajar siswa/i. Penelitian tersebut menunjukkan bahwa data hasil belajar dapat diolah menggunakan metode klasifikasi untuk menghasilkan informasi prediktif yang membantu proses evaluasi pendidikan. Temuan tersebut relevan dengan penelitian ini karena prediksi kelulusan siswa SMP Kartika VI Jayapura juga menggunakan data akademik dan nonakademik, seperti nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar. Dengan demikian, prediksi kelulusan dalam penelitian ini diarahkan sebagai alat bantu pendukung keputusan bagi sekolah dalam melakukan pembinaan dan pendampingan siswa secara lebih tepat sasaran.

2.7. *Research Gap*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Berdasarkan penelitian terdahulu, algoritma *Naïve Bayes* telah banyak digunakan dalam bidang pendidikan untuk keperluan prediksi dan klasifikasi. Penelitian yang dilakukan Ramadhani et al. [22] membandingkan algoritma Naïve Bayes, K-NN, ID3, dan SVM dalam prediksi kelulusan siswa pada domain pendidikan kejuruan. Penelitian tersebut menunjukkan bahwa data nilai semester dan ujian sekolah dapat digunakan sebagai atribut untuk membangun model prediksi kelulusan siswa. Peneliti Baridah et al. [23] membandingkan model prediksi kelulusan mahasiswa menggunakan algoritma Decision Tree dan Naïve Bayes, serta menunjukkan bahwa algoritma klasifikasi dapat digunakan untuk menghasilkan informasi prediktif pada data pendidikan. Peneliti Ramadhani et al. [22] juga menerapkan *Naïve Bayes* untuk memprediksi hasil belajar siswa pada bidang kejuruan komputer dan jaringan. Sementara itu, Ramadhani et al. [22], [24], dan [25] memprediksi kelulusan siswa berdasarkan data demografis dan akademik, serta menunjukkan bahwa atribut seperti usia, riwayat kegagalan, dan kehadiran berperan dalam proses prediksi.

Meskipun demikian, masih terdapat celah penelitian yang perlu dikaji lebih lanjut. Penelitian sebelumnya umumnya berfokus pada jenjang SMK, SD, atau perguruan tinggi, sedangkan penelitian yang secara khusus membahas prediksi kelulusan siswa SMP masih relatif terbatas. Selain itu, sebagian besar penelitian terdahulu lebih banyak menekankan variabel akademik seperti nilai ujian, transkrip nilai, atau data demografis. Penelitian ini menggunakan kombinasi atribut akademik dan nonakademik, yaitu nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar. Oleh karena itu, penelitian berjudul "Pendukung Keputusan Naïve Bayes untuk Prediksi Kelulusan Siswa SMP: Studi SMP Kartika VI Jayapura" dilakukan untuk mengisi kesenjangan dari sisi jenjang pendidikan, kelengkapan atribut, dan konteks objek penelitian.



Gambar 2. Alur metode penelitian naïve bayes

3. METODOLOGI PENELITIAN

Metode penelitian merupakan tahapan yang digunakan sebagai pedoman agar proses penelitian dapat dilakukan secara sistematis, terarah, dan sesuai dengan tujuan yang ingin dicapai. Pada penelitian ini digunakan metode *data mining* dengan pendekatan klasifikasi menggunakan algoritma *Naïve Bayes* untuk memprediksi kelulusan siswa. Data penelitian diperoleh dari arsip akademik dan nonakademik siswa SMP Kartika VI Jayapura pada Tahun Ajaran 2024/2025. Data yang digunakan dalam penelitian ini berjumlah 289 data siswa, yang kemudian dibagi menjadi 202 data latih dan 87 data uji. Data tersebut terdiri dari dua kelas

p-ISSN : 2803-1507 e-ISSN : 2803-1531

target, yaitu Lulus dan Tidak Lulus. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi kelulusan siswa. Sebelum proses pemodelan dilakukan, data diperiksa terlebih dahulu untuk memastikan tidak terdapat nilai kosong, data duplikat, maupun format atribut yang tidak sesuai. Apabila ditemukan nilai kosong, penanganan dilakukan menggunakan operator *Replace Missing Values* pada *RapidMiner*. Pembagian data dilakukan dengan komposisi 70% data latih dan 30% data uji, yaitu 202 data latih dan 87 data uji. Adapun tahapan penelitian ini meliputi pengumpulan data, pemilihan data, proses awal atau *preprocessing*, transformasi data, penerapan algoritma *Naïve Bayes*, serta evaluasi hasil model. Evaluasi performa model dilakukan menggunakan nilai *accuracy*, *precision*, *recall*, dan *confusion matrix*. Melalui tahapan tersebut, penelitian ini diharapkan dapat menghasilkan model prediksi yang akurat serta dapat digunakan sebagai alat bantu pendukung keputusan bagi pihak sekolah dalam mengidentifikasi kemungkinan kelulusan siswa secara lebih dini.

Alur metode penelitian yang diterapkan dalam penelitian ini ditunjukkan pada Gambar 2. Alur tersebut menggambarkan tahapan mulai dari pengumpulan data hingga evaluasi hasil model *Naïve Bayes*.

