

PERBANDINGAN METODE ALGORITMA KLASIFIKASI (DECISION TREE DAN NAIVE BAYES) UNTUK MEMPREDIKSI AKURASI TINGKAT KELULUSAN SISWA DI SEKOLAH SDN 008 SUNGAI DUSUN

Rahmathidayatullah

¹Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri
Jl. Provinsi Parit 1 Tembilahan Hulu, Indragiri Hilir, Riau, Indonesia
Email: rahmathidayahrahmat92@gmail.com

ABSTRAK

Penelitian ini membandingkan performa algoritma Decision Tree dan Naive Bayes dalam memprediksi tingkat kelulusan siswa di SDN 008 Sungai Dusun berdasarkan data nilai akademik. Dataset terdiri dari 212 data siswa dengan 12 atribut mata pelajaran serta satu atribut target kelulusan. Proses penelitian meliputi pra-pemrosesan data, normalisasi, transformasi data, dan pengujian menggunakan teknik 10-Fold Stratified Cross Validation. Hasil pengujian menunjukkan bahwa kedua algoritma memiliki performa yang sangat baik. Decision Tree memperoleh akurasi 95,26%, precision 95,63%, dan recall 96,67%, sementara Naive Bayes mencapai akurasi 95,76%, precision 98,00%, dan recall 95,00%. Meskipun Naive Bayes unggul dari segi akurasi dan precision, Decision Tree memberikan nilai recall lebih tinggi serta menghasilkan model yang mudah dipahami oleh pihak sekolah. Dengan demikian, Decision Tree lebih direkomendasikan untuk diterapkan sebagai sistem pendukung keputusan kelulusan siswa karena mampu mendeteksi kelulusan secara lebih konsisten dan interpretatif. Penelitian ini diharapkan dapat membantu sekolah melakukan evaluasi kelulusan secara lebih objektif dan berbasis data.

Kata Kunci: *Decision Tree, Naive Bayes, Data Mining, Prediksi Kelulusan, Cross Validation.*

ABSTRACT

This study compares the performance of the Decision Tree and Naive Bayes algorithms in predicting student graduation at SDN 008 Sungai Dusun based on academic score data. The dataset consists of 212 student records with 12 subject attributes and one graduation target attribute. The research process includes data preprocessing, normalization, data transformation, and evaluation using the 10-Fold Stratified Cross Validation technique. The results show that both algorithms achieve excellent performance. The Decision Tree algorithm achieved an accuracy of 95.26%, precision of 95.63%, and recall of 96.67%, while the Naive Bayes algorithm obtained an accuracy of 95.76%, precision of 98.00%, and recall of 95.00%. Although Naive Bayes outperforms in terms of accuracy and precision, Decision Tree provides a higher recall and generates a more interpretable model for school stakeholders. Therefore, Decision Tree is recommended as the preferred algorithm for developing a decision support system for student graduation prediction due to its better detection capability and interpretability. This research is expected to assist schools in conducting a more objective and data-driven evaluation process.

Keywords: *Decision Tree, Naive Bayes, Data Mining, Graduation Prediction, Cross Validation.*

1 PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pemanfaatan data secara lebih sistematis dalam berbagai bidang, termasuk bidang pendidikan[1]. Salah satu pendekatan yang banyak digunakan adalah *data mining*, yaitu proses penggalian pola dan pengetahuan dari kumpulan data untuk mendukung pengambilan keputusan yang lebih objektif dan berbasis data. Dalam konteks pendidikan, data mining dapat dimanfaatkan untuk menganalisis data akademik siswa guna memprediksi hasil belajar dan tingkat kelulusan secara lebih akurat[2].

Pada tingkat sekolah dasar, penentuan kelulusan siswa merupakan proses yang sangat penting karena berkaitan langsung dengan keberlanjutan pendidikan siswa ke jenjang berikutnya[3]. Namun, pada praktiknya, proses evaluasi kelulusan di beberapa sekolah masih dilakukan secara manual dan sangat bergantung pada pertimbangan subjektif guru. Kondisi ini berpotensi menimbulkan ketidakkonsistenan dalam pengambilan keputusan, terutama ketika jumlah siswa cukup banyak dan data akademik yang tersedia belum dimanfaatkan secara optimal[4].

SDN 008 Sungai Dusun merupakan salah satu sekolah dasar yang masih melakukan penentuan kelulusan siswa berdasarkan penilaian manual tanpa dukungan sistem pendukung keputusan berbasis data. Padahal, sekolah memiliki data akademik siswa yang cukup lengkap, seperti nilai mata pelajaran inti dan muatan lokal, yang dapat diolah lebih lanjut untuk menghasilkan informasi yang bernilai[5]. Oleh karena itu, diperlukan suatu pendekatan berbasis data mining yang mampu membantu pihak sekolah dalam memprediksi kelulusan siswa secara objektif dan konsisten.

Algoritma klasifikasi merupakan salah satu teknik dalam data mining yang banyak digunakan untuk memprediksi suatu kelas atau kategori berdasarkan data historis[6]. Dua algoritma klasifikasi yang sering diterapkan dalam bidang pendidikan adalah *Decision Tree* dan *Naive Bayes*. *Decision Tree* memiliki keunggulan dalam kemudahan interpretasi hasil karena menghasilkan aturan keputusan yang jelas, sedangkan *Naive Bayes* dikenal memiliki performa yang stabil dan efisien pada data numerik dengan ukuran dataset kecil hingga menengah[7].

Beberapa penelitian terdahulu telah membuktikan bahwa algoritma *Decision Tree* dan *Naive Bayes* mampu memberikan hasil yang cukup baik dalam memprediksi prestasi maupun kelulusan siswa[8]. Namun, sebagian besar penelitian tersebut dilakukan pada jenjang sekolah menengah atau menggunakan dataset publik, sehingga belum sepenuhnya merepresentasikan karakteristik akademik siswa sekolah dasar dengan data lokal[9]. Selain itu, beberapa penelitian hanya berfokus pada perbandingan nilai akurasi tanpa menjelaskan secara rinci proses pra-pemrosesan data, parameter algoritma, serta teknik validasi model yang digunakan, sehingga sulit untuk direproduksi dan dievaluasi secara mendalam.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membandingkan performa algoritma *Decision Tree* dan *Naive Bayes* dalam memprediksi tingkat kelulusan siswa sekolah dasar menggunakan data akademik lokal SDN 008 Sungai Dusun[10]. Penelitian ini diharapkan dapat memberikan kontribusi berupa analisis komparatif yang lebih sistematis serta rekomendasi algoritma klasifikasi yang paling sesuai untuk diterapkan sebagai sistem pendukung keputusan dalam proses penentuan kelulusan siswa di sekolah dasar.

