

Perbandingan Algoritma K-Nearest Neighbor dan Naive Bayes Untuk Klasifikasi Prestasi Akademik Pada Dataset Student Performance

Alvananta Farhan Maulana^{*1}, Zachol Fatah², Ahmad Homaidi³

^{1,2,3} Universitas Ibrahimy

Email: ¹alvananta1007@gmail.com, ²zacholfatah@gmail.com, ³Ahmadhomaidi@ibrahimiy.ac.id

^{*}alvananta1007@gmail.com

Abstrak

Prestasi akademik siswa merupakan indikator penting dalam proses pendidikan yang dapat digunakan untuk mengevaluasi keberhasilan pembelajaran. Namun, proses analisis prestasi akademik masih sering dilakukan secara manual sehingga kurang efektif dalam mengolah data dalam jumlah besar. Penelitian ini bertujuan membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi prestasi akademik menggunakan dataset Student Performance. Metode penelitian menggunakan Knowledge Discovery in Database (KDD) yang meliputi pengumpulan data, preprocessing, klasifikasi, evaluasi model, dan interpretasi hasil. Dataset terdiri dari sekitar 1000 data siswa dengan atribut gender, parental level of education, lunch, test preparation course, math score, reading score, dan writing score. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memperoleh performa terbaik dengan accuracy 92,75%, precision 92,34%, recall 92,80%, dan classification error 7,25%. Sedangkan algoritma KNN memperoleh accuracy 91,12%. Hasil penelitian juga diimplementasikan dalam aplikasi berbasis web menggunakan framework Flask untuk mempermudah proses prediksi prestasi akademik siswa.

Kata kunci: Data Mining, Naive Bayes, KNN, Prestasi Akademik, Flask

Abstract

Student academic achievement is an important indicator in the educational process that can be used to evaluate learning success. However, the process of analyzing academic achievement is still often done manually, making it less effective in processing large amounts of data. This study aims to compare the performance of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms in classifying academic achievement using the Student Performance dataset. The research method uses Knowledge Discovery in Database (KDD) which includes data collection, preprocessing, classification, model evaluation, and interpretation of results. The dataset consists of approximately 1000 student data with attributes of gender, parental level of education, lunch, test preparation course, math score, reading score, and writing score. Evaluation was carried out using accuracy, precision, recall, and confusion matrix. The results showed that the Naive Bayes algorithm obtained the best performance with an accuracy of 92.75%, precision of 92.34%, recall of 92.80%, and classification error of 7.25%. Meanwhile, the KNN algorithm obtained an accuracy of 91.12%. The research results were also implemented in a web-based application using the Flask framework to facilitate the process of predicting student academic achievement.

Keywords: Data Mining, Naive Bayes, KNN, Academic Achievement, Flask

I. PENDAHULUAN

Transformasi digital dalam bidang pendidikan telah mendorong meningkatnya penggunaan teknologi untuk membantu proses pengolahan dan analisis data akademik siswa. Aktivitas pembelajaran, evaluasi hasil belajar, hingga pengelolaan informasi akademik kini menghasilkan data dalam jumlah besar yang dapat dimanfaatkan sebagai sumber informasi penting untuk mendukung pengambilan keputusan di lingkungan pendidikan[1]. Pemanfaatan data akademik secara optimal dapat membantu institusi pendidikan dalam memahami pola prestasi siswa, mengidentifikasi faktor-faktor yang memengaruhi hasil belajar, serta memberikan evaluasi yang lebih objektif dan efisien[2].

Prestasi akademik siswa merupakan salah satu indikator utama keberhasilan proses pendidikan. Tingkat prestasi siswa umumnya diukur berdasarkan hasil evaluasi pembelajaran seperti nilai ujian, tugas, maupun rata-rata capaian akademik pada periode tertentu. Prestasi akademik tidak hanya menjadi tolak ukur kemampuan siswa dalam memahami materi pembelajaran, tetapi juga menjadi bahan evaluasi bagi sekolah dalam meningkatkan kualitas proses belajar mengajar. Namun, dalam praktiknya masih banyak institusi pendidikan yang mengalami kesulitan dalam melakukan analisis prestasi akademik secara cepat

dan tepat karena proses evaluasi masih dilakukan secara manual dan belum memanfaatkan teknik analisis data secara optimal[3].