1. Pengumpulan Data
   Pada tahap pengumpulan data, peneliti mengumpulkan data dari sumber yang relevan, seperti dari pihak sekolah. Data yang digunakan dalam penelitian ini dapat berupa nilai rapor siswa, tingkat kehadiran, sikap atau perilaku, nilai ujian, data ekstrakurikuler, serta status kelulusan siswa. Tahap ini bertujuan untuk memperoleh data yang lengkap dan sesuai agar dapat dijadikan sebagai bahan analisis dalam proses penelitian.
2. Pemilihan
   Pada tahap pemilihan, dilakukan seleksi terhadap data yang benar-benar dibutuhkan dalam penelitian. Tidak semua data yang telah terkumpul digunakan, melainkan hanya data yang relevan dengan tujuan penelitian. Contohnya, peneliti memilih atribut seperti nilai akademik, absensi, dan sikap, serta menghapus data yang tidak memiliki hubungan langsung dengan prediksi kelulusan. Tahap ini bertujuan agar proses analisis menjadi lebih fokus, efektif, dan efisien.
3. Proses Awal
   Tahap proses awal merupakan tahap pra-pemrosesan data atau *preprocessing*. Pada tahap ini, data yang telah dipilih akan dibersihkan dan diperiksa kembali agar siap untuk diolah. Kegiatan yang dilakukan biasanya meliputi pembersihan data yang kosong, perbaikan data yang tidak konsisten, penghapusan data ganda, serta penyamaan format data. Tahap ini sangat penting karena kualitas data yang baik akan berpengaruh pada keakuratan hasil analisis.
4. Transformasi
   Pada tahap transformasi, data yang sudah bersih diubah ke dalam bentuk yang sesuai untuk proses data mining. Bentuk transformasi dapat berupa pengelompokan nilai ke dalam kategori tertentu, pengkodean data kategori menjadi bentuk yang dapat diproses oleh sistem, serta penyusunan data ke dalam bentuk tabel dataset. Tujuan dari tahap ini adalah agar data dapat diproses dengan baik menggunakan algoritma *Naive Bayes*.
5. *Data Mining* (Algoritma *Naive Bayes*)
   Tahap data mining merupakan tahap inti dalam penelitian ini, yaitu proses pengolahan data menggunakan algoritma Naive Bayes untuk melakukan klasifikasi prediksi kelulusan siswa. Pada tahap ini, data siswa diklasifikasikan ke dalam dua kelas, yaitu Lulus dan Tidak Lulus, berdasarkan sejumlah atribut yang digunakan sebagai variabel input. Atribut yang digunakan dalam penelitian ini terdiri atas Nilai Matematika, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Sikap, Kehadiran Persen, Pelanggaran Tata Tertib, Keaktifan Ekstrakurikuler, dan Mengikuti Bimbingan Belajar. Dengan demikian, terdapat sepuluh atribut yang digunakan untuk menentukan kelas kelulusan setiap siswa.
   Dalam metode Naive Bayes, data seorang siswa dinyatakan sebagai $X = (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10})$, dengan setiap X_i merepresentasikan atribut yang digunakan dalam proses klasifikasi. Perhitungan probabilitas kelas dilakukan berdasarkan Teorema Bayes dengan rumus (1):

$$P(C|X) = [P(X|C) \times P(C)] / P(X) \qquad (1)$$

   dengan C merupakan kelas yang akan diprediksi, yaitu Lulus atau Tidak Lulus; P(C|X) merupakan probabilitas suatu data X termasuk ke dalam kelas C setelah mempertimbangkan seluruh atribut; P(X|C) merupakan probabilitas kemunculan data X apabila diketahui berada pada kelas C; P(C) merupakan probabilitas awal atau prior dari kelas C; dan P(X) merupakan probabilitas kemunculan data X. Karena algoritma Naive Bayes mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya ketika berada dalam suatu kelas tertentu, maka probabilitas P(X|C) dapat diuraikan menjadi

p-ISSN : 2803-1507 e-ISSN : 2803-1531

perkalian probabilitas masing-masing atribut terhadap kelas. Penyederhanaan tersebut dinyatakan dengan rumus (2):

$$P(X|C) = P(X_1|C) \times P(X_2|C) \times P(X_3|C) \times P(X_4|C) \times P(X_5|C) \times P(X_6|C) \times P(X_7|C) \times P(X_8|C) \times P(X_9|C) \times P(X_{10}|C) \quad (2)$$

Bentuk tersebut dapat dituliskan secara lebih ringkas menggunakan notasi perkalian sebagai rumus (3):

$$P(X|C) = \prod_{i=1}^{10} P(X_i|C) \quad (3)$$

Selanjutnya, nilai probabilitas setiap kelas dibandingkan berdasarkan hasil perhitungan pada masing-masing atribut. Kelas dengan probabilitas posterior terbesar ditetapkan sebagai hasil prediksi kelulusan siswa. Dengan demikian, proses klasifikasi tidak hanya mempertimbangkan satu atribut, tetapi menggabungkan informasi dari seluruh atribut yang tersedia.

Adapun kesepuluh atribut yang digunakan dalam proses klasifikasi adalah sebagai berikut:

X_1 = Nilai Matematika
X_2 = Nilai Bahasa Indonesia
X_3 = Nilai Bahasa Inggris
X_4 = Nilai IPA
X_5 = Nilai IPS
X_6 = Sikap
X_7 = Kehadiran Persen
X_8 = Pelanggaran Tata Tertib
X_9 = Keaktifan Ekstrakurikuler
X_{10} = Mengikuti Bimbingan Belajar

$$P(C|X) = P(C) \times P(X_1|C) \times P(X_2|C) \times P(X_3|C) \times P(X_4|C) \times P(X_5|C) \times P(X_6|C) \times P(X_7|C) \times P(X_8|C) \times P(X_9|C) \times P(X_{10}|C) \quad (4)$$

Selanjutnya, untuk menentukan hasil klasifikasi, dilakukan perbandingan nilai probabilitas antara kelas Lulus dan Tidak Lulus. Kelas yang memiliki nilai probabilitas terbesar akan menjadi hasil prediksi akhir. Proses penentuan kelas tersebut dirumuskan sebagai berikut:

$$\text{Prediksi}(X) = \arg\max_{C \in \{\text{Lulus, Tidak Lulus}\}} P(C) \prod_{i=1}^{10} P(X_i|C) \quad (5)$$

Dengan demikian, algoritma *Naive Bayes* digunakan untuk mempelajari pola data siswa berdasarkan atribut-atribut yang telah ditentukan, kemudian menghitung probabilitas masing-masing kelas untuk memperoleh prediksi kelulusan siswa secara tepat.