2 TINJAUAN PUSTAKA

Data mining merupakan proses penggalian pola dan pengetahuan dari kumpulan data dalam jumlah besar dengan tujuan mendukung pengambilan keputusan secara objektif dan akurat[11]. Dalam bidang pendidikan, data mining banyak dimanfaatkan untuk menganalisis data akademik siswa, seperti nilai mata pelajaran, kehadiran, dan hasil evaluasi belajar[12]. Pemanfaatan data mining memungkinkan pihak sekolah untuk melakukan prediksi terhadap prestasi dan kelulusan siswa berdasarkan data historis yang tersedia, sehingga proses evaluasi tidak hanya bergantung pada penilaian subjektif[13]. Penerapan data mining dalam pendidikan diharapkan mampu meningkatkan kualitas pengambilan keputusan, khususnya dalam menentukan kelulusan siswa secara adil dan berbasis data.

Algoritma *Decision Tree* merupakan salah satu algoritma klasifikasi yang bekerja dengan membentuk struktur pohon keputusan berdasarkan atribut yang paling berpengaruh terhadap kelas target[14]. Setiap simpul dalam pohon merepresentasikan atribut, sedangkan cabang menunjukkan hasil pengujian atribut tersebut. Algoritma ini memiliki keunggulan dalam hal interpretabilitas, karena hasil klasifikasi dapat diterjemahkan ke dalam bentuk aturan yang mudah dipahami.

Dalam bidang pendidikan, Decision Tree sering digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi prestasi dan kelulusan siswa. Namun, kelemahan algoritma ini adalah sensitif terhadap perubahan data dan berpotensi mengalami overfitting apabila data pelatihan tidak dikelola dengan baik[15].

Algoritma Naive Bayes adalah algoritma klasifikasi berbasis probabilitistik yang menggunakan Teorema Bayes dengan asumsi independensi antar atribut. Meskipun asumsi tersebut jarang terpenuhi sepenuhnya pada data nyata, Naive Bayes tetap mampu menghasilkan performa yang baik, khususnya pada data numerik dan dataset berukuran kecil hingga menengah.

Dalam konteks pendidikan, Naive Bayes banyak digunakan karena kesederhanaan model, efisiensi komputasi, serta kemampuannya dalam memprediksi hasil belajar dan kelulusan siswa secara stabil.

Tabel 1 Tinjauan Pustaka

Peneliti	Judul Penelitian	Rangkuman
1. Aldi Sholihin Fauzan (2024)	Analisis Perbandingan Algoritma Decision Tree Dan Naïve Untuk Mengevaluasi Prestasi Belajar Siswa Studi Kasus: Smk Al-Musyawirin	Penelitian ini membandingkan algoritma Decision Tree dan Naïve Bayes dalam mengevaluasi prestasi belajar siswa SMK Al-Musyawirin menggunakan pendekatan data mining. Evaluasi dilakukan dengan teknik cross validation menggunakan metrik akurasi, precision, dan recall. Hasil penelitian menunjukkan bahwa Decision Tree memiliki performa yang lebih baik dibandingkan Naïve Bayes.
2. Yus Susanti, Muhammad Gus Choyyin, Aziz Priyatna, Sri Lestari (2023)	Perbandingan Penerapan Algoritma Decision Tree C4.5 dan Naïve Bayes dalam Analisa Kelulusan Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan	Penelitian ini membandingkan penerapan algoritma Decision Tree C4.5 dan Naïve Bayes dalam memprediksi kelulusan siswa SMK Swadhipa 2 Natar. Data yang digunakan berasal dari transkrip nilai siswa meliputi nilai rata-rata Bahasa Indonesia, Bahasa Inggris, dan Matematika dari semester 1 hingga semester 5 serta riwayat Surat Peringatan (SP). Evaluasi model dilakukan menggunakan confusion matrix dengan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Decision Tree C4.5 memiliki

Peneliti	Judul Penelitian	Rangkuman
		akurasi (80%) dan presisi (82%) yang lebih tinggi dibandingkan Naïve Bayes, sedangkan Naïve Bayes memiliki nilai recall yang lebih tinggi (80%).
3.Abdul Rahman (2023)	Klasifikasi Akademik Menggunakan Metode Decision Tree dan Naive Bayes	Performa Siswa Metode Naive Bayes
		Penelitian ini bertujuan untuk memprediksi performa akademik siswa sejak dini guna memberikan penanganan khusus terkait prestasi belajar. Dataset yang digunakan adalah Experience Application Programming Interface (xAPI) dari sistem e-learning Kalboard 360, mencakup 480 catatan siswa dengan 16 atribut (demografi, akademik, dan perilaku). Pengelompokan data dibagi menjadi tiga kategori: Rendah (L), Sedang (M), dan Tinggi (H). Pengujian dilakukan menggunakan perangkat lunak RapidMiner dengan metode split data (70% training, 30% testing). Hasil penelitian menunjukkan bahwa metode Naive Bayes memiliki akurasi lebih tinggi sebesar 85,97% dibandingkan Decision Tree sebesar 83,89%. Oleh karena itu, Naive Bayes direkomendasikan untuk pembuatan sistem aplikasi pembentukan kelompok belajar siswa.
4.Najiyah Nailil Amani, Martanto, & Umi Hayati (2024)	Penggunaan Algoritma Decision Tree untuk Prediksi Prestasi Siswa di Sekolah Dasar Negeri 3 Bayalangu Kidul	Algoritma Decision Tree
		Penelitian ini mengembangkan model prediksi prestasi belajar siswa menggunakan algoritma Decision Tree dengan bantuan perangkat lunak RapidMiner. Data yang digunakan adalah nilai rapor dari 108 siswa di SDN 3