Perkembangan teknologi data mining dan machine learning memberikan solusi dalam mengatasi permasalahan tersebut. Data mining merupakan proses penggalian informasi dari sekumpulan data dalam jumlah besar untuk menemukan pola, hubungan, dan pengetahuan baru yang bermanfaat dalam pengambilan keputusan. Salah satu teknik yang banyak digunakan dalam data mining adalah klasifikasi. Teknik klasifikasi digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang diperoleh dari data sebelumnya. Dalam konteks pendidikan, klasifikasi dapat digunakan untuk memprediksi atau menentukan tingkat prestasi akademik siswa berdasarkan atribut-atribut yang dimiliki[4].

Algoritma klasifikasi yang sering digunakan dalam penelitian data mining diantaranya adalah K-Nearest Neighbor (KNN) dan Naive Bayes. Algoritma K-Nearest Neighbor merupakan metode supervised learning yang bekerja berdasarkan kedekatan jarak antar data untuk menentukan kelas suatu objek[5]. KNN memiliki kelebihan dalam mengenali pola data dengan baik dan cukup efektif digunakan pada data yang memiliki kemiripan karakteristik. Sementara itu, Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen. Algoritma ini dikenal memiliki proses komputasi yang cepat dan mampu memberikan performa yang baik pada data kategorikal maupun numerik[3].

Dataset Student Performance merupakan salah satu dataset pendidikan yang banyak digunakan dalam penelitian klasifikasi prestasi akademik. Dataset ini berisi informasi mengenai latar belakang siswa, tingkat pendidikan orang tua, jenis makan siang, kursus persiapan ujian, serta nilai akademik siswa seperti math score, reading score, dan writing score. Data tersebut dapat dimanfaatkan untuk melakukan klasifikasi tingkat prestasi akademik siswa berdasarkan kategori tertentu, seperti rendah, sedang, dan tinggi.

Beberapa penelitian sebelumnya telah melakukan penerapan algoritma klasifikasi pada data akademik siswa. Penelitian oleh Asih Winanti dan Chusnul Khatimah membandingkan metode Naive Bayes dan K-Nearest Neighbor dalam memprediksi prestasi siswa dan menunjukkan bahwa kedua algoritma memiliki performa yang baik dalam klasifikasi data akademik. Penelitian lain oleh Abdul Rahman juga menunjukkan bahwa algoritma Naive Bayes mampu menghasilkan tingkat akurasi yang cukup tinggi dalam klasifikasi performa akademik siswa[6]. Selain itu, penelitian Afif Muhaimin dkk. membuktikan bahwa metode KNN dapat digunakan untuk mengklasifikasikan prestasi akademik siswa dengan hasil evaluasi yang baik.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk membandingkan performa algoritma K-Nearest Neighbor dan Naive Bayes dalam klasifikasi prestasi akademik pada dataset Student Performance. Penelitian ini diharapkan dapat mengetahui algoritma yang memiliki tingkat akurasi dan performa terbaik dalam melakukan klasifikasi prestasi akademik siswa, sehingga hasil penelitian dapat menjadi referensi dalam pengembangan sistem pendukung keputusan di bidang pendidikan[7].

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif untuk membandingkan performa algoritma klasifikasi, yaitu K-Nearest Neighbor (K-NN) dan Naive Bayes, dalam mengklasifikasikan prestasi akademik siswa. Penelitian ini bertujuan untuk menganalisis perbedaan performa dalam melakukan klasifikasi berdasarkan parameter data yang sudah ditentukan. kedua algoritma diuji dan dibandingkan kinerjanya dalam mengklasifikasikan data student performance.

Metode Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi dengan pendekatan supervised learning yang digunakan untuk proses prediksi maupun klasifikasi data. Prinsip kerja algoritma KNN adalah mencari sejumlah tetangga terdekat dari suatu data berdasarkan nilai kedekatan jarak terhadap data pelatifa[8].

Algoritma KNN mengklasifikasikan objek baru berdasarkan mayoritas kelas dari tetangga terdekatnya. Salah satu kelebihan metode KNN adalah memiliki tingkat akurasi yang cukup baik dan tidak memerlukan asumsi khusus terhadap data. Pada penelitian ini, perhitungan jarak menggunakan metode Euclidean Distance dengan rumus sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (1)$$

Keterangan:

$d(x,y)$: Jarak

x_{1i} : Sample data

x_{2i} : Data testing

i : Variabel data

n : Dimensi Data

Metode Algoritma Naive Bayes

Naive Bayes merupakan metode klasifikasi berbasis supervised learning yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen. Algoritma ini banyak digunakan dalam proses klasifikasi karena memiliki konsep sederhana dan mampu memberikan hasil yang baik pada berbagai jenis data[9].