6. Evaluasi

Setelah model klasifikasi terbentuk, tahap selanjutnya adalah evaluasi untuk mengetahui tingkat kinerja model tersebut. Evaluasi dilakukan untuk melihat apakah hasil prediksi yang dihasilkan sudah baik atau belum. Pengukuran evaluasi biasanya menggunakan akurasi*, precision, recall*, dan *confusion matrix*. Tujuan dari tahap ini adalah untuk mengetahui seberapa tepat algoritma *Naive Bayes* dalam memprediksi kelulusan siswa sehingga dapat dinilai kelayakannya dalam penelitian.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pada penelitian ini, data dikumpulkan dari arsip resmi SMP Kartika VI Jayapura yang memuat 289 data siswa beserta informasi akademik dan nonakademik yang berkaitan dengan kelulusan. Untuk menjaga fokus analisis, atribut identitas seperti No, NIS, dan Nama_Siswa tidak digunakan sebagai dasar prediksi, karena penelitian ini lebih menitikberatkan pada faktor-faktor yang memengaruhi kelulusan siswa. Label yang diprediksi adalah Status_Kelulusan dengan dua kategori, yaitu Lulus dan Tidak Lulus, yang merepresentasikan hasil akhir kelulusan siswa berdasarkan data sekolah. Adapun atribut yang digunakan sebagai prediktor meliputi Nilai_Matematika, Nilai_Bahasa_Indonesia, Nilai_Bahasa_Inggris, Nilai_IPA, Nilai_IPS, Sikap, Kehadiran_Persen, Pelanggaran_Tata_Tertib, Keaktifan_Ekstrakurikuler, dan Mengikuti_Bimbingan_Belajar.

Komposisi atribut tersebut dinilai memadai untuk merepresentasikan kondisi akademik, kedisiplinan, kehadiran, serta aktivitas pendukung siswa dalam memprediksi kelulusan. Dari total 289 data siswa, dataset dibagi menjadi 202 data latih dan 87 data uji. Data latih digunakan untuk membangun model klasifikasi menggunakan algoritma *Naive Bayes*, sedangkan data uji digunakan untuk mengukur performa model dalam

p-ISSN : 2803-1507 e-ISSN : 2803-1531

memprediksi Status_Kelulusan siswa. Tabel 1 memperlihatkan cuplikan data awal siswa SMP Kartika VI Jayapura beserta atribut-atribut yang digunakan dalam penelitian ini.

Tabel 1. Tabel Data Siswa SMP VI Kartika Jayapura

No	NIS	Nama_Siswa	JK	Usia	Kelas	Nilai_MTK	...	K. Ekstrakurikuler	M.B.B	Tunggakan_SPP	Pelanggaran_T.T	Status S
1	2301001	Alfin Wenda	P	14	IX-A	51	...	Cukup Aktif	Ya	Ya	0	Lulus
2	2301002	Elsa Mofu	P	15	IX-D	77	...	Aktif	Ya	Tidak	1	Lulus
3	2301003	Intan Ayomi	L	15	IX-B	69,7	...	Kurang Aktif	Ya	Tidak	1	Lulus
4	2301004	Celine Ayomi	L	16	IX-B	66,4	...	Aktif	Ya	Tidak	5	Lulus
5	2301005	Rafi Kareth	P	15	IX-F	85,3	...	Cukup Aktif	Tidak	Tidak	2	Lulus
6	2301006	Dimas Isir	P	15	IX-C	73	...	Cukup Aktif	Tidak	Tidak	1	Lulus
7	2301007	Haris Isir	L	16	IX-B	69,8	...	Cukup Aktif	Tidak	Tidak	6	Tidak Lulus
8	2301008	Patricia Yoteni	L	15	IX-E	73,6	...	Cukup Aktif	Tidak	Ya	4	Lulus
9	2301009	Ardi Womsiwor	P	15	IX-C	68,6	...	Cukup Aktif	Tidak	Tidak	0	Lulus
10	2301010	Yohana Kareth	L	14	IX-B	91	...	Kurang Aktif	Tidak	Ya	4	Lulus
...	...	...	...	...	...	...	...	...	...	...	..	..

4.2 *Pembersihan Data*

Tahap pembersihan data dilakukan sebelum proses klasifikasi menggunakan algoritma *Naive Bayes* dengan tujuan untuk memastikan bahwa data yang digunakan berada dalam kondisi baik, konsisten, dan layak untuk diproses lebih lanjut. Dataset awal yang digunakan memuat data siswa SMP Kartika VI Jayapura dengan sejumlah atribut akademik dan nonakademik, seperti jenis kelamin, usia, kelas, nilai mata pelajaran, rata-rata rapor, kehadiran, jumlah ketidakhadiran, sikap, status ekonomi keluarga, keaktifan ekstrakurikuler, keikutsertaan bimbingan belajar, tunggakan SPP, pelanggaran tata tertib, NIS, serta status kelulusan. Pada tahap ini dilakukan pemeriksaan terhadap kelengkapan data untuk mengetahui ada atau tidaknya nilai kosong, nilai tidak valid, maupun inkonsistensi penulisan pada setiap atribut. Data yang tidak valid, seperti nilai yang bukan berupa angka pada atribut numerik, diperbaiki atau diganti menggunakan pendekatan *Replace Missing Values* di *RapidMiner* agar tidak mengganggu proses pembentukan model. Selain itu, dilakukan penyesuaian format data agar setiap atribut memiliki tipe data yang sesuai, yaitu atribut nilai, kehadiran, jumlah alfa, jumlah izin, jumlah sakit, dan pelanggaran tata tertib dipastikan terbaca sebagai data numerik, sedangkan atribut seperti jenis kelamin, kelas, sikap, status ekonomi keluarga, keaktifan ekstrakurikuler, mengikuti bimbingan belajar, tunggakan SPP, dan status kelulusan diperlakukan sebagai data kategorikal.

Selanjutnya, dilakukan seleksi atribut untuk menghilangkan atribut yang tidak berkontribusi langsung terhadap proses klasifikasi. Pada penelitian ini, atribut identitas seperti NIS tidak digunakan sebagai dasar pengambilan keputusan model karena hanya berfungsi sebagai penanda data siswa. Setelah proses pembersihan data dilakukan, tahap berikutnya adalah seleksi atribut. Atribut identitas seperti NIS dan Nama_Siswa tidak digunakan karena hanya berfungsi sebagai penanda data siswa. Atribut tambahan seperti Jenis_Kelamin, Usia, Kelas, Rata_Rata_Rapor, Jumlah_Alfa, Jumlah_Izin, Jumlah_Sakit, Status_Ekonomi_Keluarga, dan Tunggakan_SPP juga tidak digunakan dalam pemodelan akhir agar atribut penelitian tetap konsisten dengan rancangan model. Dengan demikian, atribut final yang digunakan sebagai prediktor adalah Nilai_Matematika, Nilai_Bahasa_Indonesia, Nilai_Bahasa_Inggris, Nilai_IPA, Nilai_IPS, Sikap, Kehadiran_Persen, Pelanggaran_Tata_Tertib, Keaktifan_Ekstrakurikuler, dan Mengikuti_Bimbingan_Belajar, sedangkan Status_Kelulusan digunakan sebagai label. Selain itu, data juga diperiksa agar tidak terdapat duplikasi baris maupun kesalahan input yang dapat memengaruhi hasil klasifikasi. Hasil dari tahap pembersihan data adalah dataset yang lebih rapi, konsisten, dan siap digunakan pada tahap berikutnya, yaitu pembentukan model klasifikasi *Naive Bayes*. Dengan adanya pembersihan data, kualitas data menjadi lebih baik sehingga hasil prediksi yang diperoleh dapat lebih akurat dan dapat