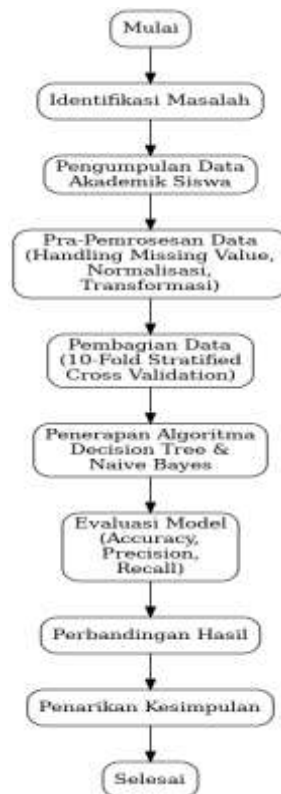
Peneliti	Judul Penelitian	Rangkuman
		<p>Bayalangu Kidul. Metodologi penelitian mengikuti tahapan standar data mining: pengumpulan data, preprocessing (pembersihan data), transformasi, data mining, dan evaluasi. Hasil evaluasi menunjukkan model ini sangat efektif dengan tingkat akurasi mencapai 97,22%. Secara spesifik, prediksi kelulusan memiliki presisi 99,01% dan recall 98,04%, sementara prediksi ketidakkelulusan memiliki presisi 71,43% dan recall 83,33%. Temuan ini memberikan informasi berharga mengenai faktor-faktor yang mempengaruhi kelulusan siswa guna meningkatkan kualitas pendidikan secara keseluruhan.</p>
5. Nurul Elvida & Adika May Sari (2024)	Penerapan Data Mining untuk Prediksi Kelulusan Siswa SD Negeri Sukaresmi Kota Bogor Dengan Alogaritma C4.5	<p>Penelitian ini bertujuan untuk memprediksi kelayakan kelulusan siswa di SDN Sukaresmi Kota Bogor dengan menggunakan Algoritma C4.5 untuk membentuk pohon keputusan (decision tree). Data yang dianalisis meliputi nilai harian, UTS, UAS, dan perilaku dari 55 siswa. Pengolahan data dilakukan menggunakan alat RapidMiner melalui tahapan seleksi, pembersihan, dan transformasi data. Hasil perhitungan menunjukkan bahwa atribut Nilai UAS memiliki Gain tertinggi (0,5456) dan terpilih sebagai Node Akar. Penelitian ini menghasilkan tingkat akurasi sebesar 94,67%. Berdasarkan hasil pengujian, sebanyak 37 siswa diprediksi lulus dan 18 siswa diprediksi tidak lulus.</p>

Berdasarkan hasil kajian terhadap penelitian terdahulu pada Tabel 1, dapat disimpulkan bahwa algoritma Decision Tree dan Naive Bayes telah banyak digunakan dalam berbagai konteks untuk klasifikasi dan prediksi di bidang pendidikan. Penelitian oleh Aldi Sholihin Fauzan (2024) dan Yus Susanti et al. (2023) menunjukkan bahwa kedua algoritma tersebut efektif dalam mengevaluasi prestasi dan kelulusan siswa pada jenjang sekolah menengah, dengan perbedaan performa yang dipengaruhi oleh karakteristik data dan metode evaluasi yang digunakan. Sementara itu, Abdul Rahman (2023) memanfaatkan dataset e-learning berskala besar dengan atribut demografis dan perilaku, sehingga fokus penelitian lebih menekankan pada lingkungan pembelajaran daring.

Meskipun memberikan hasil yang cukup baik, penelitian-penelitian terdahulu tersebut masih memiliki beberapa keterbatasan. Pertama, mayoritas penelitian dilakukan pada jenjang sekolah menengah atau menggunakan dataset publik, sehingga belum sepenuhnya merepresentasikan kondisi akademik siswa sekolah dasar dengan karakteristik data yang bersifat lokal. Kedua, beberapa penelitian hanya menitikberatkan pada nilai akurasi tanpa menjelaskan secara rinci tahapan pra-pemrosesan data dan skema validasi model, sehingga penelitian sulit direproduksi dan dievaluasi secara mendalam. Ketiga, penelitian pada konteks sekolah dasar umumnya hanya menggunakan satu algoritma klasifikasi, seperti C4.5, tanpa melakukan perbandingan langsung dengan algoritma probabilistik seperti Naive Bayes.

Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan melakukan perbandingan kinerja algoritma Decision Tree dan Naive Bayes dalam memprediksi tingkat kelulusan siswa sekolah dasar menggunakan data akademik lokal SDN 008 Sungai Dusun. Penelitian ini tidak hanya membandingkan performa kedua algoritma, tetapi juga menerapkan tahapan pra-pemrosesan data dan teknik validasi model yang jelas dan sistematis. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis berupa rekomendasi metode klasifikasi yang paling sesuai untuk diterapkan sebagai sistem pendukung keputusan kelulusan siswa sekolah dasar berbasis data akademik dan konteks lokal sekolah.

3 METODE PENELITIAN



Gambar 1 menunjukkan tahapan penelitian yang dilakukan secara sistematis mulai dari identifikasi masalah hingga penarikan kesimpulan.

Identifikasi Masalah

Tahap awal penelitian dimulai dengan identifikasi masalah yang terjadi di SDN 008 Sungai Dusun. Permasalahan utama yang ditemukan adalah proses penentuan kelulusan siswa yang masih dilakukan secara manual dan cenderung subjektif, karena hanya berdasarkan penilaian guru tanpa dukungan analisis data secara sistematis. Kondisi ini berpotensi menimbulkan ketidakkonsistenan dalam pengambilan keputusan. Oleh karena itu, diperlukan suatu pendekatan berbasis data mining untuk membantu memprediksi kelulusan siswa secara lebih objektif.