Prinsip kerja Naive Bayes adalah menghitung probabilitas suatu data terhadap kelas tertentu berdasarkan data pelatihan yang tersedia[10]. Kelas dengan probabilitas terbesar akan dipilih sebagai hasil klasifikasi. Rumus Teorema Bayes ditunjukkan sebagai berikut:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (2)$$

Keterangan:

$P(A|B)$: Probabilitas posterior

$P(B|A)$: Likelihood

$P(A)$: Prior probability

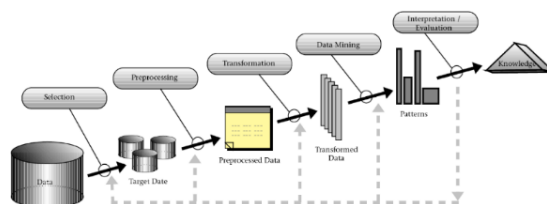
$P(B)$: Evidence

Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini menggunakan studi pustaka dan data sekunder. Studi pustaka dilakukan dengan mengumpulkan referensi dari jurnal, artikel ilmiah, buku, dan penelitian terdahulu yang berkaitan dengan data mining, klasifikasi, algoritma K-Nearest Neighbor (KNN), dan Naive Bayes. Tahap ini bertujuan untuk memperoleh landasan teori yang mendukung penelitian.

Data yang digunakan dalam penelitian ini merupakan dataset Student Performance yang diperoleh dari platform Kaggle. Dataset tersebut berisi sekitar 1000 data siswa dengan beberapa atribut, seperti gender, parental level of education, test preparation course, math score, reading score, dan writing score. Data ini dipilih karena memiliki struktur yang lengkap dan relevan untuk proses klasifikasi prestasi akademik siswa.

Tahap Penelitian



Gambar 1. Alur Knowledge Discovery in Database

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD) sebagai tahapan pengolahan data. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

a. Pengumpulan Data

Tahap pertama dilakukan dengan mengumpulkan dataset Student Performance dari Kaggle sebagai data penelitian.

b. Preprocessing Data

Pada tahap ini dilakukan pembersihan data, pemilihan atribut, transformasi data, serta penambahan atribut rata-rata nilai dan kelas prestasi akademik sebagai label klasifikasi.

c. Proses Klasifikasi

Data yang telah diproses kemudian dibagi menjadi data training dan data testing. Selanjutnya dilakukan penerapan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes untuk membangun model klasifikasi prestasi akademik siswa.

d. Evaluasi Model

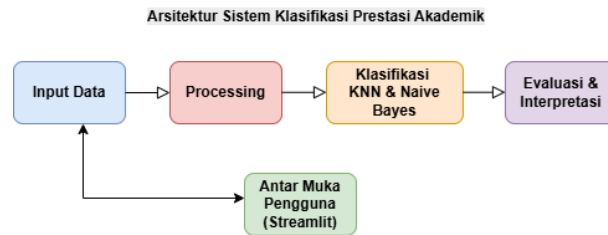
Hasil klasifikasi dievaluasi menggunakan beberapa metrik pengujian seperti accuracy, precision, recall, dan confusion matrix untuk mengetahui performa masing-masing algoritma.

e. Interpretasi Hasil

Tahap terakhir dilakukan dengan membandingkan hasil evaluasi algoritma KNN dan Naive Bayes untuk menentukan algoritma yang memiliki performa terbaik dalam klasifikasi prestasi akademik siswa.

Perancangan Sistem

Perancangan sistem pada penelitian ini bertujuan untuk membangun proses klasifikasi prestasi akademik siswa menggunakan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes. Sistem dimulai dari proses input dataset, preprocessing data, proses klasifikasi, hingga evaluasi hasil model. Data yang telah diproses akan digunakan sebagai data training dan data testing untuk menghasilkan model klasifikasi yang dapat menentukan kategori prestasi akademik siswa.



Gambar 2. Perancangan / alur sistem

Deskripsi Dataset

Tabel 1. Atribut Pada Dataset Student Performance

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Rata-Rata	Kelas prestasi
female	group B	bachelor's degree	standard	none	72	72	74	72,67	Sedang
female	group C	some college	standard	completed	69	90	88	82,33	Tinggi
female	group B	master's degree	standard	none	90	95	93	92,67	Tinggi
male	group A	associate's degree	free/reduced	none	47	57	44	49,33	Rendah
male	group C	some college	standard	none	76	78	75	76,33	Sedang
female	group B	associate's degree	standard	none	71	83	78	77,33	Sedang
female	group B	some college	standard	completed	88	95	92	91,67	Tinggi
male	group B	some college	free/reduced	none	40	43	39	40,67	Rendah
male	group D	high school	free/reduced	completed	64	64	67	65,00	Sedang
female	group B	high school	free/reduced	none	38	60	50	49,33	Rendah
male	group C	associate's degree	standard	none	58	54	52	54,67	Rendah
male	group D	associate's degree	standard	none	40	52	43	45,00	Rendah
female	group B	high school	standard	none	65	81	73	73,00	Sedang
male	group A	some college	standard	completed	78	72	70	73,33	Sedang