p-ISSN : 2803-1507 e-ISSN : 2803-1531

dipertanggungjawabkan. Hasil data yang telah melalui proses pembersihan dan penyesuaian format ditampilkan pada Gambar 3. Data bersih ini kemudian digunakan sebagai dasar dalam proses pemodelan klasifikasi.

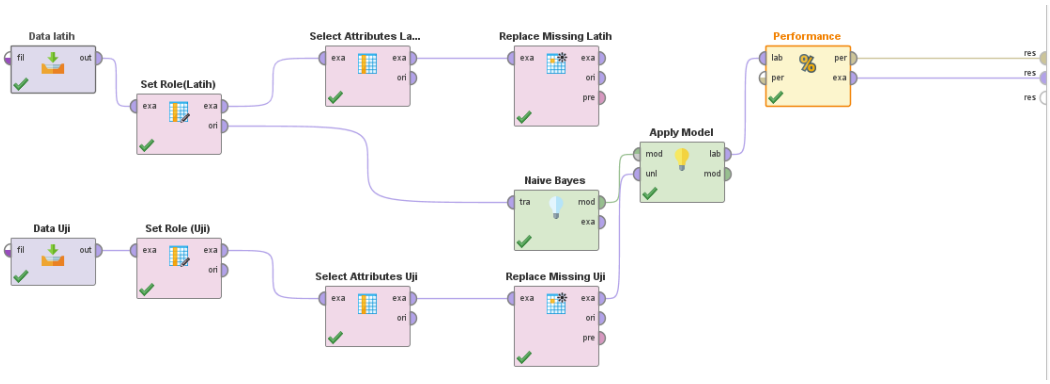
Jenis_Kel	Usia	Kelas	Nilai_M	Nilai_Ba	Nilai_Ba	Nilai_IP	Nilai_IP	Rata_Ra	Kehadir	Jumlah_	Jumlah_	Jumlah_	Sikap	Status_I	Keaktifa	Mengik	Tunggal	Pelanggaran	NIS	Status_Kelulusan
L	14	IX-F	69,2	68,5	80,0	66,6	75,5	72,0	88,5	4	2	2	C	Rendah	Cukup A	Tidak	Tidak	4	2301203,0	Lulus
L	14	IX-B	73,1	79,5	80,5	86,4	78,5	79,6	85,4	6	3	2	C	Menenç	Kurang	Ya	Tidak	5	2301204,0	Tidak Lulus
P	15	IX-E	66,5	67,6	67,8	73,1	73,2	69,6	98,6	3	0	0	B	Menenç	Aktif	Tidak	Tidak	0	2301205,0	Lulus
L	14	IX-A	60,4	66,9	59,0	65,6	57,6	61,9	93,3	2	1	1	B	Tinggi	Aktif	Tidak	Tidak	1	2301206,0	Lulus
L	15	IX-C	75,3	80,0	67,6	78,3	68,2	73,9	99,3	0	0	0	B	Rendah	Aktif	Tidak	Ya	3	2301207,0	Lulus
P	15	IX-D	66,4	82,2	75,9	75,3	83,3	76,6	92,5	3	2	1	B	Menenç	Cukup A	Ya	Tidak	2	2301208,0	Lulus
L	16	IX-C	63,7	64,0	60,8	65,1	72,1	65,1	77,8	7	3	3	C	Tinggi	Kurang	Tidak	Tidak	8	2301209,0	Tidak Lulus
L	15	IX-D	93,7	91,4	95,9	88,6	88,3	91,6	100,0	0	0	1	B	Tinggi	Kurang	Ya	Tidak	1	2301210,0	Lulus
L	15	IX-F	70,6	79,9	78,8	64,8	63,6	71,5	100,0	2	0	0	B	Menenç	Aktif	Ya	Tidak	0	2301211,0	Lulus
P	16	IX-C	80,8	80,4	91,5	90,8	94,3	87,6	100,0	1	0	0	B	Tinggi	Cukup A	Tidak	Tidak	0	2301212,0	Lulus
L	14	IX-A	51,9	60,4	45,0	64,3	51,2	54,6	90,3	3	1	1	B	Menenç	Kurang	Tidak	Tidak	6	2301213,0	Tidak Lulus
P	15	IX-A	76,9	77,1	68,8	75,7	79,1	75,5	95,3	1	1	1	B	Menenç	Kurang	Ya	Tidak	2	2301214,0	Lulus
L	15	IX-B	78,2	74,4	65,7	70,0	62,8	70,2	100,0	0	0	1	A	Menenç	Aktif	Ya	Tidak	0	2301215,0	Lulus
L	15	IX-A	51,8	57,4	48,3	48,9	51,4	51,6	91,6	3	1	2	B	Menenç	Kurang	Tidak	Tidak	3	2301216,0	Tidak Lulus
L	14	IX-D	48,4	52,3	47,1	56,1	45,0	49,8	89,0	3	2	2	C	Menenç	Aktif	Ya	Ya	4	2301217,0	Tidak Lulus

Gambar 3. Data Set Bersih

4.3 *Pemodelan Data*

Pada penelitian ini, proses klasifikasi di *RapidMiner* diawali dengan memisahkan dataset menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membentuk model klasifikasi, sedangkan data uji digunakan untuk menguji kemampuan model dalam memprediksi status kelulusan siswa. Pada alur di gambar, data latih terlebih dahulu dimasukkan melalui operator *Read CSV*, kemudian dilanjutkan ke *Set Role* (Latih) untuk menentukan peran setiap atribut. Pada tahap ini, atribut Status_Kelulusan ditetapkan sebagai label karena menjadi target yang akan diprediksi, sedangkan atribut seperti NIS dapat dijadikan sebagai id atau tidak digunakan langsung dalam proses pembelajaran model. Setelah itu, dilakukan *Select Attributes* Latih untuk memilih atribut-atribut yang relevan terhadap prediksi kelulusan dan menghapus atribut yang tidak diperlukan, seperti atribut identitas atau atribut yang tidak berpengaruh langsung terhadap hasil klasifikasi. Tahap berikutnya adalah *Replace Missing* Latih, yang berfungsi untuk menangani data kosong, data tidak valid, atau data yang tidak konsisten agar data latih menjadi lebih bersih dan siap digunakan pada proses pembentukan model.