Pengumpulan Data

Setelah permasalahan teridentifikasi, tahap selanjutnya adalah pengumpulan data. Data yang digunakan dalam penelitian ini berupa data akademik siswa SDN 008 Sungai Dusun yang diperoleh dari arsip sekolah. Dataset terdiri dari 212 data siswa dengan atribut nilai mata pelajaran, yaitu PAI(Agama), PPKN(Pendidikan Pancasila dan Kewarganegaraan), Bahasa Indonesia, Matematika, IPA(Ilmu Pengatahuan Alam), IPS(Ilmu Pengatahuan Sosial), SBDP(Seni Budaya dan Prakarya), PJOK(Pendidikan Jasmani, Olahraga, dan Kesehatan), ARMEL (Arab Melayu), BMR (Budaya Melayu Riau), Bahasa Inggris, serta satu atribut target yaitu Kelulusan. Data ini dipilih karena dianggap merepresentasikan kemampuan akademik siswa secara menyeluruh.

Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik sebelum diterapkan pada algoritma klasifikasi. Tahapan ini sangat penting karena kualitas data akan berpengaruh langsung terhadap hasil prediksi model.

1. Handling Missing Value

Pada tahap ini, seluruh atribut dalam dataset diperiksa untuk mengetahui apakah terdapat nilai yang hilang (*missing value*). Pemeriksaan dilakukan menggunakan AI Studio dengan melihat kelengkapan data pada setiap kolom atribut. Berdasarkan hasil pemeriksaan, diketahui bahwa dataset yang digunakan relatif lengkap dan tidak ditemukan *missing value* yang signifikan. Oleh karena itu, pada penelitian ini tidak dilakukan proses imputasi data agar tidak mengubah distribusi data asli.

2. Normalisasi Data

Selanjutnya dilakukan proses normalisasi data terhadap seluruh atribut numerik menggunakan metode *Min-Max Normalization*. Metode ini mengubah nilai data ke dalam rentang 0–1. Normalisasi dilakukan untuk menyamakan skala nilai antar atribut sehingga tidak ada atribut tertentu yang mendominasi proses pembelajaran algoritma akibat perbedaan rentang nilai.

3. Transformasi Data

Pada tahap transformasi data, atribut target Kelulusan dikonversi menjadi data kategorikal dengan dua kelas, yaitu Lulus dan Tidak Lulus. Transformasi ini dilakukan agar data sesuai dengan kebutuhan algoritma klasifikasi yang digunakan dalam penelitian.

Pembagian Data (Cross Validation)

Setelah data melalui tahap pra-pemrosesan, dilakukan pembagian data menggunakan teknik *10-Fold Stratified Cross Validation*. Dataset dibagi menjadi 10 bagian (*fold*) dengan proporsi kelas Lulus dan Tidak Lulus yang tetap terjaga pada setiap *fold*. Pada setiap iterasi, satu *fold* digunakan sebagai data pengujian dan sembilan *fold* lainnya digunakan sebagai data pelatihan. Teknik ini dipilih untuk menghasilkan evaluasi model yang lebih objektif dan mengurangi bias akibat pembagian data secara acak.

Penerapan Algoritma Klasifikasi

Tahap berikutnya adalah penerapan algoritma klasifikasi menggunakan AI Studio sebagai alat bantu pengolahan data. Dua algoritma yang digunakan dalam penelitian ini adalah *Decision Tree* dan *Naïve Bayes*.

1. Decision Tree

Algoritma *Decision Tree* digunakan untuk membentuk model klasifikasi dalam bentuk struktur pohon keputusan. Pemilihan atribut pada setiap node dilakukan menggunakan kriteria *Entropy (Information Gain)*, yaitu dengan memilih atribut yang mampu mengurangi tingkat ketidakpastian

data secara maksimal. Kriteria ini dipilih karena efektif dalam menangani data pendidikan yang memiliki variasi nilai akademik dan menghasilkan model yang mudah dipahami. Proses pembentukan pohon keputusan dilakukan menggunakan AI Studio dengan parameter bawaan (*default*), tanpa pembatasan kedalaman pohon.

2. Naive bayes

Algoritma Naive Bayes yang digunakan dalam penelitian ini adalah Gaussian Naive Bayes, karena seluruh atribut prediktor berupa data numerik kontinu berupa nilai akademik siswa. Algoritma ini bekerja berdasarkan *Teorema Bayes* dengan mengasumsikan bahwa setiap atribut bersifat independen satu sama lain. Pada Gaussian Naive Bayes, probabilitas setiap kelas dihitung berdasarkan nilai rata-rata (*mean*) dan standar deviasi dari masing-masing atribut pada setiap kelas dengan menggunakan distribusi normal (*Gaussian*). Proses klasifikasi dilakukan menggunakan AI Studio dengan parameter bawaan (*default*), tanpa modifikasi tambahan pada model.

Evaluasi Model

Evaluasi performa model dilakukan menggunakan confusion matrix dengan tiga metrik utama: *Accuracy*, *Precision*, dan *Recall*. Nilai evaluasi diperoleh dari rata-rata hasil 10 fold pengujian untuk masing-masing algoritma, sehingga hasil yang diperoleh lebih stabil dan dapat dipertanggungjawabkan.

Perbandingan Hasil

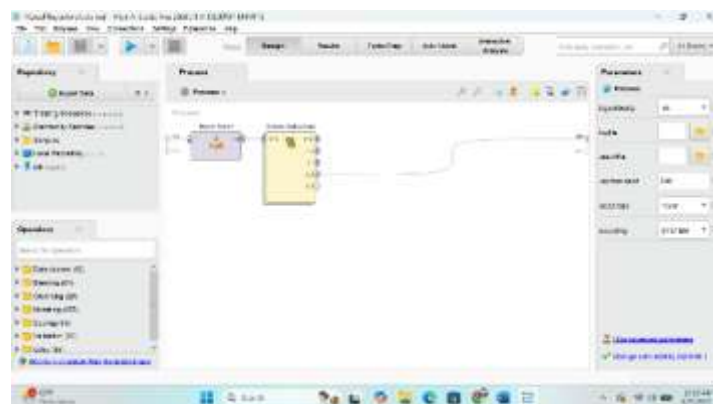
Tahap perbandingan hasil bertujuan untuk mengevaluasi dan membandingkan performa algoritma Decision Tree dan Naive Bayes dalam memprediksi tingkat kelulusan siswa. Perbandingan dilakukan berdasarkan metrik evaluasi accuracy, precision, dan recall yang dihasilkan melalui penerapan teknik 10-Fold Stratified Cross Validation.

