Tersedia Online : http://e-journals.unmul.ac.id/

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : http://e-journals2.unmul.ac.id/index.php/atasi/index

Perbandingan Kinerja Algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes dalam Pengklasifikasian Skor Ujian Mahasiswa Berdasarkan Faktor Akademik dan Non-Akademik

Mohammad Hadi Sujatmiko [1)*], Zaehol Fatah [2)], Ahmad Homaidi [3)]
[1,2,3)] Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy

E-Mail : m.hadisujatmiko060607@gmail.com [1)]; zaeholfatah@gmail.com [2)]; ahmadhomaidi@ibrahimy.ac.id [3)]

ARTICLE INFO

Article history:
Received : 08/05/2026
Revised : 21/05/2026
Accepted : 27/05/2026
Available online : 31/08/2026

Keywords:
Academic performance
KNN
Naïve Bayes
Classification
Model accuracy

ABSTRACT

Student academic achievement is an important indicator for assessing the success of the learning process in higher education. Students' exam scores are influenced not only by academic factors, such as study hours, class attendance, study methods, exam difficulty, and course, but also by non-academic factors, such as age, gender, internet access, sleep duration, sleep quality, and learning facilities. This study aims to compare the performance of the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms in classifying students' exam scores into two categories: Poor and Good. The dataset used in this study consists of 5,716 records with 12 attributes. The research process includes data preprocessing, target transformation, data splitting into training and testing sets, algorithm implementation, and performance evaluation using accuracy, precision, recall, and classification error. The results show that Naïve Bayes achieved an accuracy of 83.02%, precision of 82.85%, recall of 82.79%, and classification error of 16.98%, making it more effective than KNN in classifying students' exam scores.

Kata Kunci :
Prestasi akademik mahasiswa
K-Nearest Neighbor (KNN)
Naïve Bayes
Klasifikasi skor ujian
Evaluasi performa model

ABSTRAK

Prestasi akademik mahasiswa merupakan salah satu indikator penting dalam menilai keberhasilan proses pembelajaran di perguruan tinggi. Skor ujian mahasiswa tidak hanya dipengaruhi oleh faktor akademik, seperti jam belajar, kehadiran kelas, metode belajar, tingkat kesulitan ujian, dan program studi, tetapi juga dipengaruhi oleh faktor non-akademik, seperti usia, jenis kelamin, akses internet, durasi tidur, kualitas tidur, dan fasilitas belajar. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes dalam mengklasifikasikan skor ujian mahasiswa ke dalam kategori Jelek dan Bagus. Dataset yang digunakan berjumlah 5.716 data dengan 12 atribut. Proses penelitian meliputi preprocessing data, transformasi target, pembagian data latih dan data uji, penerapan algoritma, serta evaluasi performa menggunakan accuracy, precision, recall, dan classification error. Hasil penelitian menunjukkan bahwa Naïve Bayes memperoleh accuracy sebesar 83,02%, precision 82,85%, recall 82,79%, dan classification error 16,98%, sehingga lebih unggul dibandingkan KNN dalam klasifikasi skor ujian mahasiswa.

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong transformasi signifikan dalam dunia pendidikan tinggi, khususnya dalam pengelolaan proses pembelajaran yang semakin berbasis data (Muhammad Aram, S.Pd. 2024). Institusi pendidikan tidak lagi hanya berperan sebagai penyedia pengetahuan, tetapi juga dituntut untuk mampu melakukan monitoring dan evaluasi perkembangan akademik mahasiswa secara lebih efektif, objektif, dan komprehensif (Muhammad Aram, 2024). Salah satu permasalahan yang sering dihadapi adalah keterlambatan dalam mendeteksi penurunan performa akademik mahasiswa, yang umumnya baru diketahui setelah hasil ujian akhir diumumkan (Warinda, 2025). Kondisi ini menyebabkan upaya intervensi akademik menjadi kurang optimal

*) Corresponding Author
https://doi.org/10.30872/atasi.v5i2.4652

karena dilakukan pada tahap yang sudah terlambat. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengklasifikasikan potensi hasil belajar mahasiswa secara dini, sehingga langkah preventif dan perbaikan dapat dilakukan secara lebih proaktif.