Setelah data latih selesai diproses, data tersebut dimasukkan ke operator *Naive Bayes* untuk membentuk model klasifikasi. Algoritma *Naive Bayes* bekerja dengan menghitung peluang setiap kelas berdasarkan data pelatihan yang tersedia. Dalam penelitian ini, model belajar dari pola hubungan antara atribut siswa, seperti nilai akademik, kehadiran, sikap, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar, terhadap Status_Kelulusan. Sementara itu, data uji juga diproses dengan tahapan yang serupa agar susunan atributnya sama dengan data latih. Data uji dibaca melalui *Read CSV*, kemudian peran atribut diatur pada *Set Role* (Uji), selanjutnya atribut diseleksi kembali pada *Select Attributes* Uji, dan dibersihkan pada *Replace Missing* Uji. Kesamaan tahapan antara data latih dan data uji sangat penting agar model dapat membaca data uji dengan struktur yang sama seperti saat model dilatih. Dengan demikian, model yang dihasilkan dari data latih dapat diterapkan secara tepat pada data uji melalui operator *Apply* Model.



Gambar 4. Pemodelan Data

Pada tahap *Apply* Model, model *Naive Bayes* yang telah terbentuk diterapkan pada data uji untuk menghasilkan prediksi status kelulusan siswa. Hasil prediksi tersebut kemudian dikirim ke operator *Performance* untuk dievaluasi. Pada tahap evaluasi ini, *RapidMiner* menghasilkan ukuran performa model

p-ISSN : 2803-1507 e-ISSN : 2803-1531

seperti akurasi, *precision, recall*, dan *confusion matrix*. Nilai akurasi menunjukkan seberapa besar persentase prediksi yang benar dari seluruh data uji, sedangkan *precision* dan *recall* memberikan gambaran lebih rinci mengenai kemampuan model dalam mengenali masing-masing kelas, yaitu Lulus dan Tidak Lulus. *Confusion matrix* juga membantu melihat jumlah data yang diprediksi dengan benar maupun yang masih salah klasifikasi. Berdasarkan alur tersebut, dapat dipahami bahwa proses pada *RapidMiner* dilakukan secara sistematis, dimulai dari pembacaan data, pengaturan atribut, pembersihan data, pembentukan model, penerapan model pada data uji, hingga evaluasi hasil. Dengan alur ini, penelitian dapat menunjukkan bahwa algoritma *Naive Bayes* mampu digunakan untuk memprediksi kelulusan siswa berdasarkan data yang telah dipersiapkan sebelumnya. Proses pemodelan data menggunakan *RapidMiner* ditunjukkan pada Gambar 4. Gambar tersebut memperlihatkan alur pembacaan data, penentuan *label*, seleksi atribut, penanganan nilai kosong, penerapan algoritma *Naïve Bayes*, hingga evaluasi performa model.

4.4 *Evaluasi*

Berdasarkan hasil penelitian yang telah dilakukan, implementasi *data mining* menggunakan algoritma klasifikasi *Naïve Bayes* dapat digunakan sebagai alat bantu pendukung keputusan dalam memprediksi tingkat kelulusan siswa SMP pada SMP Kartika VI Jayapura. Hal ini ditunjukkan oleh nilai akurasi sebesar 82,76%, yang berarti bahwa model mampu memberikan prediksi yang benar terhadap sebagian besar data uji yang digunakan. Hasil tersebut menunjukkan bahwa data akademik dan nonakademik siswa yang sebelumnya hanya tersimpan sebagai arsip sekolah dapat diolah menjadi informasi prediktif yang berguna. Dengan demikian, penerapan *data mining* dalam penelitian ini tidak hanya berfokus pada implementasi algoritma, tetapi juga diarahkan untuk membantu sekolah dalam mengidentifikasi siswa yang berpotensi mengalami kendala kelulusan secara lebih cepat, sistematis, dan berbasis data.

Jika ditinjau dari hasil *confusion matrix*, model berhasil memprediksi 60 siswa yang benar-benar berstatus Lulus sebagai Lulus, serta 12 siswa yang benar-benar berstatus Tidak Lulus sebagai Tidak Lulus. Namun, masih terdapat 13 siswa yang sebenarnya Lulus tetapi diprediksi Tidak Lulus, dan 2 siswa yang sebenarnya Tidak Lulus tetapi diprediksi Lulus. Hasil ini menunjukkan bahwa model *Naïve Bayes* telah mampu mengenali pola hubungan antara atribut siswa dengan status kelulusannya. Atribut seperti nilai mata pelajaran, kehadiran, sikap, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar memiliki kontribusi dalam membentuk model prediksi. Dalam konteks penelitian ini, hasil tersebut memperkuat bahwa algoritma *Naïve Bayes* relevan digunakan sebagai pendekatan awal dalam membangun alat bantu pendukung keputusan untuk prediksi kelulusan siswa SMP.

accuracy: 82.76%

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	60	2	96.77%
pred. Tidak Lulus	13	12	48.00%
class recall	82.19%	85.71%	