Penarikan kesimpulan

Tahap penarikan kesimpulan dilakukan berdasarkan hasil perbandingan performa algoritma Decision Tree dan Naive Bayes yang diperoleh pada tahap sebelumnya. Kesimpulan penelitian ditentukan dengan mempertimbangkan nilai metrik evaluasi, yaitu accuracy, precision, dan recall, dari masing-masing algoritma secara keseluruhan. Kesimpulan ini digunakan untuk menjawab tujuan penelitian serta memberikan rekomendasi algoritma klasifikasi yang paling sesuai untuk diterapkan dalam sistem pendukung keputusan kelulusan siswa di SDN 008 Sungai Dusun.

4 HASIL DAN PEMBAHASAN

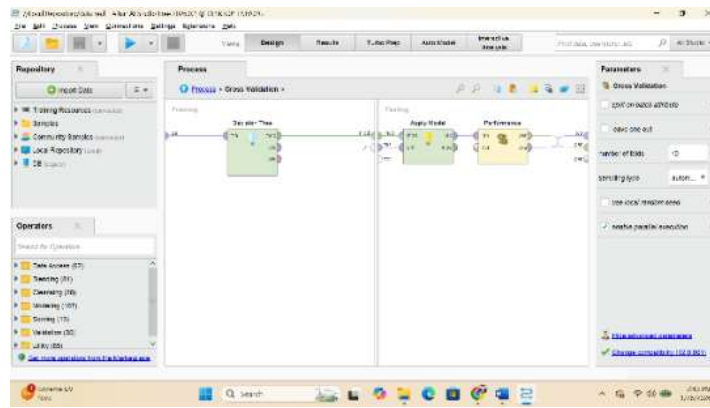
4.1 Hasil



Gambar 2 Proses Pengujian Model Klasifikasi Menggunakan Cross Validation di Altair AI Studio

Gambar 2 menunjukkan proses pengujian model klasifikasi yang dilakukan menggunakan aplikasi Altair AI Studio. Data akademik siswa dibaca dari file Excel menggunakan operator *Read Excel*, kemudian diproses menggunakan operator *Cross Validation* untuk melakukan evaluasi model. Pada tahap *Cross Validation*, data dibagi menjadi beberapa bagian untuk keperluan pelatihan dan pengujian model. Proses ini bertujuan untuk memperoleh hasil evaluasi performa algoritma klasifikasi yang lebih objektif. Hasil pengujian selanjutnya digunakan sebagai dasar perhitungan nilai metrik evaluasi, yaitu accuracy, precision, dan recall.

Rahmathidayatullah, Perbandingan Metode Algoritma Klasifikasi (Decision Tree Dan Naive Bayes) Untuk Memprediksi Akurasi Tingkat Kelulusan Siswa Di Sekolah SDN 008 Sungai Dusun



Gambar 3 Penerapan Algoritma Decision Tree pada Proses Cross Validation

Gambar 3 menunjukkan penerapan algoritma Decision Tree pada tahap pelatihan dan pengujian model dalam skema Cross Validation. Pada bagian pelatihan (*training*), algoritma Decision Tree digunakan untuk membentuk model klasifikasi berdasarkan data latih yang tersedia. Model yang terbentuk selanjutnya diterapkan pada data uji (*testing*) melalui proses Apply Model.

Evaluasi performa model Decision Tree dilakukan menggunakan operator Performance untuk memperoleh nilai metrik evaluasi, yaitu accuracy, precision, dan recall. Nilai evaluasi yang dihasilkan merupakan rata-rata dari seluruh fold pengujian, sehingga mencerminkan performa model secara keseluruhan.

Hasil Pengujian Algoritma Decision Tree



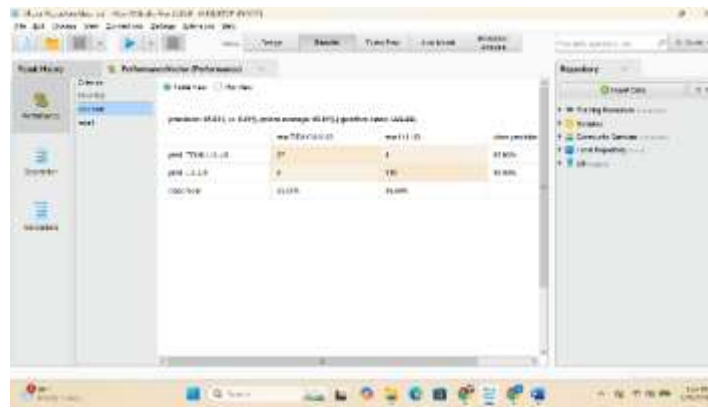
Gambar 4 hasil Pengujian Algoritma Decision Tree Accuracy

Berdasarkan hasil pengujian menggunakan teknik 10-Fold Stratified Cross Validation, algoritma Decision Tree memperoleh nilai accuracy sebesar $95,26\% \pm 5,02\%$, dengan nilai micro average sebesar $95,28\%$. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan data kelulusan siswa dengan tingkat ketepatan yang sangat tinggi, yaitu sekitar 95%. Simbol $\pm 5,02\%$ menggambarkan variasi performa pada setiap fold selama proses validasi. Untuk memahami lebih jauh performa model, berikut ringkasan confusion matrix hasil pengujian:

Tabel 1. Confusion matrix

Keterangan	Jumlah
Lulus → diprediksi Lulus	115
Lulus → diprediksi Tidak Lulus	4
Tidak Lulus → diprediksi Tidak Lulus	87
Tidak Lulus → diprediksi Lulus	6

Hasil tersebut menunjukkan bahwa model berhasil mengklasifikasikan 115 siswa Lulus dan 87 siswa Tidak Lulus secara benar, dengan jumlah kesalahan prediksi yang relatif kecil (4 dan 6 data). Temuan ini mengindikasikan bahwa algoritma Decision Tree memiliki kemampuan prediksi yang sangat baik terhadap data akademik siswa pada SDN 008 Sungai Dusun.