Dataset yang digunakan dalam penelitian ini adalah Student Performance yang diperoleh dari platform Kaggle. Dataset terdiri dari sekitar 1000 data siswa dengan beberapa atribut, yaitu gender, race/ethnicity, parental level of education, lunch, test preparation course, math score, reading score, dan writing score. Selain itu, ditambahkan atribut rata-rata nilai dan kelas prestasi sebagai label klasifikasi dengan kategori rendah, sedang, dan tinggi.

Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan data sebelum proses klasifikasi. Proses ini meliputi pembersihan data, pemilihan atribut, transformasi data, serta penambahan atribut rata-rata nilai dan kelas prestasi akademik. Nilai rata-rata diperoleh dari perhitungan math score, reading score, dan writing score. Selanjutnya, nilai rata-rata dikategorikan menjadi kelas prestasi rendah, sedang, dan tinggi.

Implementasi Model

Implementasi model dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes. Pada algoritma KNN, proses klasifikasi dilakukan berdasarkan kedekatan jarak antar data menggunakan metode Euclidean Distance. Sedangkan algoritma Naive Bayes melakukan klasifikasi berdasarkan probabilitas kemunculan atribut terhadap kelas tertentu. Proses implementasi dilakukan menggunakan data training dan data testing untuk menghasilkan model klasifikasi prestasi akademik siswa.

Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa algoritma dalam proses klasifikasi. Pengujian dilakukan menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan confusion matrix. Hasil evaluasi digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan prestasi akademik siswa.

Perbandingan Hasil

Tahap perbandingan hasil dilakukan dengan membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naive Bayes berdasarkan hasil evaluasi model. Perbandingan dilakukan menggunakan nilai accuracy, precision, recall, dan confusion matrix untuk mengetahui algoritma yang memiliki performa terbaik dalam klasifikasi prestasi akademik pada dataset *Student Performance*.

III. HASIL DAN PEMBAHASAN

Deskriptif Dataset

Dataset yang digunakan dalam penelitian ini adalah Student Performance yang diperoleh dari platform Kaggle. Dataset terdiri dari sekitar 1000 data siswa dengan beberapa atribut, yaitu gender, race/ethnicity, parental level of education, lunch, test preparation course, math score, reading score, dan writing score.

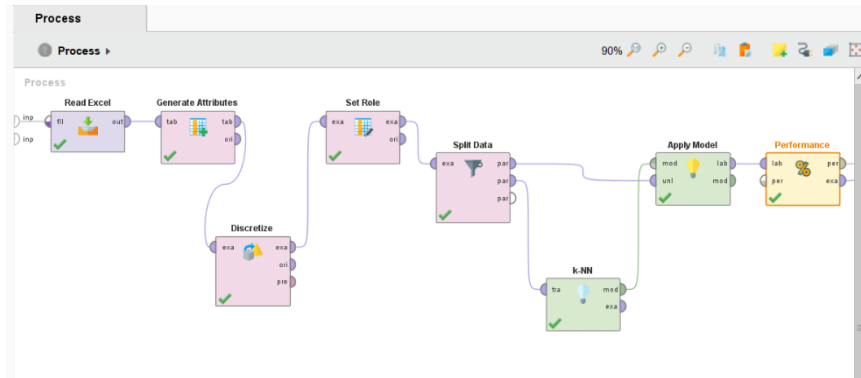
Pada penelitian ini dilakukan penambahan atribut rata-rata nilai yang diperoleh dari perhitungan math score, reading score, dan writing score. Selanjutnya, nilai rata-rata tersebut diklasifikasikan ke dalam tiga kategori prestasi akademik, yaitu rendah, sedang, dan tinggi. Proses pengolahan data dilakukan menggunakan RapidMiner melalui tahapan preprocessing, pembagian data training dan testing, serta implementasi algoritma klasifikasi.