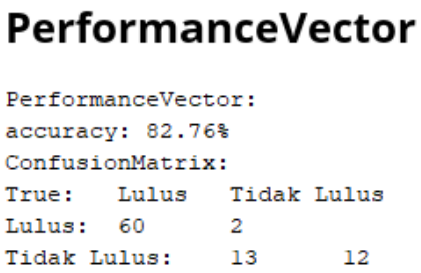
Gambar 5. Hasil Model Uji

Hasil pengujian model terhadap data uji ditampilkan pada Gambar 5. Gambar tersebut memperlihatkan nilai akurasi, *class precision*, dan *class recall* dari masing-masing kelas prediksi.nilai *class precision* pada kelas Lulus sebesar 96,77% menunjukkan bahwa hampir seluruh siswa yang diprediksi Lulus oleh model memang benar-benar termasuk dalam kategori Lulus. Hal ini menandakan bahwa model memiliki ketepatan prediksi yang sangat baik pada kelas tersebut. Sementara itu, nilai *class recall* untuk kelas Tidak Lulus sebesar 85,71% menunjukkan bahwa model cukup baik dalam mendeteksi siswa yang benar-benar berada pada kategori Tidak Lulus. Meskipun demikian, nilai *precision* untuk kelas Tidak Lulus hanya sebesar 48,00%. Artinya, dari seluruh siswa yang diprediksi Tidak Lulus, masih banyak yang sebenarnya termasuk kategori Lulus. Kondisi ini menunjukkan bahwa model belum memiliki performa yang seimbang pada kedua kelas, terutama dalam memberikan ketepatan prediksi terhadap siswa berisiko tidak lulus. Ringkasan hasil evaluasi model dalam bentuk *Performance Vector* ditampilkan pada Gambar 6. Informasi tersebut digunakan untuk melihat performa model berdasarkan nilai akurasi dan *confusion matrix*.

Hasil *Performance Vector* menunjukkan bahwa model *Naïve Bayes* memperoleh nilai akurasi sebesar 82,76%. Nilai tersebut menunjukkan bahwa secara umum model mampu melakukan klasifikasi dengan cukup baik terhadap data uji. Namun, hasil evaluasi tidak hanya dapat dilihat dari akurasi, karena akurasi yang tinggi belum tentu menunjukkan bahwa model memiliki kemampuan yang seimbang dalam mengenali seluruh kelas. Oleh karena itu, nilai *precision, recall,* dan *confusion matrix* perlu diperhatikan untuk memahami

p-ISSN : 2803-1507 e-ISSN : 2803-1531

performa model secara lebih rinci. Berdasarkan hasil *Performance Vector*, model memiliki kemampuan yang lebih kuat dalam mengenali kelas Lulus dibandingkan kelas Tidak Lulus. Hal ini terlihat dari nilai *class precision* kelas Lulus sebesar 96,77%, sedangkan nilai *precision* kelas Tidak Lulus hanya sebesar 48,00%. Kondisi tersebut menunjukkan bahwa prediksi pada kelas Tidak Lulus masih memiliki tingkat kesalahan yang cukup tinggi. Dengan kata lain, masih terdapat siswa yang sebenarnya Lulus tetapi diprediksi sebagai Tidak Lulus oleh model.



Gambar 6. *Performance Vector*

Temuan ini penting karena tujuan utama penelitian adalah membantu sekolah mengidentifikasi siswa yang berisiko tidak lulus. Oleh karena itu, meskipun hasil akurasi tergolong cukup baik, model ini tidak dapat dijadikan sebagai dasar keputusan tunggal dalam menentukan kelulusan siswa. Model *Naïve Bayes* yang dihasilkan lebih tepat digunakan sebagai alat bantu peringatan dini atau pendukung keputusan awal. Hasil prediksi tetap perlu divalidasi oleh guru, wali kelas, dan pihak sekolah sebelum digunakan dalam pengambilan keputusan. Dengan demikian, hasil *Performance Vector* menunjukkan bahwa model cukup bermanfaat untuk membantu sekolah menentukan siswa yang membutuhkan perhatian lebih, tetapi masih memerlukan pertimbangan manusia dalam penerapannya.

5. DISKUSI

Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* mampu menghasilkan model prediksi kelulusan siswa SMP Kartika VI Jayapura dengan akurasi sebesar 82,76%. Nilai tersebut menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar. Atribut yang digunakan dalam penelitian ini meliputi data akademik dan nonakademik, seperti nilai mata pelajaran, sikap, kehadiran, pelanggaran tata tertib, keaktifan ekstrakurikuler, dan keikutsertaan bimbingan belajar. Hal ini menunjukkan bahwa prediksi kelulusan siswa tidak hanya dapat dilihat dari aspek akademik, tetapi juga perlu mempertimbangkan faktor kedisiplinan, kehadiran, serta keterlibatan siswa dalam kegiatan pendukung pembelajaran. Jika dibandingkan dengan penelitian terdahulu yang membahas prediksi kelulusan pada bidang pendidikan, hasil penelitian ini menunjukkan bahwa *Naïve Bayes* masih relevan digunakan sebagai metode klasifikasi. Beberapa penelitian sebelumnya juga menunjukkan bahwa data pendidikan dapat diolah menjadi informasi prediktif untuk membantu institusi dalam memperkirakan kemungkinan kelulusan peserta didik. Kesamaan penelitian ini dengan penelitian sebelumnya terletak pada penggunaan data akademik sebagai dasar prediksi, sedangkan perbedaannya terletak pada penggunaan atribut nonakademik yang lebih beragam dan objek penelitian yang berfokus pada siswa SMP.

Dibandingkan dengan penelitian yang menggunakan algoritma lain seperti *Decision Tree*, hasil akurasi penelitian ini masih dapat dikatakan cukup baik meskipun belum menjadi yang tertinggi. Perbedaan hasil tersebut dapat dipengaruhi oleh jumlah data, karakteristik atribut, distribusi kelas, serta cara kerja algoritma yang digunakan. *Naïve Bayes* bekerja berdasarkan perhitungan probabilitas dan asumsi independensi antaratribut, sedangkan algoritma lain dapat membentuk pola keputusan berdasarkan percabangan atau aturan tertentu. Oleh karena itu, hasil akurasi 82,76% menunjukkan bahwa *Naïve Bayes* layak digunakan sebagai pendekatan awal dalam prediksi kelulusan siswa. Meskipun akurasi model tergolong cukup baik, hasil evaluasi menunjukkan bahwa model lebih kuat dalam mengenali siswa pada kategori Lulus dibandingkan kategori Tidak Lulus. Kondisi ini penting diperhatikan karena tujuan utama penelitian adalah membantu sekolah mengidentifikasi siswa yang berisiko tidak lulus. Oleh karena itu, model *Naïve Bayes* dalam penelitian ini lebih tepat digunakan sebagai alat bantu awal atau sistem peringatan dini, bukan sebagai dasar keputusan tunggal dalam menentukan kelulusan siswa. Hasil prediksi tetap perlu dipertimbangkan bersama

p-ISSN : 2803-1507 e-ISSN : 2803-1531

penilaian guru, wali kelas, dan pihak sekolah agar pembinaan serta pendampingan belajar dapat diberikan secara lebih tepat sasaran.