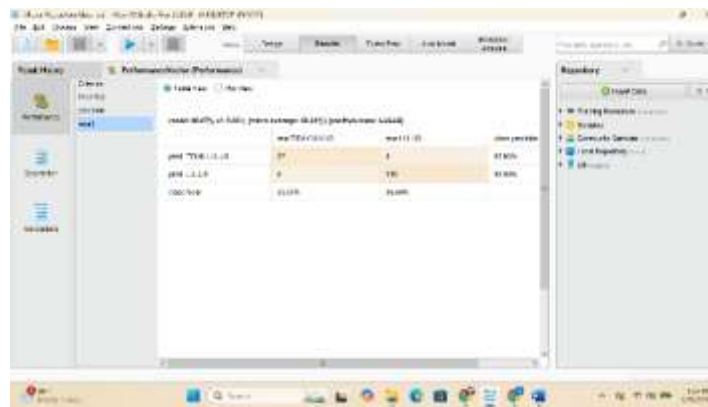
Gambar 5 hasil Pengujian Algoritma Decision Tree Precision

Berdasarkan hasil pengujian menggunakan teknik 10-Fold Stratified Cross Validation, algoritma Decision Tree menghasilkan nilai precision sebesar 95,63% ± 6,64%, dengan nilai micro average sebesar 95,04%. Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengidentifikasi prediksi positif yang benar, khususnya untuk kelas LULUS yang ditetapkan sebagai kelas positif Rincian nilai precision per kelas adalah sebagai berikut:

Tabel 2. Nilai precision per kelas

Prediksi	True TIDAK LULUS	True LULUS	Precision
TIDAK LULUS	87	4	95,60%
LULUS	6	115	95,04%

Nilai precision kelas TIDAK LULUS sebesar 95,60% menunjukkan bahwa sebagian besar data yang diprediksi sebagai “Tidak Lulus” adalah benar. Sementara itu, precision kelas LULUS sebesar 95,04% menunjukkan bahwa model juga sangat akurat dalam memprediksi siswa yang benar-benar lulus. Secara keseluruhan, tingginya nilai precision pada kedua kelas mengindikasikan bahwa algoritma Decision Tree mampu meminimalkan kesalahan prediksi positif (false positive), sehingga sangat efektif digunakan untuk memprediksi kelulusan siswa berdasarkan data akademik di SDN 008 Sungai Dusun.



Gambar 6 hasil Pengujian Algoritma Decision Tree Recall

Berdasarkan hasil pengujian menggunakan teknik 10-Fold Stratified Cross Validation, algoritma Decision Tree menghasilkan nilai recall sebesar 96,67% ± 5,83%, dengan nilai micro average sebesar 96,64%. Nilai recall yang tinggi ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi data positif dengan benar, khususnya untuk kelas LULUS yang dijadikan kelas positif. Rincian recall per kelas ditunjukkan pada tabel berikut:

Tabel 3. Recall per kelas

Prediksi	True TIDAK LULUS	True LULUS	Recall
TIDAK LULUS	87	6	93,55%
LULUS	4	115	96,64%

Nilai *recall* kelas TIDAK LULUS sebesar 93,55% menunjukkan bahwa sebagian besar siswa yang benar-benar tidak lulus berhasil teridentifikasi dengan tepat oleh model. Sementara itu, *recall* kelas LULUS sebesar 96,64% mengindikasikan bahwa hampir seluruh siswa yang benar-benar lulus dapat dikenali dengan sangat akurat. Secara keseluruhan, nilai *recall* yang tinggi pada kedua kelas menunjukkan bahwa algoritma Decision Tree sangat efektif dalam meminimalkan kesalahan *false negative* dan mampu menangkap mayoritas data aktual, sehingga layak digunakan sebagai metode prediksi kelulusan siswa pada SDN 008 Sungai Dusun.

Hasil Pengujian Algoritma Naïve Bayes



Gambar 7 Hasil Pengujian Algoritma Naïve Bayes accuracy

Berdasarkan hasil pengujian menggunakan 10-Fold Stratified Cross Validation, algoritma Naive Bayes memperoleh nilai *accuracy* sebesar 95,76% ± 6,09%, dengan nilai *micro average* sebesar 95,75%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan data kelulusan siswa dengan akurasi yang tinggi dan performa yang relatif stabil pada setiap *fold*. Adapun ringkasan confusion matrix adalah sebagai berikut:

Tabel 4. Confusion matrix

Keterangan	Jumlah
LULUS → diprediksi LULUS	113
LULUS → diprediksi TIDAK LULUS	3
TIDAK LULUS → diprediksi TIDAK LULUS	90
TIDAK LULUS → diprediksi LULUS	6

Model Naive Bayes berhasil mengklasifikasikan 113 data LULUS dan 90 data TIDAK LULUS secara benar. Kesalahan prediksi pada kedua kelas relatif kecil, yaitu hanya 3 data Lulus dan 6 data Tidak Lulus yang salah diklasifikasikan. Hal ini menunjukkan bahwa model mampu mengenali pola data akademik siswa dengan cukup baik.



Gambar 8 hasil Pengujian Algoritma Naïve Bayes Precision

Hasil pengujian menunjukkan bahwa Naive Bayes menghasilkan nilai *precision* sebesar 98,00% ± 6,32%, dengan nilai *micro average* sebesar 97,41%. Nilai *precision* ini menunjukkan bahwa prediksi

positif yang diberikan model sangat akurat, terutama pada kelas LULUS yang dijadikan kelas positif. Rincian precision untuk tiap kelas adalah sebagai berikut:

Tabel 5. Precision untuk tiap kelas

Prediksi	True TIDAK LULUS	True LULUS	Precision
TIDAK LULUS	90	6	93,75%
LULUS	3	113	97,41%

Precision kelas LULUS sebesar 97,41% menunjukkan bahwa hampir seluruh prediksi “Lulus” adalah benar. Precision kelas TIDAK LULUS sebesar 93,75% juga menunjukkan tingkat ketepatan yang baik dalam mendeteksi siswa yang tidak lulus. Nilai precision yang tinggi ini mengindikasikan kemampuan Naive Bayes untuk meminimalkan kesalahan prediksi positif (*false positive*).