Prestasi akademik mahasiswa yang direpresentasikan melalui skor ujian tidak hanya dipengaruhi oleh faktor akademik, tetapi juga oleh faktor non-akademik yang saling berinteraksi secara kompleks (Arumdani 2025). Faktor akademik seperti jam belajar (*study hours*) dan kehadiran kelas (*class attendance*) secara umum dianggap sebagai indikator utama dalam menentukan keberhasilan belajar. Namun, penelitian menunjukkan bahwa faktor non-akademik seperti durasi tidur (*sleep hours*), kualitas tidur (*sleep quality*), akses internet (*internet access*), serta metode belajar (*study method*) juga memiliki pengaruh signifikan terhadap kesiapan belajar, tingkat konsentrasi, dan daya serap mahasiswa. Bahkan, aspek psikologis seperti motivasi belajar turut berkontribusi dalam menentukan capaian akademik (Achmad Efendi dan Dahlia, 2025). Dengan demikian, pendekatan klasifikasi skor ujian mahasiswa perlu mempertimbangkan integrasi antara faktor akademik dan non-akademik untuk menghasilkan model yang lebih komprehensif dan akurat.

Dalam konteks ini, penerapan machine learning menjadi solusi yang relevan untuk mengatasi kompleksitas analisis data multidimensi. Machine learning memiliki kemampuan untuk mengidentifikasi pola tersembunyi serta hubungan non-linear dalam data yang sulit dianalisis secara manual (Gopinda, 2026). Melalui proses pembelajaran dari data historis, algoritma *machine learning* mampu membangun model klasifikasi yang dapat digunakan untuk memprediksi atau mengelompokkan data baru secara otomatis (Wijaya dan Bismi, 2025). Dalam bidang pendidikan, pendekatan ini telah banyak digunakan untuk mendukung sistem prediksi performa akademik mahasiswa secara lebih akurat dan efisien.

Di antara berbagai algoritma *machine learning*, algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* merupakan dua metode yang sering digunakan dalam permasalahan klasifikasi. KNN merupakan algoritma berbasis kedekatan (*instance-based learning*) yang mengklasifikasikan data berdasarkan kemiripan terhadap data latih terdekat (Muthohharoh, Lidimilah, dan Homaidi 2025). Algoritma ini memiliki keunggulan dalam menangani pola data yang kompleks tanpa memerlukan asumsi distribusi tertentu. Sementara itu, *Naïve Bayes* merupakan algoritma berbasis probabilitas yang menggunakan *Teorema Bayes* dengan asumsi independensi antar fitur. Keunggulan metode ini terletak pada efisiensi komputasi, kemampuan menangani data berdimensi tinggi, serta performa yang cukup stabil pada berbagai kondisi data (Zulkifli et al, 2026).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengimplementasikan serta membandingkan kinerja algoritma KNN dan *Naïve Bayes* dalam mengklasifikasikan skor ujian mahasiswa. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset publik yang mencakup berbagai variabel akademik dan non-akademik. Proses penelitian meliputi tahapan pengumpulan data, *preprocessing*, penerapan algoritma, serta evaluasi kinerja menggunakan metrik klasifikasi seperti *akurasi, presisi, recall*, dan *confusion matrix.*

Kebaruan (*novelty*) dalam penelitian ini terletak pada integrasi faktor akademik dan non-akademik dalam proses klasifikasi skor ujian mahasiswa menggunakan perbandingan algoritma KNN dan *Naïve Bayes*. Berbeda dengan penelitian sebelumnya yang umumnya hanya berfokus pada faktor akademik atau performa algoritma secara umum, penelitian ini menggabungkan variabel seperti kualitas tidur, akses internet, dan fasilitas belajar sebagai bagian dari proses klasifikasi. Selain itu, model terbaik hasil klasifikasi juga diimplementasikan ke dalam aplikasi berbasis *web* menggunakan Python dan Flask sehingga hasil penelitian tidak hanya bersifat teoritis tetapi juga aplikatif.

Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi berbasis bukti mengenai algoritma yang paling efektif dalam mengklasifikasikan skor ujian mahasiswa. Selain itu, penelitian ini juga diharapkan dapat berkontribusi dalam pengembangan sistem pendukung keputusan di bidang pendidikan, khususnya dalam membantu proses monitoring dan evaluasi performa akademik mahasiswa secara lebih dini, akurat, dan holistik.

2. TINJAUAN PUSTAKA

A. Study Literatur

Penelitian yang dilakukan oleh Sathe dan Adamuthe (2021) yang berjudul *Comparative Study of Supervised Algorithms for Predicting Student Performance* bertujuan untuk membandingkan kinerja tujuh *algoritma supervised learning*, yaitu C5.0, J48, CART, *Naïve Bayes,* KNN, *Random Forest, dan Support Vector Machine (SVM)* dalam memprediksi performa akademik siswa. Metode yang digunakan adalah klasifikasi dengan tiga dataset berbeda, yaitu data siswa tingkat sekolah sebanyak 395 record dengan 33 atribut, data mahasiswa sebanyak 133 record dengan 22 atribut, serta data e-learning sebanyak 480 record dengan 16 atribut. Tahapan penelitian meliputi seleksi atribut menggunakan *korelasi Pearson* untuk menentukan variabel yang paling berpengaruh terhadap nilai akhir, dilanjutkan dengan penerapan algoritma klasifikasi serta penyesuaian parameter. Evaluasi dilakukan menggunakan *metrik akurasi, precision, recall, true positive rate (TPR),* dan *false positive rate (FPR).*

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* dan *C5.0* memiliki performa terbaik dibandingkan algoritma lainnya. *Random Forest* mencapai akurasi sebesar 99,75% pada dataset sekolah, 100% pada dataset perguruan tinggi, dan 98,12% pada dataset *e-learning*, sedangkan *C5.0* memperoleh akurasi sebesar 98,48%, 100%, dan 93,33%. Algoritma lain seperti J48, CART, *Naïve Bayes*, KNN, dan SVM menunjukkan performa yang lebih rendah. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma berbasis *ensemble* seperti *Random Forest* lebih unggul dalam menangani kompleksitas data dan menghasilkan akurasi yang tinggi, serta menunjukkan bahwa pemilihan algoritma sangat dipengaruhi oleh karakteristik dataset dan parameter yang digunakan.

https://doi.org/10.30872/atasi.v5i2.4652

Penelitian lain yang dilakukan oleh (Munarsih dan Ningsi, 2023) yang berjudul Perbandingan Kinerja *Klasifikasi Data Mining* Algoritma untuk Prediksi Prestasi Akademik Mahasiswa bertujuan untuk memprediksi prestasi akademik mahasiswa menggunakan beberapa algoritma klasifikasi yaitu *Naïve Bayes,* C4.5, dan KNN. Variabel yang digunakan dalam penelitian ini meliputi jenis kelamin, usia, fakultas, daerah asal, status pekerjaan, partisipasi organisasi, jenis sekolah asal, jarak tempat tinggal, serta profesi orang tua sebagai atribut independen, sedangkan Indeks Prestasi Kumulatif (IPK) digunakan sebagai variabel dependen. Metode yang digunakan dalam penelitian ini adalah teknik data mining dengan pendekatan klasifikasi serta validasi menggunakan metode *cross validation*. Data yang digunakan berasal dari hasil kuesioner terhadap 87 mahasiswa yang kemudian diolah menggunakan bantuan perangkat lunak Python. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki tingkat akurasi tertinggi sebesar 56,25%, diikuti oleh algoritma *Naïve Bayes dan C4.5* yang masing-masing memiliki tingkat akurasi sebesar 50,00% .