6. KESIMPULAN DAN SARAN

1.1 Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi *data mining* menggunakan algoritma klasifikasi *Naïve Bayes* dapat digunakan untuk memprediksi tingkat kelulusan siswa di SMP Kartika VI Jayapura dan telah mencapai tujuan penelitian dengan baik. Data yang digunakan berjumlah 289 data siswa dengan atribut prediktor berupa Nilai Matematika, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Sikap, Kehadiran Persen, Pelanggaran Tata Tertib, Keaktifan Ekstrakurikuler, dan Mengikuti Bimbingan Belajar, sedangkan label yang diprediksi adalah Status Kelulusan. Hasil pengujian menunjukkan bahwa model *Naïve Bayes* memperoleh nilai akurasi sebesar 82,76%, sehingga dapat dikategorikan cukup baik dalam memprediksi kelulusan siswa. Hasil *confusion matrix* menunjukkan bahwa model lebih kuat dalam mengenali kategori Lulus, tetapi tetap mampu mendeteksi siswa pada kategori Tidak Lulus. Dengan demikian, model yang dihasilkan layak digunakan sebagai alat bantu awal bagi pihak sekolah dalam mendukung identifikasi dini terhadap siswa yang berpotensi mengalami kendala kelulusan, sehingga sekolah dapat mengambil langkah pembinaan, pendampingan, dan pengawasan secara lebih terarah untuk meningkatkan tingkat kelulusan siswa.

1.2 Saran

Penelitian selanjutnya disarankan untuk menambah jumlah data dan memperluas periode pengambilan data agar model yang dihasilkan menjadi lebih representatif dan mampu menggambarkan kondisi siswa secara lebih menyeluruh. Selain itu, penelitian berikutnya dapat menambahkan atribut lain yang dianggap berpengaruh terhadap kelulusan, seperti dukungan orang tua, motivasi belajar, hasil ujian sekolah, atau faktor sosial lainnya agar hasil prediksi menjadi lebih akurat. Untuk meningkatkan performa model, penelitian selanjutnya juga dapat membandingkan algoritma *Naïve Bayes* dengan metode klasifikasi lain, seperti *Decision Tree*, *K-Nearest Neighbor*, atau *Random Forest*, sehingga dapat diketahui metode yang paling sesuai untuk prediksi kelulusan siswa. Pihak sekolah juga disarankan untuk memanfaatkan hasil penelitian ini sebagai dasar pengembangan sistem pendukung keputusan sederhana yang dapat membantu guru dan manajemen sekolah dalam memantau siswa yang berisiko tidak lulus sejak dini.

DAFTAR PUSTAKA

[1] D. A. Punkastyo, F. Septian, and A. Syaripudin, “Implementasi Data Mining menggunakan Algoritma Naïve Bayes untuk Prediksi Kelulusan Siswa,” Journal of System and Computer Engineering (JSCE), vol. 5, no. 1, pp. 24–35, Jan. 2024, doi: 10.61628/jsce.v5i1.1073.

[2] T. Adrian and N. Suarna, “Implementasi Data Mining untuk Mengklasifikasi Hasil Kelulusan Madrasah menggunakan Algoritma Naive Bayes,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 1, pp. 1000–1004, Mar. 2024, doi: 10.36040/JATI.V8I1.8892.

[3] R. M. Apsari, “Penerapan Metode Naïve bayes dalam Memprediksi Prestasi Siswa,” Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), vol. 4, no. 2, pp. 38–46, Aug. 2024, doi: 10.55382/JURNALPUSTAKAAI.V3I3.760.

[4] A. David Imanuel, N. Nawaningtyas Pusparini, and A. Sani, “Klasifikasi untuk Memprediksi Tingkat Kelulusan Mahasiswa STMIK Widuri Menggunakan Algoritma Naive Bayes,” JURNAL ILMIAH INFORMATIKA, vol. 12, no. 01, pp. 1–7, Mar. 2024, doi: 10.33884/jif.v12i01.8201.

[5] M. Hakim et al., “Analisis Prediksi Kelulusan Mahasiswa Universitas Dinamika Bangsa menggunakan Metode Naïve Bayes,” ejournal.unama.ac.idMF Hakim, S SaputraJurnal Informatika Dan Rekayasa Komputer (JAKAKOM), 2025•ejournal.unama.ac.id, doi: 10.33998/JAKAKOM.2025.5.1.1999.

[6] A. Hussen et al., “Modeling of Student Graduation Prediction using the Naive Bayes Classifier Algorithm,” ieeexplore.ieee.orgAAME Hussen, A Saikhu2024 3rd International Conference on Creative Communication and, 2024•ieeexplore.ieee.org, Accessed: Jun. 09, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10701117/

[7] Junaidi, R. Valicia Anggela, and I. Fadhli, “Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Metode Data Mining dengan Algoritma Naive Bayes,” Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika, vol. 9, no. 2, pp. 65–74, Sep. 2023, doi:

p-ISSN : 2803-1507 e-ISSN : 2803-1531

10.22202/EI.2023.V9I2.7324.

[8] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Educational Data Mining on Student Academic Performance Prediction: A Survey: Educational Data Mining pada Prediksi Kinerja Akademik Mahasiswa : Sebuah Survey," Procedia of Social Sciences and Humanities, vol. 3, pp. 1442–1447, Jul. 2022, doi: 10.21070/PSSH.V3I.344.

[9] T. Le Quy, T. H. Nguyen, G. Friege, and E. Ntoutsi, "Evaluation of Group Fairness Measures in Student Performance Prediction Problems," SpringerT Le Quy, TH Nguyen, G Friege, E NtoutsiJoint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022•Springer, vol. 1752 CCIS, pp. 119–136, 2023, doi: 10.1007/978-3-031-23618-1_8.

