JIMU: Jurnal Ilmiah Multi Disiplin
Vol. 03, No. 02, Tahun 2025, Hal. 1319-1329, ISSN: 3031-9498 (Online)

Implementasi *Naive Bayes* dan *K-Nearest Neighbor* (KNN) untuk Identifikasi Dini Siswa Berpotensi *Dropout*

Yudisti Prayigo Permana[1], Muhammad Ihsan Ashari[2]

[12]Universitas Pamulang
E-mail: dosen03142@unpam.ac.id[1] , dosen03154@unpam.ac.id[2]

Article Info

Article history:

Received Juni 01, 2025
Revised Juni 13, 2025
Accepted Juni 20, 2025

Keywords:

Dropout, Naive Bayes, K-Nearest Neighbor, Data Mining, Academic Prediction, Machine Learning

ABSTRACT

The phenomenon of school dropouts remains a critical issue in the education sector, particularly in efforts to maintain student learning continuity. Therefore, a predictive approach is needed to identify students with a potential for dropping out early on, so that preventive measures can be taken immediately. This study utilizes the Naive Bayes and K-Nearest Neighbor (KNN) classification algorithms to analyze the likelihood of students dropping out based on behavioral data, academic achievements, and socio-economic conditions. A total of 500 student data were used, with 100 of them serving as the testing sample. The analysis process was conducted using RapidMiner software. From the test results, the KNN algorithm showed higher performance than Naive Bayes in terms of dropout prediction, with an accuracy of 98%, while Naive Bayes achieved an accuracy of 97%. KNN also demonstrated high recall capability in detecting at-risk students, making it more suitable for systems requiring high accuracy in sensitive cases such as dropout. However, Naive Bayes remains a viable alternative due to its processing speed and simplicity of implementation. The most significant variables in determining dropout potential include high absenteeism, low academic performance, minimal parental support, and an unfavorable economic background. These findings emphasize the importance of schools and authorities paying special attention to these factors. Thus, the application of machine learning technology in the education system can provide strategic solutions to anticipate dropout cases more effectively. This prediction system is expected to support more targeted decision-making in efforts to prevent students from dropping out of school.

Article Info

Article history:

Received Juni 01, 2025
Revised Juni 13, 2025
Accepted Juni 20, 2025

Keywords:

Dropout, Naive Bayes, K-Nearest Neighbor, Data Mining, Prediksi Akademik, Machine Learning

ABSTRACT

Fenomena putus sekolah masih menjadi isu kritis dalam sektor pendidikan, khususnya dalam upaya mempertahankan kelangsungan belajar peserta didik. Untuk itu, diperlukan suatu pendekatan prediktif yang mampu mengidentifikasi siswa dengan potensi dropout sejak dini, sehingga tindakan preventif dapat segera dilakukan. Penelitian ini memanfaatkan algoritma klasifikasi Naive Bayes dan K-Nearest Neighbor (KNN) dalam menganalisis kemungkinan siswa mengalami dropout berdasarkan data perilaku, capaian akademik, serta kondisi sosial ekonomi. Sebanyak 500 data siswa digunakan, dengan 100 di antaranya dijadikan sampel pengujian. Proses analisis dilakukan menggunakan perangkat lunak RapidMiner. Dari hasil pengujian, algoritma KNN menunjukkan performa yang lebih tinggi dibandingkan Naive Bayes dalam hal prediksi dropout, dengan akurasi

mencapai 98%, sementara Naive Bayes memperoleh akurasi 97%. KNN juga menunjukkan kemampuan recall yang tinggi dalam mendeteksi siswa berisiko, sehingga lebih sesuai diterapkan dalam sistem yang membutuhkan akurasi tinggi terhadap kasus-kasus sensitif seperti dropout. Meski demikian, Naive Bayes tetap menjadi alternatif yang dapat dipertimbangkan berkat kecepatan proses dan kesederhanaan implementasinya. Beberapa variabel yang paling signifikan dalam menentukan potensi dropout antara lain adalah jumlah ketidakhadiran yang tinggi, nilai akademik yang rendah, dukungan orang tua yang minim, serta latar belakang ekonomi yang tidak mendukung. Temuan ini menekankan pentingnya perhatian khusus dari sekolah dan pihak berwenang terhadap faktor-faktor tersebut. Dengan demikian, penerapan teknologi machine learning dalam sistem pendidikan dapat memberikan solusi strategis untuk mengantisipasi kasus dropout secara lebih efektif. Sistem prediksi ini diharapkan dapat mendukung pengambilan keputusan yang lebih tepat sasaran dalam upaya mencegah siswa putus sekolah.