Perbandingan Kinerja Model Klasifikasi

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu Naive Bayes dan K-Nearest Neighbor (KNN). Evaluasi model dilakukan menggunakan nilai accuracy, precision, recall, dan confusion matrix untuk mengetahui performa masing-masing algoritma dalam mengklasifikasikan prestasi akademik siswa.

Hasil pengujian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi dengan baik. Namun, terdapat perbedaan tingkat performa pada masing-masing model. Berdasarkan hasil evaluasi, algoritma Naive Bayes memperoleh tingkat akurasi yang lebih tinggi dibandingkan algoritma KNN.

3.1. Metode Algoritma K-Nearest Neighbor



Gambar 3. Alur Proses K-Nearest Neighbor

Criterion: ☒ Table View ☐ Plot View

accuracy: 91.12%

	true rendah	true sedang	true tinggi	class precision
pred. rendah	205	15	0	93.18%
pred. sedang	14	375	22	91.24%
pred. tinggi	0	20	149	88.17%
class recall	93.61%	91.46%	87.13%	

Gambar 4. Hasil Accuracy K-Nearest Neighbor

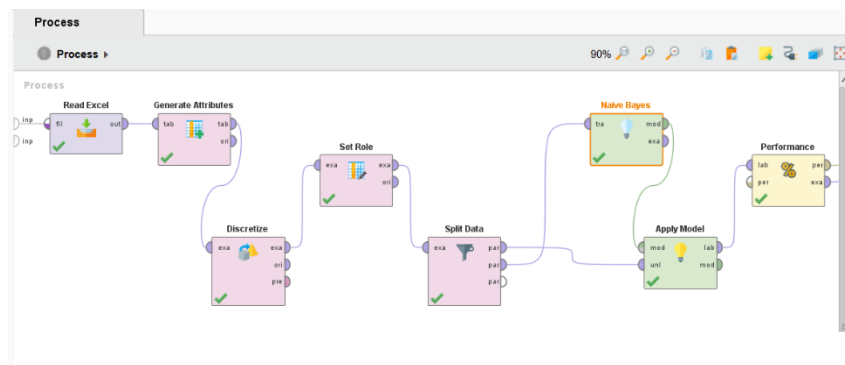
Performance	Description	Annotations
PerformanceVector:	accuracy: 91.12%	
ConfusionMatrix:	True: rendah sedang tinggi rendah: 205 15 0 sedang: 14 375 22 tinggi: 0 20 149 classification_error: 8.88%	
ConfusionMatrix:	True: rendah sedang tinggi rendah: 205 15 0 sedang: 14 375 22 tinggi: 0 20 149 weighted_mean_recall: 90.74%, weights: 1, 1, 1	
ConfusionMatrix:	True: rendah sedang tinggi rendah: 205 15 0 sedang: 14 375 22 tinggi: 0 20 149 weighted_mean_precision: 90.86%, weights: 1, 1, 1	
ConfusionMatrix:	True: rendah sedang tinggi rendah: 205 15 0 sedang: 14 375 22 tinggi: 0 20 149	

Gambar 5. Hasil Performance K-Nearest Neighbor

Berdasarkan hasil pengujian menggunakan algoritma K-Nearest Neighbor (KNN), diperoleh nilai accuracy sebesar 91,12% dengan classification error sebesar 8,88%. Nilai weighted mean recall yang diperoleh sebesar 90,74% dan weighted mean precision sebesar 90,86%.

Hasil confusion matrix menunjukkan bahwa pada kelas rendah terdapat 205 data yang berhasil diklasifikasikan dengan benar dan 15 data salah klasifikasi ke kelas sedang. Pada kelas sedang terdapat 375 data yang berhasil diklasifikasikan dengan benar, sedangkan 14 data salah diklasifikasikan ke kelas rendah dan 22 data ke kelas tinggi. Pada kelas tinggi terdapat 149 data yang berhasil diklasifikasikan dengan benar dan 20 data salah diklasifikasikan ke kelas sedang.