Penelitian lain juga dilakukan oleh (Amra, Ihsan A. Abu dan Ashraf Y. A. Maghari, 2017) yang berjudul *Predicting Student Performance Using KNN and Naïve Bayesian* bertujuan untuk memprediksi kinerja siswa dengan menggunakan teknik data mining pada dataset pendidikan sekolah menengah yang diperoleh dari Kementerian Pendidikan di Jalur Gaza. Dalam penelitian ini digunakan dua metode klasifikasi yaitu KNN dan *Naïve Bayes* untuk mengetahui algoritma yang memiliki performa terbaik dalam melakukan prediksi kinerja siswa.

Atribut yang digunakan dalam penelitian ini meliputi beberapa faktor yang memengaruhi kinerja siswa seperti jenis kelamin, pekerjaan ayah, lokasi tempat tinggal, nilai rata-rata siswa pada tahun sebelumnya, serta beberapa atribut pendidikan lainnya. Dataset yang digunakan berjumlah 500 data yang telah melalui proses *preprocessing* dengan reduksi atribut dari 14 menjadi 8 atribut penting guna meningkatkan kualitas data. Proses klasifikasi dilakukan dengan membagi data menjadi 70% data latih dan 30% data uji, serta menggunakan parameter evaluasi berupa *akurasi, precision, dan recall*. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes* memiliki performa yang lebih baik dibandingkan dengan KNN dengan tingkat akurasi sebesar 93,17% hingga 93,6%, sedangkan metode KNN memperoleh akurasi sekitar 63% . Hal ini menunjukkan bahwa *Naïve Bayes* lebih efektif dalam menangani dataset dengan karakteristik atribut yang memiliki hubungan probabilistik yang kuat terhadap kelas target.

Berdasarkan beberapa penelitian terdahulu, sebagian besar penelitian hanya berfokus pada perbandingan performa algoritma klasifikasi tanpa melakukan analisis gabungan antara faktor akademik dan non-akademik secara lebih komprehensif. Selain itu, implementasi hasil model ke dalam aplikasi prediksi berbasis web masih jarang dilakukan. Oleh karena itu, penelitian ini tidak hanya membandingkan performa algoritma KNN dan Naïve Bayes, tetapi juga menganalisis pengaruh faktor non-akademik terhadap performa akademik mahasiswa serta mengimplementasikan model terbaik ke dalam sistem prediksi berbasis web.

B. Landasan Teori

1. *Data Mining*

Data mining merupakan proses penggalian informasi atau pola tersembunyi dari sekumpulan data dalam jumlah besar dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, serta *machine learning*. Konsep ini berkembang seiring meningkatnya kebutuhan dalam mengolah data yang kompleks agar dapat menghasilkan informasi yang berguna bagi pengambilan keputusan. *Data mining* tidak hanya berfokus pada pengumpulan data, tetapi juga pada proses menemukan hubungan, tren, dan pola tertentu yang sebelumnya sulit diketahui secara manual.

Dalam penerapannya, *data mining* digunakan pada berbagai bidang, seperti pendidikan, kesehatan, bisnis, hingga perbankan. Proses data mining biasanya melibatkan beberapa tahapan, seperti seleksi data, pembersihan data, transformasi data, proses penggalian pola, dan evaluasi hasil. Dengan adanya data mining, data yang awalnya hanya berupa kumpulan informasi dapat diubah menjadi pengetahuan yang bernilai sehingga mampu mendukung proses analisis dan prediksi (Feri SLN, 2023).

2. Perbandingan

Perbandingan merupakan suatu metode analisis yang digunakan untuk menilai perbedaan atau persamaan antara dua objek, metode, sistem, atau variabel tertentu. Dalam penelitian, perbandingan sering digunakan untuk mengetahui tingkat efektivitas, akurasi, efisiensi, atau performa dari beberapa metode yang diuji. Dengan melakukan perbandingan, peneliti dapat menentukan metode yang paling sesuai berdasarkan hasil pengukuran tertentu (Rohman dan Wibowo, 2024).