[10] F. R. Al-Giffary and M. Martanto, "Klasifikasi Kelulusan Siswa Tahun 2024 Menggunakan Metode Decision Tree (Studi Kasus SMA Islam Alazhar 5 Cirebon)," Jurnal Manajamen Informatika Jayakarta, vol. 4, no. 2, p. 195, Apr. 2024, doi: 10.52362/jmijayakarta.v4i2.1408.

[11] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means untuk Klasterisasi Data Obat pada Rumah Sakit ASRI," NUANSA INFORMATIKA, vol. 16, no. 1, pp. 125–133, Jan. 2022, doi: 10.25134/NUANSA.V16I1.5294.

[12] E. Fitriana et al., "Clustering Kos dengan Algoritma K-Means Untuk Rekomendasi Tempat Berdasarkan Harga dan Fasilitas," Jurnal Media Informatika, vol. 6, no. 3, pp. 1990–1995, Jun. 2025, doi: 10.55338/JUMIN.V6I3.6188.

[13] S. Rachardian and E. Sediyono, "Prediksi Kelulusan Tepat Waktu Mahasiswa untuk Pemantauan Program Studi Menggunakan Metode Data Mining," AITI, vol. 21, no. 2, pp. 168–182, Sep. 2024, doi: 10.24246/aiti.v21i2.168-182.

[14] A. Wijaya, Anita, M. C. Manurung, and Y. D. S. Sitanggang, "Penerapan Data Mining untuk Prediksi Kelulusan Siswa Sekolah Dasar Menggunakan Algoritma Naïve Bayes Classifier," Jurnal Algoritma, vol. 23, no. 1, May 2026, doi: 10.33364/algoritma/v.23-1.3448.

[15] W. Lastari and J. Jasmir, "Penerapan Data Mining Untuk Memprediksi Prestasi Siswa SMA pada Dinas Pendidikan Provinsi Jambi," Jurnal Manajemen Sistem Informasi, vol. 8, no. 2, pp. 310–321, Jun. 2023, Accessed: Jul. 22, 2026. [Online]. Available: https://ejournal.unama.ac.id/index.php/jurnalmsi/article/view/864

[16] A. David Imanuel, N. Nawaningtyas Pusparini, and A. Sani, "Klasifikasi untuk Memprediksi Tingkat Kelulusan Mahasiswa STMIK Widuri Menggunakan Algoritma Naive Bayes," JURNAL ILMIAH INFORMATIKA, vol. 12, no. 01, pp. 1–7, Mar. 2024, doi: 10.33884/jif.v12i01.8201.

[17] Pusparini et al., "Klasifikasi untuk Memprediksi Tingkat Kelulusan Mahasiswa STMIK Widuri Menggunakan Algoritma Naive Bayes," ejournal.upbatam.ac.idAD Imanuel, NN Pusparini, A SaniJurnal Ilmiah Informatika, 2024•ejournal.upbatam.ac.id, vol. 12, no. 01, pp. 1–7, Mar. 2024, doi: 10.33884/JIF.V12I01.8201.

[18] S. M. Hudzaifah, D. Gunawan, and D. Iskandar, "Implementasi Algoritma Naïve Bayes dalam Memprediksi Tingkat Kelulusan Siswa pada Sertifikasi Mikrotik Certified Network Associate (MTCNA)," Jurnal Indonesia : Manajemen Informatika dan Komunikasi, vol. 5, no. 3, pp. 3225–3236, Sep. 2024, doi: 10.35870/jimik.v5i3.1012.

[19] F. Fauziah and A. S. Ramadhantya, "Penggunaan Rapidminer untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes," Jurnal Teknologi Informasi, vol. 10, no. 1, pp. 52–60, Jun. 2024, doi: 10.52643/JTI.V10I1.4671.

[20] A. Riyanto and E. S. Ompusunggu, "Implementasi Data Mining untuk Mengklasifikasi Hasil Belajar Siswa/i dengan Metode Naïve Bayes," Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP), vol. 7, no. 2, pp. 129–141, Oct. 2024, doi: 10.34012/JUTIKOMP.V7I2.5237.

[21] H. Latifah and Sri Mujiyono, "Perbandingan Algoritma Naive Bayes , K-NN , ID3 , dan SVM dalam Menentukan Prediksi Kelulusan Siswa di SMK Muhamadiah Majenang," Jurnal Mahasiswa Teknik Informatika, vol. 2, no. 1, pp. 38–45, Apr. 2023, doi: 10.35473/jamastika.v2i1.1871.

[22] Ramadhani Zidan Arifin, Hasbi Firmansyah, and Wahyu Asriyani, "Prediksi Kelulusan Siswa

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Berdasarkan Data Demografis dan Akademik pada Dataset Student Performance," Jurnal Pengabdian Masyarakat dan Riset Pendidikan, vol. 4, no. 2, pp. 13300–13307, Dec. 2025, doi: 10.31004/jerkin.v4i2.4251.

[23] L. Baridah and R. A. Putri, "Predicting Student Learning Outcomes in Vocational Computer and Network Engineering Using Naïve Bayes," Indonesian Journal of Artificial Intelligence and Data Mining, vol. 8, no. 3, pp. 667–676, Nov. 2025, doi: 10.24014/IJAIDM.V8I3.38333.

[24] R. Arifin, H. Firmansyah, "Prediksi Kelulusan Siswa berdasarkan Data Demografis dan Akademik pada Dataset Student Performance: Penelitian," jerkin.orgRZ Arifin, H Firmansyah, W AsriyaniJurnal Pengabdian Masyarakat dan Riset Pendidikan, 2025•jerkin.org, vol. 4, no. 2, pp. 13300–13307, Dec. 2025, doi: 10.31004/JERKIN.V4I2.4251.

[25] E. Kuznetsova and M. Nkosi, "Adaptive Fuzzy Logic Integration for Optimizing Decision Support Systems under Data Uncertainty," Journal of Technology Informatics and Engineering, vol. 4, no. 3, pp. 463–477, Dec. 2025, doi: 10.51903/JTIE.V4I3.456.