Gambar 9 hasil Pengujian Algoritma Naive Bayes Recall

Nilai recall yang dihasilkan oleh algoritma Naive Bayes adalah 95,00% ± 8,96%, dengan *micro average* sebesar 94,96%. Recall yang tinggi ini menunjukkan bahwa model mampu mendeteksi sebagian besar data yang benar-benar positif dengan baik.

Rincian recall untuk tiap kelas adalah sebagai berikut:

Tabel 6. Recall untuk tiap kelas

Prediksi	True TIDAK LULUS	True LULUS	Recall
TIDAK LULUS	90	3	96,77%
LULUS	6	113	94,96%

Recall kelas TIDAK LULUS sebesar 96,77% menunjukkan bahwa model dapat mengenali siswa yang benar-benar tidak lulus dengan sangat baik. Sementara itu, recall kelas LULUS sebesar 94,96% menunjukkan bahwa sebagian besar siswa yang benar-benar lulus berhasil dideteksi oleh model.

Secara keseluruhan, nilai recall yang tinggi menggambarkan bahwa Naive Bayes mampu meminimalkan kesalahan *false negative*, sehingga efektif digunakan untuk mendeteksi siswa yang benar-benar lulus maupun tidak lulus.

4.2 Pembahasan

Bagian ini membahas hasil perbandingan kinerja algoritma Decision Tree dan Naive Bayes yang digunakan untuk memprediksi kelulusan siswa SDN 008 Sungai Dusun. Pembahasan dilakukan berdasarkan nilai akurasi, precision, dan recall yang diperoleh dari pengujian menggunakan teknik 10-Fold Stratified Cross Validation.

interpretasi akurasi

Hasil pengujian menunjukkan bahwa kedua algoritma memiliki akurasi yang sangat tinggi, yaitu: Decision Tree: 95,26% dan Naive Bayes: 95,76%

Perbedaan akurasi antara kedua algoritma sangat kecil (0,50%). Hal ini menunjukkan bahwa pola nilai akademik siswa memiliki distribusi yang cukup teratur sehingga kedua model dapat mempelajarinya dengan baik. Decision Tree mampu mencapai akurasi tinggi karena algoritma ini membagi data berdasarkan atribut yang memiliki informasi paling signifikan, sehingga model

sangat efektif memisahkan antara siswa yang lulus dan tidak lulus. Sementara itu, Naive Bayes memberikan performa stabil karena pendekatan probabilitiknya efektif meskipun terdapat variasi pada nilai siswa.

Interpretasi Precision

Nilai precision dari kedua model adalah: Decision Tree: 95,63% Naive Bayes: 98,00%. Precision Naive Bayes lebih tinggi dibandingkan Decision Tree. Artinya, ketika model memprediksi seorang siswa lulus, prediksi tersebut sangat jarang salah (false positive rendah). Ini menunjukkan keunggulan Naive Bayes dalam menghasilkan prediksi positif yang benar, terutama karena algoritma ini menghitung probabilitas setiap kelas secara detail menggunakan distribusi Gaussian. Sebaliknya, Decision Tree masih menghasilkan sedikit kesalahan prediksi pada kelas “Lulus”, namun tetap berada dalam batas sangat tinggi (>95%).

Interpretasi Recall

Decision Tree: 96,67% dan Naive Bayes: 95,00%. Recall Decision Tree lebih tinggi dibandingkan Naive Bayes. Ini berarti Decision Tree lebih mampu menangkap siswa yang benar-benar lulus maupun benar-benar tidak lulus. Tingginya recall disebabkan oleh struktur pohon keputusan yang mampu membedakan pola nilai akademik secara detail pada setiap simpul keputusan. Naive Bayes sedikit lebih rendah recall-nya karena algoritma ini mengasumsikan independensi antar atribut, padahal nilai-nilai pelajaran siswa kemungkinan tidak sepenuhnya independen (misalnya siswa dengan nilai Matematika tinggi cenderung memiliki nilai IPA tinggi). Hal ini menyebabkan beberapa siswa yang sebenarnya lulus tidak terdeteksi sepenuhnya oleh model.

Perbandingan Kinerja Kedua Algoritma

Sebelum membandingkan kedua algoritma, kita perlu memahami dulu karakter masing-masing. Decision Tree unggul dalam mendeteksi pola karena membagi data berdasarkan atribut paling informatif, sedangkan Naive Bayes unggul pada prediksi positif karena menggunakan pendekatan probabilitas. Perbedaan mekanisme inilah yang membuat nilai akurasi, precision, dan recall keduanya berbeda. Setelah memahami karakteristik ini, barulah dilakukan perbandingan kinerja kedua algoritma.

Tabel 7. Perbandingan Kinerja Kedua Algoritma

Aspek	Decision Tree	Naive Bayes	Algoritma Terbaik
Akurasi	95,26%	95,76%	Naive Bayes (sedikit lebih tinggi)
Precision	95,63%	98,00%	Naive Bayes
Recall	96,67%	95,00%	Decision Tree
Interpretasi Model	Sangat mudah dipahami	Tidak polanya	terlihat Decision Tree

Tidak ada algoritma yang benar-benar unggul di semua aspek. Namun: Naive Bayes unggul dalam precision dan akurasi. Decision Tree unggul dalam recall dan interpretabilitas. Jika sekolah menginginkan model yang mudah dipahami oleh guru, maka Decision Tree lebih cocok. Jika sekolah menginginkan prediksi kelulusan yang sangat minim kesalahan positif, maka Naive Bayes dapat menjadi pilihan.

Hasil penelitian ini selaras dengan penelitian terdahulu: Aldi Sholihin (2024) menemukan bahwa Decision Tree unggul dibanding Naive Bayes pada evaluasi prestasi siswa SMK. Yus Susanti (2023) menemukan Decision Tree memiliki akurasi lebih tinggi, namun Naive Bayes unggul recall-nya. Abdul Rahman (2023) menemukan Naive Bayes lebih stabil pada dataset besar.