Dalam konteks data mining, perbandingan umumnya dilakukan terhadap algoritma klasifikasi atau prediksi untuk mengetahui algoritma mana yang memberikan hasil terbaik. Dalam konteks penelitian ini, perbandingan dilakukan untuk mengevaluasi kinerja dua algoritma klasifikasi, yaitu KNN dan *Naïve Bayes Classifier*. Evaluasi kinerja dilakukan menggunakan metrik klasifikasi seperti *akurasi, presisi*, *recall*, dan *confusion matrix*, sehingga dapat diketahui algoritma mana yang memiliki performa lebih baik dalam mengklasifikasikan skor ujian mahasiswa.

3. Klasifikasi

Klasifikasi merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimiliki. Proses klasifikasi terdiri dari dua tahap utama, yaitu tahap pembelajaran (*training*) dan tahap pengujian (*testing*) (Pebdika, Herdiana, dan Solihudin, 2023). Pada tahap pembelajaran, sistem membangun model berdasarkan data latih yang telah diketahui label kelasnya, sedangkan pada tahap pengujian, model tersebut digunakan untuk memprediksi kelas pada data baru yang belum diketahui labelnya. Dalam penelitian ini, klasifikasi digunakan untuk mengelompokkan skor ujian mahasiswa ke dalam kategori tertentu berdasarkan faktor akademik dan non-akademik.

https://doi.org/10.30872/atasi.v5i2.4652

4. *K-Nearest Neighbor* (*KNN*)

KNN merupakan salah satu algoritma klasifikasi yang termasuk dalam kategori *instance-based learning* atau *lazy learning*. Prinsip kerja KNN adalah menentukan kelas suatu data berdasarkan kedekatannya dengan sejumlah *k* data terdekat dalam data latih. Kedekatan antar data umumnya dihitung menggunakan metode jarak, seperti jarak *Euclidean.* Data baru akan diklasifikasikan ke dalam kelas yang paling banyak muncul di antara tetangga terdekatnya (*majority voting*) (Zhang, 2016).

Kelebihan dari metode KNN adalah kemudahan implementasi serta kemampuannya dalam menangani hubungan data yang bersifat non-linear tanpa memerlukan asumsi distribusi tertentu. Namun, metode ini memiliki kelemahan, seperti sensitif terhadap skala data, nilai *k* yang digunakan, serta keberadaan data *noise*. Oleh karena itu, proses normalisasi data dan pemilihan nilai *k* yang optimal menjadi hal yang penting dalam penerapan metode ini.

Adapun Salah satu rumus yang paling umum digunakan yaitu *Euclidean Distance* (Nooraeni dan Nurfalah 2022), sebagai berikut:

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2} \quad \ldots\ldots(1)$$

Keterangan:
d(x,y) : Jarak
x_{1i} : Sample data
x_{2i} : Data testing
i : Variabel data
n : Dimensi Data

5. *Naïve Bayes Classifier*

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan konsep *Teorema Bayes* dalam menentukan kemungkinan suatu data masuk ke dalam kelas tertentu. Algoritma ini mengasumsikan bahwa setiap atribut bersifat independen atau tidak saling memengaruhi satu sama lain. Walaupun asumsi tersebut tergolong sederhana, *Naive Bayes* mampu memberikan performa yang baik pada berbagai jenis data (Watratan, B, dan Moeis, 2020).

Dalam proses klasifikasi, *Naive Bayes* menghitung probabilitas setiap kelas berdasarkan data yang dimiliki, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi. Algoritma ini banyak digunakan dalam klasifikasi teks, analisis sentimen, diagnosis penyakit, dan prediksi akademik. Keunggulan *Naive Bayes* terletak pada kecepatan proses perhitungan serta kemampuan bekerja dengan baik pada dataset berukuran besar.