3.2. Metode Algoritma Naive Bayes



Gambar 6. Alur Proses Naive Bayes

Criterion: ☒ Table View ☐ Plot View

accuracy: 92.75%

	true rendah	true sedang	true tinggi	class precision
pred rendah	208	15	0	93.27%
pred sedang	11	378	15	93.56%
pred tinggi	0	17	156	90.17%
class recall	94.98%	92.20%	91.23%	

Gambar 7. Hasil Accuracy Naive Bayes

Performance	Description	Annotations
PerformanceVector	PerformanceVector: accuracy: 92.75% ConfusionMatrix: True: rendah sedang tinggi rendah: 208 15 0 sedang: 11 378 15 tinggi: 0 17 156 classification_error: 7.25% ConfusionMatrix: True: rendah sedang tinggi rendah: 208 15 0 sedang: 11 378 15 tinggi: 0 17 156 weighted_mean_recall: 92.80%, weights: 1, 1, 1 ConfusionMatrix: True: rendah sedang tinggi rendah: 208 15 0 sedang: 11 378 15 tinggi: 0 17 156 weighted_mean_precision: 92.34%, weights: 1, 1, 1 ConfusionMatrix: True: rendah sedang tinggi rendah: 208 15 0 sedang: 11 378 15 tinggi: 0 17 156	

Gambar 8. Hasil Performance Naive Bayes

Berdasarkan hasil pengujian menggunakan algoritma Naive Bayes dengan menggunakan tols Rapid Miner, diperoleh nilai accuracy sebesar 92,75% dengan classification error sebesar 7,25%. Selain itu, nilai weighted mean recall yang diperoleh sebesar 92,80% dan weighted mean precision sebesar 92,34%.

Hasil confusion matrix menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Pada kelas rendah terdapat 208 data yang berhasil diklasifikasikan dengan benar dan 15 data salah klasifikasi ke kelas sedang. Pada kelas sedang terdapat 378 data yang berhasil diklasifikasikan dengan benar, sedangkan 11 data salah diklasifikasikan ke kelas rendah dan 15 data ke kelas tinggi. Pada kelas tinggi terdapat 156 data yang berhasil diklasifikasikan dengan benar dan 17 data salah diklasifikasikan ke kelas sedang.

Berdasarkan hasil tersebut, algoritma Naive Bayes menunjukkan performa yang sangat baik dalam melakukan klasifikasi prestasi akademik siswa pada dataset Student Performance.

3.3. Hasil dan Perbandingan Algoritma

Hasil pengujian menggunakan algoritma K-Nearest Neighbor (KNN) menunjukkan nilai accuracy sebesar 91,12% dengan classification error sebesar 8,88%. Nilai weighted mean recall yang diperoleh sebesar 90,74%, sedangkan nilai weighted mean precision sebesar 90,86%.

Tabel 2. Hasil Pengujian Algoritma K-Nearest Neighbor

Metode	Accuracy	Precision	Recall	Classification Error
KNN	91.12%	90.86%	90.74%	8,88%

Berdasarkan hasil confusion matrix, pada kelas rendah terdapat 205 data yang berhasil diklasifikasikan dengan benar dan 15 data salah diklasifikasikan ke kelas sedang. Pada kelas sedang, sebanyak 375 data berhasil dikenali dengan benar, sedangkan 14 data salah diklasifikasikan ke kelas rendah dan 22 data ke kelas tinggi. Pada kelas tinggi, sebanyak 149 data berhasil diklasifikasikan dengan tepat dan 20 data salah diklasifikasikan ke kelas sedang.

Hasil tersebut menunjukkan bahwa algoritma KNN juga memiliki kemampuan klasifikasi yang baik, meskipun tingkat akurasi masih sedikit lebih rendah dibandingkan algoritma Naive Bayes.

Sedangkan hasil pengujian menggunakan algoritma Naive Bayes, diperoleh nilai accuracy sebesar 92,75% dengan nilai classification error sebesar 7,25%. Selain itu, model menghasilkan nilai weighted mean recall sebesar 92,80% dan weighted mean precision sebesar 92,34%.

Tabel 3. Hasil Pengujian Algoritma Naive Bayes

Metode	Accuracy	Precision	Recall	Classification Error
Naive Bayes	92.75%	92.34%	92.80%	7.25%

Hasil confusion matrix menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Pada kelas rendah, sebanyak 208 data berhasil diklasifikasikan dengan tepat dan 15 data salah diklasifikasikan ke kelas sedang. Pada kelas sedang, sebanyak 378 data berhasil diklasifikasikan dengan benar, sedangkan 11 data salah diklasifikasikan ke kelas rendah dan 15 data ke kelas tinggi. Sementara itu, pada kelas tinggi, sebanyak 156 data berhasil dikenali dengan benar dan 17 data salah diklasifikasikan ke kelas sedang. Berdasarkan hasil tersebut, algoritma Naive Bayes menunjukkan performa yang sangat baik dalam mengenali pola data dan mengklasifikasikan tingkat prestasi akademik siswa.