5 KESIMPULAN

Penelitian ini bertujuan untuk membandingkan performa algoritma Decision Tree dan Naive Bayes dalam memprediksi tingkat kelulusan siswa di SDN 008 Sungai Dusun menggunakan data nilai akademik. Berdasarkan hasil pengujian menggunakan teknik 10-Fold Stratified Cross Validation, kedua algoritma menunjukkan performa yang sangat baik dengan akurasi di atas 95%.

Rahmathidayatullah, Perbandingan Metode Algoritma Klasifikasi (Decision Tree Dan Naive Bayes) Untuk Memprediksi Akurasi Tingkat Kelulusan Siswa Di Sekolah SDN 008 Sungai Dusun

Algoritma Decision Tree memperoleh akurasi sebesar 95,26%, precision 95,63%, dan recall 96,67%. Sementara itu, algoritma Naive Bayes menunjukkan akurasi 95,76%, precision 98,00%, dan recall 95,00%. Hasil ini menunjukkan bahwa Naive Bayes unggul dalam akurasi dan precision, sedangkan Decision Tree unggul pada nilai recall dan kemudahan interpretasi model.

Secara keseluruhan, kedua algoritma layak diterapkan sebagai metode prediksi kelulusan siswa. Namun, algoritma Decision Tree lebih direkomendasikan karena model yang dihasilkan mudah dipahami oleh pihak sekolah dan memberikan kemampuan deteksi yang lebih baik terhadap siswa yang benar-benar lulus maupun tidak lulus. Penelitian ini diharapkan dapat mendukung pengembangan sistem pendukung keputusan kelulusan siswa berbasis data mining sehingga proses evaluasi menjadi lebih objektif, konsisten, dan data-driven.

REFERENSI

- [1] S. Kota, B. Dengan, C. Alogaritma, N. Elvida, A. M. Sari, and U. Lampung, "Bulletin of Community Engagement," vol. 4, no. 3, 2024.
- [2] N. N. Amani, U. Hayati, T. Informatika, M. Informatika, K. Cirebon, and A. D. Tree, "PENGUNAAN ALGORITMA DECISION TREE UNTUK PREDIKSI PRESTASI," vol. 8, no. 1, pp. 473–479, 2024.
- [3] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," vol. 9, pp. 22–31, 2023.
- [4] J. Simada, S. Informasi, B. Data, and V. No, "PERBANDINGAN PENERAPAN ALGORITMA DECISION TREE C . 45 DAN NAÏVE BAYES DALAM ANALISA KELULUSAN SISWA PADA SMK SWADHIPA 2 NATAR," vol. 6, no. 2, pp. 117–123, 2023.
- [5] A. Sholihin Fauzan, A. Irma Purnama Sari, and I. Ali, "Analisis Perbandingan Algoritma Decision Tree Dan Naïve Untuk Mengevaluasi Prestasi Belajar Siswa," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 741–747, 2024, doi: 10.36040/jati.v8i1.8403.
- [6] M. D. Noverta Effendi*1, Witrihan Ramadhani2, Fitri Farida3, "Jurnal Computer Science and Information Technology (CoSciTech) things," J. Comput. Sci. Inf. Technol., vol. 5, no. 2, pp. 358–366, 2024.
- [7] U. Pamulang, U. Pamulang, and D. M. Decision, "Scientia Sacra : Jurnal Sains , Teknologi dan Masyarakat Analisis Perbandingan Algoritma Data Mining Metode Decision Tree C4 . 5 Dengan Naïve Bayes dalam Penjurusan Siswa (Studi Kasus MAN 1 Kota Tangerang Selatan)," vol. 2, no. 1, pp. 44–61, 2022.
- [8] J. Sisingamangaraja, J. Karangkajen, and K. Mergangsan, "PERBANDINGAN ALGORITMA KLASIFIKASI K-NN DENGAN VARIASI JARAK , NAIVE BAYES , LOGISTIC REGRESSION , DAN DECISION TREE," vol. 9, no. 4, pp. 7057–7064, 2025.
- [9] A. Supriyadi, "Perbandingan Algoritma Naive Bayes dan Decision Tree (C4 . 5) dalam Klasifikasi Dosen Berprestasi," vol. 7, no. 1, pp. 39–49, 2023.
- [10] N. K. Juniari, N. Luh, G. Pivin, and G. H. Setiawan, "Analisis Perbandingan Metode Decision Tree Dan Naive Bayes Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa," vol. 2, no. 1, pp. 691–696, 2025.
- [11] K. Berdasarkan, D. Mahasiswa, U. Lancang, A. Zamsuri, and D. Muladi, "Perbandingan Akurasi Algoritma C4 . 5 Dan Naïve Bayes Terhadap Klasifikasi Kelulusan Berdasarkan Data Mahasiswa Universitas Lancang Kuning," no. x, pp. 398–406, 2015.
- [12] L. L. N. L. Dnduld, Z. D. Q. UlvqdgI, S. Lql, and G. Shpdvdudq, "\$ QDOLVVLV 3HUEDQGLQJDQ \$ OJRULWPD ' HFLVLRQ 7UHH 1DwYH % D \ HV GDQ N 1H DUHVV".
- [13] A. Fakh, M. A. Hamzami, M. R. Hadianto, and N. I. Shafira, "Perbandingan Akurasi Algoritma C4 . 5 dan K-NN Untuk Prediksi," vol. 3, no. c, pp. 18–25, 2025.

Rahmathidayatullah, Perbandingan Metode Algoritma Klasifikasi (Decision Tree Dan Naive Bayes) Untuk Memprediksi Akurasi Tingkat Kelulusan Siswa Di Sekolah SDN 008 Sungai Dusun

- [14] M. Zainuddin, “Perbandingan 4 Algoritma Berbasis Particle Swarm Optimization (pso) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa,” vol. 13, no. 1, pp. 1–12, 2019.
- [15] G. Divva, M. Zulma, and N. Chamidah, “Perbandingan Metode Klasifikasi Naive Bayes , Decision Tree Dan K- Nearest Neighbor Pada Data Log Firewall,” no. April, pp. 679–688, 2021.