Adapun Perhitungan probabilitas pada algoritma *Naive Bayes* menggunakan rumus *teorema Bayes* (Maryvonne Lumley & James Wilkinson, 2014) sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad \ldots\ldots(2)$$

Keterangan:
X = data dengan class yang belum diketahui
H = hipotesis data *class* spesifik
P(H|X) = probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)
P(H) = probabilitas hipotesis H (*prior probability*)
P(X|H) = probabilitas X berdasarkan kondisi H (*likelihood probability*)
P(X) = probabilitas X (*evidence probability*)

6. Faktor *akademik* dan non-akademik

Faktor akademik dan non-akademik merupakan variabel penting dalam menentukan performa belajar mahasiswa. Skor ujian mahasiswa dipengaruhi oleh berbagai faktor yang saling berkaitan. Faktor akademik meliputi aspek yang secara langsung berhubungan dengan proses pembelajaran, seperti jam belajar (*study hours*), kehadiran di kelas (*class attendance*), tingkat kesulitan ujian (*exam difficulty*), dan metode belajar (*study method*). Sementara itu, faktor non-akademik mencakup aspek pendukung di luar kegiatan belajar formal, seperti durasi dan kualitas tidur (*sleep hours* dan *sleep quality*), akses internet (*internet access*), serta fasilitas pendukung belajar (*facility rating*). Kombinasi kedua faktor tersebut membentuk kondisi belajar mahasiswa secara menyeluruh, sehingga analisis skor akademik yang akurat perlu mempertimbangkan keduanya secara simultan.

7. RapidMiner

RapidMiner merupakan perangkat lunak yang digunakan untuk proses *data mining*, *machine learning*, dan analisis data. RapidMiner menyediakan tampilan berbasis visual sehingga pengguna dapat melakukan pengolahan data tanpa harus menulis kode program secara langsung. Dalam proses *data mining*, RapidMiner dapat digunakan untuk melakukan *preprocessing data*, pemilihan atribut, pembagian data latih dan data uji, penerapan algoritma klasifikasi, serta evaluasi performa model (Sudarsono et al., 2021).

Dalam penelitian ini, RapidMiner digunakan untuk membangun dan menguji model klasifikasi skor ujian mahasiswa menggunakan algoritma KNN dan Naïve Bayes. Penggunaan RapidMiner membantu peneliti dalam

https://doi.org/10.30872/atasi.v5i2.4652

membandingkan hasil performa kedua algoritma berdasarkan nilai *accuracy, precision, recall*, dan *classification error*.

8. Python

Python merupakan bahasa pemrograman tingkat tinggi yang banyak digunakan dalam pengembangan aplikasi, *analisis data, data mining,* dan *machine learning* (Runimeirati, Abdul Muis, dan Muhammad, 2023). Python memiliki *sintaks* yang sederhana sehingga mudah dipahami dan digunakan oleh pengembang. Selain itu, Python juga memiliki banyak pustaka pendukung seperti *pandas, scikit-learn*, dan *joblib* yang dapat digunakan untuk mengolah data, membangun model machine learning, serta menyimpan model hasil pelatihan.

9. Flask

Flask merupakan *framework web* berbasis Python yang digunakan untuk membangun aplikasi *web* secara sederhana dan ringan. *Flask* termasuk *micro-framework* karena menyediakan komponen inti yang minimal, tetapi tetap dapat dikembangkan sesuai kebutuhan aplikasi. Dengan Flask, pengembang dapat membuat halaman *web*, mengatur proses input data, memproses data melalui model *machine learning*, dan menampilkan hasil prediksi kepada pengguna (Runimeirati et al., 2023).

3. METODE PENELITIAN

A. Prosedur Penelitian

Adapun prosedur penelitian dapat direpresentasikan pada gambar ini

Gambar 1. *Flowchart* Penelitian (Moch. Rizky Yuliansyah, B, dan Franz, 2022)

1. Pengumpulan Data

Data diperoleh dari repository online pada *platform Kaggle* yakni *Exam Score Prediction*. Dataset ini berisi data skor ujian mahasiswa berdasarkan faktor akademik dan non-akademik yang digunakan untuk menganalisis dan mengklasifikasikan tingkat prestasi akademik berdasarkan beberapa faktor pendukung.