Corresponding Author:

Yudisti Prayigo Permana
Universitas Pamulang
Email: dosen03142@unpam.ac.id

PENDAHULUAN

Pendidikan merupakan fondasi utama dalam membangun sumber daya manusia yang berkualitas. Namun, salah satu tantangan besar yang dihadapi oleh institusi pendidikan, terutama di Indonesia, adalah tingkat dropout (putus sekolah) yang masih cukup tinggi. Berdasarkan data Kementerian Pendidikan dan Kebudayaan (Kemdikbud), pada tahun 2022, terdapat lebih dari 200.000 siswa di Indonesia yang tidak dapat menyelesaikan pendidikan dasar dan menengah karena berbagai faktor, termasuk kesulitan ekonomi, rendahnya prestasi akademik, dan masalah perilaku. Dropout tidak hanya berdampak pada perkembangan akademik siswa, tetapi juga memengaruhi kualitas sumber daya manusia secara jangka panjang. Beberapa faktor yang berkontribusi terhadap risiko dropout antara lain absensi tinggi, nilai akademik rendah, kondisi sosial-ekonomi yang kurang mendukung, pelanggaran disiplin, hingga kurangnya dukungan dari orang tua (Amri et al., 2023). Seiring perkembangan teknologi, pendekatan educational data mining dan machine learning mulai digunakan dalam sistem pendidikan untuk mengidentifikasi siswa yang berpotensi mengalami masalah akademik termasuk dropout. Salah satu teknik yang populer adalah klasifikasi data untuk mendeteksi siswa berisiko tinggi (Romero & Ventura, 2024). Studi oleh Kamila dan Subastian (2020) menunjukkan bahwa algoritma Naive Bayes mampu memprediksi potensi dropout mahasiswa dengan akurasi tinggi. Sementara itu, algoritma K-Nearest Neighbor (KNN) juga terbukti efektif dalam mengklasifikasikan kelulusan atau ketidaklulusan siswa dengan akurasi lebih dari 95% (Amandha et al., 2023). Kedua metode ini bersifat supervised dan cocok untuk mengklasifikasi variabel biner seperti status dropout (ya/tidak). Dampak dari dropout tidak

hanya dirasakan oleh individu, tetapi juga oleh masyarakat dan perekonomian negara, karena berkurangnya kesempatan kerja dan meningkatnya ketimpangan sosial. Sekolah selama ini mengandalkan identifikasi manual oleh guru atau konselor (BK) untuk mendeteksi siswa yang berisiko dropout. Namun, pendekatan ini memiliki beberapa kelemahan, seperti (1) Subjektivitas – Penilaian sering bergantung pada pengamatan personal tanpa analisis data yang komprehensif. (2) Ketertinggalan Deteksi – Intervensi biasanya baru dilakukan setelah siswa menunjukkan tanda-tanda parah (seperti bolos berkali-kali atau nilai anjlok). (3) Ketidakmampuan Memproses Banyak Variabel – Faktor penyebab dropout sangat kompleks (akademik, ekonomi, psikologis), sehingga sulit dianalisis secara manual. Untuk mengatasi masalah ini, penerapan teknologi machine learning seperti Naive Bayes dan K-Nearest Neighbor (KNN) dapat menjadi solusi. Kedua algoritma ini dipilih karena Naive Bayes efektif untuk klasifikasi berbasis probabilitas, cocok untuk data perilaku dan kategorikal seperti status ekonomi, frekuensi bolos selain itu KNN mampu menangani hubungan non-linear antara variabel seperti interaksi antara nilai akademik dan dukungan orang tua. Penelitian ini akan memanfaatkan data historis siswa (seperti nilai rapor, kehadiran, partisipasi di kelas, dan latar belakang keluarga) untuk membangun model prediktif. Dengan sistem ini, sekolah dapat mengidentifikasi siswa berisiko lebih dini sebelum kondisinya semakin buruk dan menyusun intervensi yang tepat seperti bantuan beasiswa, pendampingan belajar, atau konseling berdasarkan faktor dominan yang memengaruhi risiko dropout juga dapat mengoptimalkan alokasi sumber daya dengan fokus pada siswa yang paling membutuhkan. Penelitian ini diharapkan tidak hanya memberikan kontribusi dalam bidang edutech, tetapi juga mendukung Tujuan Pembangunan Berkelanjutan (SDGs) poin ke-4, yaitu pendidikan inklusif dan berkualitas bagi semua. Dengan pendekatan berbasis data, upaya pencegahan dropout dapat dilakukan secara lebih objektif, akurat, dan proaktif.