Sehingga hasil evaluasi, algoritma Naive Bayes memperoleh performa yang lebih baik dibandingkan algoritma K-Nearest Neighbor pada seluruh metrik pengujian. Hal ini menunjukkan bahwa pendekatan probabilitas pada Naive Bayes lebih efektif dalam mengenali pola data akademik pada dataset Student Performance.

Tabel 4. Hasil Perbandingan Algoritma K-Nearest Neighbor dan Naive Bayes

Metode	Accuracy	Precision	Recall	Classification Error
K-Nearest Neighbor	91.12%	90.86%	90.74%	8,88%

Naïve Bayes	92.75%	92.34%	92.80%	7.25%
----------------	--------	--------	--------	-------

3.4. Pembuatan Aplikasi

Setelah proses klasifikasi dan pengujian model selesai dilakukan, tahap selanjutnya adalah membangun aplikasi klasifikasi prestasi akademik siswa berbasis web. Aplikasi ini dibuat untuk mengimplementasikan algoritma Naive Bayes yang telah dilatih menggunakan dataset Student Performance. Sistem dibangun menggunakan bahasa pemrograman Python dengan framework Flask serta antarmuka berbasis HTML, CSS, dan Bootstrap sehingga aplikasi dapat diakses melalui browser secara interaktif.

Model Aplikasi Klasifikasi Prestasi Akademik Siswa Pada halaman Tentang Model, aplikasi menampilkan informasi mengenai model klasifikasi yang digunakan dalam sistem. Informasi tersebut meliputi algoritma yang digunakan, dataset, jumlah data, ukuran data testing, random state, serta kategori kelas prestasi akademik siswa.

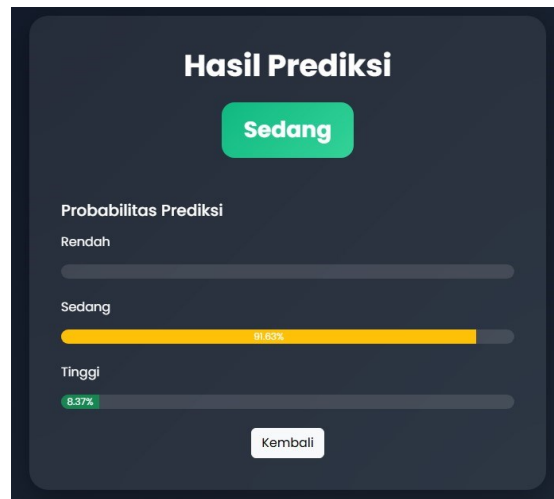
Gambar 9. Tampilan Form Prediksi

Halaman form prediksi merupakan halaman utama aplikasi yang digunakan pengguna untuk memasukkan data siswa sebelum dilakukan proses klasifikasi. Form prediksi dibuat dengan tampilan modern dan responsif sehingga memudahkan pengguna dalam melakukan input data.

Berdasarkan Gambar 9, pengguna diminta memasukkan beberapa atribut yang digunakan dalam proses klasifikasi prestasi akademik siswa. Data yang diinput meliputi gender, parental level of education, lunch, test preparation course, math score, reading score, dan writing score.

Setelah seluruh data diisi, pengguna dapat menekan tombol Prediksi Sekarang untuk memproses data menggunakan model Naive Bayes dan menghasilkan kategori prestasi akademik siswa.

Halaman hasil prediksi digunakan untuk menampilkan hasil klasifikasi prestasi akademik siswa berdasarkan data yang telah dimasukkan sebelumnya. Sistem akan menampilkan kategori hasil prediksi beserta probabilitas dari masing-masing kelas.



Gambar 10. Tampilan Hasil Prediksi

Berdasarkan Gambar 10, sistem berhasil mengklasifikasikan data siswa ke dalam kategori Sedang. Selain hasil klasifikasi utama, aplikasi juga menampilkan probabilitas prediksi untuk setiap kelas, yaitu Rendah, Sedang, dan Tinggi dalam bentuk progress bar.

Hasil probabilitas menunjukkan bahwa kategori Sedang memiliki probabilitas tertinggi sebesar 91,63%, sedangkan kategori Tinggi memiliki probabilitas sebesar 8,37%, dan kategori Rendah mendekati 0%. Hal ini menunjukkan bahwa model Naive Bayes mampu memberikan tingkat keyakinan prediksi terhadap kategori prestasi akademik siswa berdasarkan atribut yang dimasukkan pengguna.

3.5. Pengujian Sistem

Pengujian sistem dilakukan menggunakan metode black box testing, yaitu pengujian yang dilakukan dengan menguji fungsi-fungsi utama aplikasi tanpa melihat kode program di dalam sistem. Pengujian dilakukan untuk memastikan bahwa seluruh fitur aplikasi berjalan sesuai dengan kebutuhan pengguna.