2. Preprocessing

Preprocessing merupakan tahap pengolahan awal data sebelum diterapkan ke algoritma klasifikasi. Tahap ini sangat penting karena kualitas data akan memengaruhi performa model. Proses preprocessing biasanya meliputi:

a. *Cleaning Data* → menghapus data kosong, duplikat, atau tidak konsisten
b. *Transformasi Data* → mengubah format data agar seragam
c. *Normalisasi Data* → menyamakan skala nilai atribut
d. *Seleksi Fitur* → memilih atribut yang relevan
e. *Encoding Data* → mengubah data kategorikal menjadi numerik

Tujuan preprocessing adalah menghasilkan dataset yang bersih dan siap digunakan pada tahap klasifikasi.

3. Penerapan Metode *KNN* Dan *Naive Bayes*

Metode *KNN* bekerja dengan mencari sejumlah tetangga terdekat berdasarkan jarak data. Data baru akan diklasifikasikan berdasarkan mayoritas kelas dari tetangga terdekat. Adapun langkah-langkahnya sebagai berikut:

a. Menentukan nilai K
b. Menghitung jarak *Euclidean*
c. Mengurutkan jarak terkecil
d. Menentukan kelas berdasarkan label mayoritas pada K

Sedangkan *Naive Bayes* menggunakan probabilitas berdasarkan Teorema Bayes. langkah-langkahnya sebagai berikut:

a. Menghitung probabilitas *prior*
b. Menghitung *likelihood* tiap atribut
c. Menghitung *posterior* probability
d. Menentukan kelas dengan probabilitas terbesar

https://doi.org/10.30872/atasi.v5i2.4652

4. Pengukuran Performa Metode *KNN* dan *Naive Bayes*

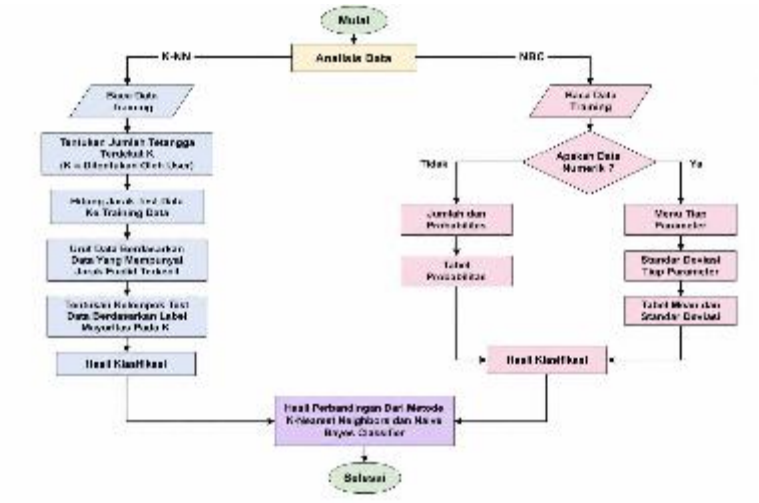
Setelah model diterapkan, langkah berikutnya adalah mengukur performa masing-masing algoritma. Pengukuran performa dilakukan untuk mengetahui kualitas model klasifikasi berdasarkan hasil prediksi. Parameter evaluasi yang digunakan adalah *Accuracy, Precision, Recall, Confusion Matrix*. Tahap ini bertujuan mengetahui tingkat keberhasilan model dalam melakukan klasifikasi.

5. Perbandingan Performa Metode *KNN* dan *Naive Bayes*

Tahap ini dilakukan untuk membandingkan hasil evaluasi kedua metode. Perbandingan dilakukan berdasarkan Tingkat akurasi, Ketepatan klasifikasi, Kecepatan proses, Tingkat error, Konsistensi hasil. Dari hasil perbandingan, peneliti dapat menentukan algoritma mana yang paling efektif untuk dataset penelitian. Tahap ini menjadi inti penelitian komparatif karena menghasilkan analisis objektif antara dua metode klasifikasi.