LANDASAN TEORI

1. Data Mining

Data Mining adalah proses mengekstraksi dan menganalisis data dalam jumlah besar dari berbagai sumber untuk mengidentifikasi pola, hubungan, atau tren yang bermakna, yang dapat digunakan untuk pengambilan keputusan bisnis, prediksi, atau optimisasi.

Komponen Utama Data Mining

a. Ekstraksi Data – Mengumpulkan data dari database, data warehouse, atau sumber lain.
b. Pemrosesan Data – Membersihkan dan mempersiapkan data untuk analisis.
c. Analisis Pola – Menggunakan algoritma statistik, machine learning, atau AI untuk menemukan pola tersembunyi.
d. Interpretasi & Visualisasi – Menyajikan hasil dalam bentuk yang mudah dipahami (grafik, laporan, dll.).

Teknik Data Mining

a. Classification (Klasifikasi data ke dalam kategori)
b. Clustering (Pengelompokan data berdasarkan kemiripan)
c. Association Rule Learning (Mencari hubungan antar variabel, seperti market basket analysis)
d. Regression (Memprediksi nilai numerik berdasarkan data historis)
e. Anomaly Detection (Mendeteksi outlier atau data tidak normal)

Menurut Wahyudi et al. (2022) dalam penelitian lokal mendefinisikan data mining sebagai proses gabungan dari berbagai disiplin (machine learning, pengenalan pola, statistik, database, visualisasi) untuk mengekstrak informasi penting dari basis data besar. Proses ini termasuk bagian dari knowledge discovery in databases (KDD).

Data mining dalam konteks pendidikan dikenal sebagai *Educational Data Mining (EDM)*, yang bertujuan untuk mengekstraksi informasi berguna dari data akademik siswa untuk mendukung pengambilan keputusan (Romero & Ventura, 2024). Teknik ini sering digunakan untuk memprediksi keberhasilan akademik, deteksi keterlambatan kelulusan, hingga identifikasi risiko dropout.

2. Naive Bayes

Naive Bayes adalah algoritma machine learning berbasis probabilitas yang digunakan untuk klasifikasi data dengan mengandalkan teorema Bayes (Bayes' Theorem). Algoritma ini disebut "naive" (naif) karena membuat asumsi sederhana bahwa semua fitur (variabel prediktor) bersifat independen (bebas) satu sama lain, meskipun dalam kenyataannya belum tentu demikian. Naive Bayes merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur. Metode ini dikenal memiliki performa baik dalam dataset berskala besar dengan proses training yang cepat (Jefri & Fatah, 2025). Dalam konteks pendidikan, Naive Bayes telah digunakan untuk memprediksi kelulusan dan dropout dengan hasil akurasi yang memuaskan (Marzuqi et al., 2021).

Konsep Dasar Naive Bayes

Rumus dasar Naive Bayes berasal dari Teorema Bayes:

$$P(A|B) = \frac{P(B|A) \, . P(A)}{P(B)}$$

Dalam konteks klasifikasi:

P(A|B) = Probabilitas kelas *A* diberikan fitur *B* (*posterior probability*).

P(B|A) = Probabilitas fitur *B* muncul di kelas *A* (*likelihood*).

P(A) = Probabilitas kelas *A* (*prior probability*).

P(B) = Probabilitas fitur *B* (*evidence*).

Algoritma ini menghitung probabilitas suatu data masuk ke dalam suatu kelas berdasarkan fitur-fiturnya dan memilih kelas dengan probabilitas tertinggi.

3. K-Nearest Neighbor (KNN)

KNN adalah algoritma klasifikasi non-parametrik yang menentukan kelas suatu objek berdasarkan mayoritas tetangganya. KNN cocok digunakan dalam lingkungan data pendidikan karena kesederhanaannya serta mampu menangani data numerik dan kategorik (Amandha et al., 2023).