Tabel 5. Pengujian Sistem

No	Pengujian	Keluaran yang Diharapkan	Keterangan
1	Input gender	Sistem menerima pilihan gender siswa	Berhasil
2	Input parental level of education	Sistem menerima pilihan pendidikan orang tua	Berhasil
3	Input lunch	Sistem menerima pilihan jenis lunch siswa	Berhasil
4	Input test preparation course	Sistem menerima pilihan preparation course	Berhasil
5	Input math score	Sistem menerima nilai matematika siswa	Berhasil
6	Input reading score	Sistem menerima nilai membaca siswa	Berhasil
7	Input writing score	Sistem menerima nilai menulis siswa	Berhasil
8	Klik tombol Prediksi Sekarang	Sistem menampilkan hasil klasifikasi prestasi akademik siswa	Berhasil
9	Menampilkan probabilitas prediksi	Sistem menampilkan probabilitas kelas Rendah, Sedang, dan Tinggi	Berhasil
10	Klik tombol Kembali	Sistem kembali ke halaman utama	Berhasil

Berdasarkan hasil pengujian pada tabel di atas, seluruh fitur aplikasi dapat berjalan dengan baik sesuai fungsi yang diharapkan. Sistem mampu menerima input data pengguna, melakukan proses klasifikasi menggunakan algoritma Naive Bayes, serta menampilkan hasil prediksi beserta probabilitas

klasifikasi secara interaktif. Dengan demikian, aplikasi klasifikasi prestasi akademik siswa berbasis web ini dapat digunakan sebagai media prediksi prestasi akademik siswa secara efektif.

IV. KESIMPULAN

Berdasarkan hasil penelitian, algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dapat digunakan untuk melakukan klasifikasi prestasi akademik siswa pada dataset Student Performance. Proses klasifikasi menggunakan atribut gender, parental level of education, lunch, test preparation course, math score, reading score, dan writing score. Hasil evaluasi menunjukkan bahwa algoritma Naive Bayes memiliki performa lebih baik dibandingkan KNN dengan nilai accuracy sebesar 92,75%, precision 92,34%, recall 92,80%, dan classification error 7,25%. Sedangkan algoritma KNN memperoleh accuracy 91,12%, precision 90,86%, recall 90,74%, dan classification error 8,88%. Berdasarkan hasil tersebut, algoritma Naive Bayes dinilai lebih efektif dalam mengenali pola prestasi akademik siswa. Hasil penelitian juga berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Flask sehingga mempermudah proses prediksi prestasi akademik siswa secara interaktif.

REFERENSI

- [1] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.
- [2] A. Muhaimin, M. Amin Hariyadi, and M. I. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 193–202, 2024, doi: 10.55338/jikomsi.v7i1.2865.
- [3] N. S. Fauziah and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [4] P. Ramadhan, Yuhandri, and J. Veri, "Eksplorasi Algoritma Decision Tree untuk Penentuan Siswa Berprestasi," *bit-Tech*, vol. 7, no. 3, pp. 826–833, 2025, doi: 10.32877/bt.v7i3.2210.
- [5] A. S. Arifianto, K. Dewi Safitri, K. Agustianto, and I. G. Wiryawan, "Pengaruh Prediksi Missing Value pada Klasifikasi Decision Tree C4.5," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, pp. 779–786, 2022, doi: 10.25126/jtiik.2022944778.
- [6] A. Winantu and C. Khatimah, "Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa," *INTEK J. Inform. dan Teknol. Inf.*, vol. 6, no. 1, pp. 58–64, 2023, doi: 10.37729/intek.v6i1.3006.
- [7] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmadden, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 273–281, 2024, doi: 10.57152/malcom.v4i1.1085.
- [8] S. N. Bakri and L. S. Harahap, "Analisis klasifikasi Algoritma K-Nearest Neighbor (K-NN) pada struktur Daerah di Kota Medan," *J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 2, pp. 182–193, 2025, doi: 10.70340/jirsi.v4i2.165.
- [9] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes," *MDP Student Conf.*, vol. 2, no. 1, pp. 297–305, 2023, doi: 10.35957/mdp-sc.v2i1.4388.
- [10] F. Sholekhah, A. D. Putri, R. Rahmadden, and L. Efrizoni, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom: Comparison of Naive Bayes and K-Nearest Neighbors Algorithms for Metabolic Syndrome Classification," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 507–514, 2024.