K-Nearest Neighbors (KNN) adalah algoritma *supervised learning* yang digunakan untuk klasifikasi dan regresi berdasarkan kedekatan (*similarity*) data. Algoritma ini bekerja dengan memprediksi kelas/nilai suatu data baru berdasarkan "K" tetangga terdekat-nya dalam ruang fitur.

Cara Kerja KNN:

a. Hitung Jarak antara data baru dengan semua data latih (menggunakan *Euclidean Distance*, *Manhattan Distance*, dll.).
b. Pilih "K" tetangga terdekat (nilai K bisa 3, 5, dst.).
c. Prediksi Kelas/Nilai:
    1) Klasifikasi: Ambil mayoritas kelas dari K tetangga.
    2) Regresi: Hitung rata-rata nilai dari K tetangga.

4. Faktor-faktor Dropout

Beberapa indikator utama dalam prediksi dropout mencakup: ketidakhadiran, prestasi akademik, jarak rumah ke sekolah, keikutsertaan ekstrakurikuler, pelanggaran disiplin, status ekonomi, dan tingkat dukungan orang tua (Alfiansyah & Soetanto, 2024; Kodratillah et al., 2021).

Dampak Dropout

a. Ekonomi: Kesulitan mendapat pekerjaan layak, berisiko masuk lingkaran kemiskinan.
b. Sosial: Meningkatnya pengangguran, kriminalitas, atau ketergantungan sosial.
c. Individu: Kehilangan kesempatan mengembangkan potensi diri.

METODE

1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen klasifikasi menggunakan algoritma Naive Bayes dan KNN. Tujuan utama adalah membandingkan performa kedua algoritma dalam memprediksi status dropout siswa.

2. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari dokumentasi akademik dan administratif sekolah, mencakup 500 siswa yang masuk pada tahun pelajaran 2023/2024. Data tersebut terdiri dari berbagai variabel yang menggambarkan karakteristik siswa dan faktor-faktor yang berpotensi memengaruhi status dropout. Variabel-variabel tersebut meliputi: Absen yang mencatat jumlah ketidakhadiran siswa dalam satu semester, Nilai yang merepresentasikan rata-rata nilai akademik siswa dalam skala 0–100, serta Status_Ekonomi yang mengkategorikan tingkat ekonomi keluarga siswa ke dalam tiga kelompok (1 = Rendah, 2 = Menengah, 3 = Tinggi). Selain itu, terdapat variabel Jarak_km yang menunjukkan jarak rumah siswa ke sekolah dalam kilometer, Ekstra yang menandakan keikutsertaan siswa dalam kegiatan ekstrakurikuler (0 = Tidak, 1 = Ya), dan Pelanggaran yang mencatat jumlah pelanggaran disiplin yang dilakukan siswa. Variabel Dukungan_OrangTua mengukur tingkat dukungan orang tua terhadap pendidikan anaknya (1 = Rendah, 2 = Sedang, 3 = Tinggi), sementara Dropout berfungsi sebagai variabel target yang menunjukkan status siswa (Tidak dropout, Dropout). Data ini dikumpulkan secara komprehensif untuk menganalisis pola dan faktor dominan yang berkontribusi terhadap kasus dropout di lingkungan sekolah.

3. Proses Analisis

- Pra-pemrosesan: Pembersihan data dan normalisasi
- Modeling: Implementasi Naive Bayes dan KNN menggunakan RapidMiner
- Evaluasi Model: Menggunakan metrik *Accuracy, Precision, Recall, dan F1-Score* untuk membandingkan kinerja model (Febriansyah & Fauzi, 2023)

4. Perangkat Lunak

RapidMiner digunakan sebagai platform utama karena kemampuannya dalam mengolah dan memvisualisasikan data dengan metode drag-and-drop serta mendukung algoritma machine learning (Elbouknify et al., 2025).

5. Pengujian Data

Pada tahap pengujian, digunakan 500 data siswa yang telah dibersihkan dari data tidak lengkap untuk mengevaluasi performa algoritma kemudian menggunakan 100 data siswa sebagai sampel pengujian. Pengujian ini bertujuan mengidentifikasi metode terbaik dengan membandingkan nilai precision, recall, dan accuracy dari setiap algoritma yang dianalisis. Berikut rumus yang digunakan untuk menghitung ketiga metrik evaluasi tersebut:

a. *Accuracy*

Akurasi adalah rasio antara prediksi yang benar (positif dan negatif) terhadap seluruh data. Rumus akurasi adalah:

$$Accuracy = \left(\frac{True\ Positif + True\ Negatif}{Total\ data}\right) x\ 100\%$$

b. *Precision*

Presisi adalah rasio antara jumlah prediksi positif yang benar terhadap seluruh hasil prediksi yang dinyatakan positif. Rumus presisi adalah:

$$Precision\ = \left(\frac{True\ Positif}{True\ Positif + False\ positif}\right) x\ 100\%$$

c. *Recall*

Recall adalah rasio antara jumlah prediksi positif yang benar terhadap total jumlah data yang sebenarnya positif. Rumus recall adalah:

$$Recall = \left(\frac{True\ Positif}{False\ Positif + True\ Negatif}\right) x\ 100\%$$

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data 500 siswa yang diperoleh melalui observasi dan dokumentasi. Data siswa yang dikumpulkan mencakup seluruh peserta didik pada tahun ajaran 2023/2024.

Tabel 1. Dataset Siswa

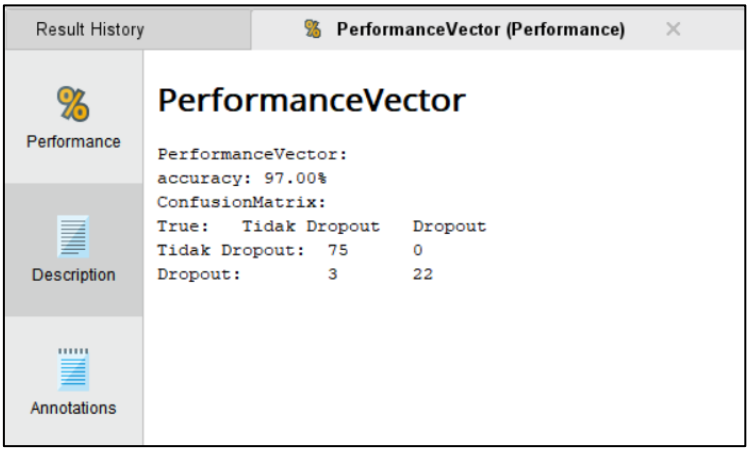
Nama	Absen	Nilai	Status Ekonomi	Jarak (km)	Ekstra	Pelanggaran	Dukungan OrangTua	Dropout
Siswa 1	4	85,47	1	0,15	0	4	2	Tidak Dropout
Siswa 2	14	59,89	1	2,07	1	5	2	Dropout
Siswa 3	10	72,21	2	7,13	1	4	2	Tidak Dropout
Siswa 4	5	81,74	1	3,87	0	3	2	Tidak Dropout
Siswa 5	2	91,95	2	6,62	1	3	2	Tidak Dropout
Siswa 6	15	39,05	1	3,47	0	3	2	Dropout
Siswa 7	4	87,79	1	3,91	0	4	1	Tidak Dropout

Siswa 8	10	70,92	1	1,59	1	1	2	Tidak Dropout
Siswa 9	7	82,27	1	2,58	0	4	3	Tidak Dropout
…	…	…	…	…	…	…	…	…
Siswa 500	0	100	1	9,13	1	1	2	Tidak Dropout

Seluruh data telah melalui proses pembersihan dan preprocessing untuk memastikan kelengkapan dan konsistensi data sebelum dilakukan analisis lebih lanjut. Pengujian model dilakukan dengan membandingkan performa berbagai algoritma machine learning berdasarkan metrik evaluasi akurasi, presisi, dan recall untuk menentukan model prediktif terbaik dalam mengidentifikasi risiko dropout siswa dengan menggunakan sampel sebanyak 100 data siswa. Pengukuran kinerja suatu algoritma penting untuk mengetahui seberapa baik suatu sistem mengorganisir data. Confusion Matrix merupakan perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. Confusion Matrix disajikan dalam bentuk tabel matriks yang menggambarkan performa model klasifikasi terhadap serangkaian data uji yang nilai aktualnya telah diketahui. Pengolahan data siswa dengan metode klasifikasi menggunakan 2 model algoritma, dengan hasil sebagai berikut:

Hasil Algoritma Naive Bayes

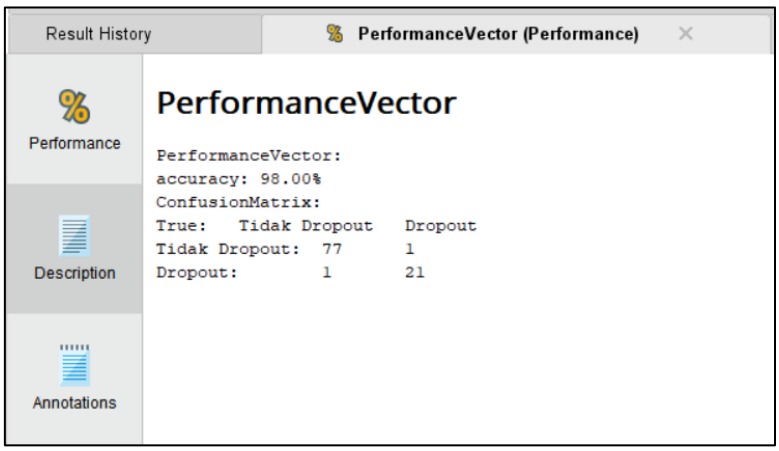
Pengukuran kinerja klasifikasi dari hasil algoritma Naïve Bayes dilakukan menggunakan confusion matrix, yaitu dengan menghitung nilai akurasi, presisi, dan recall. Nilai hasil prediksinya adalah sebagai berikut:



Gambar 1. PerformanceVector Algoritma Naive Bayes

Hasil Algoritma KNN (K-Nearest Neighbor)

Pengukuran kinerja klasifikasi dari hasil algoritma KNN dilakukan menggunakan confusion matrix, yaitu dengan menghitung nilai akurasi, presisi, dan recall. Nilai hasil prediksinya adalah sebagai berikut:



Gambar 2. PerformanceVector Algoritma K-Nearest Neighbor

Hasil Perbandingan Metode

Setelah menerapkan prediksi potensi dropout menggunakan dua metode, yaitu algoritma Naïve Bayes dan K-Nearest Neighbor dengan aplikasi RapidMiner (seperti terlihat pada Gambar 1 dan 2), metode terbaik yang ditemukan adalah algoritma KNN. Hasil pengolahan data menunjukkan perbandingan performa klasifikasi antara algoritma Naïve Bayes dan K-Nearest Neighbor berdasarkan tingkat akurasinya. Semakin tinggi nilai akurasi yang diperoleh dari hasil pengujian, semakin kuat pula validitas arah keputusan yang dihasilkan. Perbandingan nilai akurasi hasil pengolahan data untuk prediksi dropout menggunakan metode klasifikasi antara algoritma Naïve Bayes dan K-Nearest Neighbor dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan Metode

Algorithm	**# Samples**	**Results of Interests**		***Values of Confusion Matrix* (%)**		
		True	***False***	***Accuracy***	***Precision***	***Recall***
Naïve Bayes	100	97	3	**97.00**	100.00	96.15
K-Nearest Neighbor	100	98	2	**98.00**	98.72	98.72

Seperti yang terlihat pada Tabel 2 dan berdasarkan perbandingan tingkat akurasi dalam menentukan arah menggunakan metode klasifikasi antara algoritma Naive Bayes dan K-Nearest Neighbor, diperlihatkan bahwa metode klasifikasi dengan algoritma K-Nearest Neighbor merupakan metode dengan tingkat confusion matrix tertinggi dibandingkan dengan algoritma Naive Bayes, yang masing-masing memperoleh nilai akurasi sebesar 98% dan 97%. Dengan demikian, dapat disimpulkan bahwa untuk memprediksi potensi siswa yang akan dropout, algoritma yang paling tepat digunakan adalah K-Nearest Neighbor

KESIMPULAN

Hasil penelitian ini menunjukkan bahwa metode klasifikasi dalam data mining, khususnya algoritma Naive Bayes dan K-Nearest Neighbor (KNN), dapat digunakan secara efektif untuk mendeteksi secara dini siswa yang berpotensi mengalami dropout. Penelitian ini menganalisis data dari 500 siswa, dengan 100 sampel data digunakan dalam tahap pengujian menggunakan perangkat lunak RapidMiner. Dari hasil pengujian, algoritma KNN menunjukkan performa yang lebih akurat dibandingkan Naive Bayes dalam memprediksi potensi siswa dropout. KNN mampu mencapai tingkat akurasi sebesar 98%, sedangkan Naive Bayes memperoleh akurasi 97%. Dengan tingkat recall yang tinggi dalam mengenali siswa yang berisiko, KNN dinilai sangat tepat untuk digunakan dalam sistem yang menuntut sensitivitas tinggi terhadap potensi permasalahan akademik siswa. Beberapa faktor utama yang terbukti memengaruhi prediksi dropout meliputi: tingginya tingkat ketidakhadiran, rendahnya nilai akademik, status ekonomi keluarga yang rendah, dan minimnya dukungan dari orang tua. Temuan ini menegaskan pentingnya pemantauan secara rutin terhadap indikator-indikator tersebut oleh pihak sekolah dan instansi pendidikan terkait. Dengan demikian, penerapan model prediktif berbasis machine learning seperti KNN dan Naive Bayes dapat menjadi pendekatan inovatif dalam dunia pendidikan. Sistem prediksi ini memungkinkan sekolah untuk melakukan intervensi secara lebih awal, akurat, dan berbasis data, sehingga diharapkan dapat menekan angka putus sekolah dan meningkatkan efektivitas program pendampingan siswa secara tepat sasaran dan efisien..

DAFTAR PUSTAKA

Amandha, S., Rohayani, H., & Kurniawansyah, K. (2023). Implementation of Data Mining for Predicting Student Graduation Using the K-Nearest Neighbor Algorithm at Jambi Muhammadiyah University. Indonesian Journal of Artificial Intelligence and Data Mining, 3(2), 45–52. https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/26150

Amri, Z., Kusrini, K., & Kusnawi, K. (2023). Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM. Edumatic: Jurnal Pendidikan Informatika, 7(2), 187–196. https://doi.org/10.29408/edumatic.v7i2.18620

Alfiansyah, D. M., & Soetanto, H. (2024). Prediksi Keterlambatan Pembayaran SPP Siswa dengan Pendekatan Metode Naive Bayes dan K-Nearest Neighbors. Building of Informatics, Technology and Science (BITS), 5(4), 706–719. https://doi.org/10.47065/bits.v5i4.4643

Elbouknify, I., Berrada, I., Mekouar, L., Iraqi, Y., Bergou, E. H., Belhabib, H., Nail, Y., & Wardi, S. (2025). AI-based identification and support of at-risk students: A case study of the Moroccan education system. arXiv. https://arxiv.org/abs/2504.07160

Febriansyah, A., & Fauzi, I. (2023). Educational Data Mining: Comparison of Models for Predicting Non-Academic Students. Southeast Asian Journal on Open and Distance Learning, 1(2), 15–25. https://odelia-journal.seamolec.org/index.php/current/article/view/18

Jefri, & Fatah, Z. (2025). Data Mining Classification to Predict Student Graduation Using the Naive Bayes Method. Jurnal Ilmiah Teknologi dan Komputer (JITTER), 5(3), 2270–2278. https://ojs.unud.ac.id/index.php/jitter/article/view/121193

Jimenez Martinez, A. L., Sood, K., & Mahto, R. (2024). Early Detection of At-Risk Students Using Machine Learning. arXiv. https://arxiv.org/abs/2412.09483

Kodratillah, E. Y., Daririn, D., & Naya, C. (2021). Penerapan Data Mining Untuk Prediksi Kelulusan Siswa Menggunakan Algoritma Naïve Bayes Pada SMK Garuda. Jurnal SIGMA, 12(4), 45–52. https://jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/1246

Marzuqi, A., Laksitowening, K. A., & Asror, I. (2021). Temporal Prediction on Students' Graduation using Naïve Bayes and K-Nearest Neighbor Algorithm. Jurnal Media Informatika Budidarma, 5(2), 89–97. http://dx.doi.org/10.30865/mib.v5i2.2919

Romero, C., & Ventura, S. (2024). Educational data mining and learning analytics: An updated survey. arXiv. https://arxiv.org/abs/2402.07956

Wahyudi, A. K., Azizah, N., & Saputro, H. (2022). Data Mining Klasifikasi Kepribadian Siswa SMP Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4. 5. Journal of Information System and Computer, 2(2), 8-13. https://doi.org/10.34001/jister.v2i2.392