B. Analisis Perancangan Model

Adapun tahapan model dari metode KNN dan *Naïve Bayes Classifier* dapat direpresentasikan sebagai berikut (Moch. Rizky Yuliansyah et al. 2022)



Gambar 2. Analisis Perancangan Model

Berdasarkan desain model yang ditunjukkan pada Gambar 2, proses analisis dapat dijabarkan melalui beberapa tahapan yang menggambarkan penerapan metode KNN dan *Naïve Bayes Classifier*.

1. Analisis Perancangan Model KNN

Metode KNN bekerja dengan mengelompokkan data berdasarkan kedekatan jarak antara data uji dan data training. Pada *flowchart*, proses KNN dimulai dari pembacaan data *training* yang digunakan sebagai referensi klasifikasi.

Tahap pertama adalah membaca data training yang berisi atribut dan label kelas. Setelah itu ditentukan nilai K, yaitu jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi. Nilai K menentukan berapa banyak data terdekat yang dipertimbangkan dalam pengambilan keputusan. Langkah berikutnya adalah menghitung jarak antara data testing dengan seluruh data *training* menggunakan *Euclidean Distance*.

Hasil perhitungan jarak kemudian diurutkan dari nilai terkecil hingga terbesar. Sistem akan mengambil sejumlah K data dengan jarak terdekat, lalu menentukan kelas berdasarkan label mayoritas dari tetangga tersebut. Hasil akhir dari proses ini berupa klasifikasi data berdasarkan kedekatan karakteristik antar data.

2. Analisis Perancangan Model *Naive Bayes*

Metode Naive Bayes bekerja menggunakan pendekatan probabilitas berdasarkan Teorema Bayes. Proses dimulai dengan membaca data training yang digunakan untuk membangun model probabilitas setiap kelas. Setelah data dibaca, sistem mengidentifikasi jenis data, yaitu numerik atau kategorik. Pada data numerik, dilakukan perhitungan parameter statistik berupa mean dan standar deviasi untuk setiap atribut. Nilai tersebut digunakan untuk membentuk distribusi probabilitas tiap kelas. Sedangkan pada data kategorik, sistem menghitung frekuensi kemunculan atribut pada setiap kelas dan menentukan probabilitasnya menggunakan Teorema Bayes. Selanjutnya sistem membentuk tabel probabilitas sebagai dasar klasifikasi. Kelas dengan probabilitas terbesar dipilih sebagai hasil prediksi.

4. HASIL DAN PEMBAHASAN

A. Persiapan Data

Dataset yang digunakan terdiri dari 5.716 data (baris) dengan 12 atribut (kolom) yang merepresentasikan karakteristik mahasiswa serta faktor-faktor yang memengaruhi skor ujian.

1. Faktor akademik

Faktor akademik adalah variabel yang secara langsung mempengaruhi proses pembelajaran dan performa akademik mahasiswa. Variabel-variabel ini berhubungan langsung dengan aktivitas belajar, strategi akademik, dan lingkungan pendidikan formal. Adapun atribut datasetnya sebagai berikut

https://doi.org/10.30872/atasi.v5i2.4652

Tabel 1. Faktor Akademik

Study_Hours	*Class_Attendance*	*Study_Method*	*Exam_Difficulty*	*Course*	*Study_Hours*
2.78	92.9	coaching	hard	diploma	2.78
3.37	64.8	online videos	moderate	bca	3.37
7.88	76.8	coaching	moderate	b.sc	7.88
0.67	48.4	online videos	moderate	diploma	0.67
0.89	71.6	coaching	moderate	diploma	0.89
3.48	65.4	mixed	moderate	b.tech	3.48
1.35	69	online videos	hard	b.tech	1.35
5.48	51.1	self-study	moderate	b.sc	5.48
2.89	92	self-study	easy	bca	2.89
6.77	44.8	group study	moderate	bba	6.77

2. Faktor Non-akademik

Faktor non-akademik adalah variabel yang tidak langsung terkait dengan pembelajaran, tetapi dapat mempengaruhi performa secara tidak langsung. Faktor ini lebih ke kondisi pribadi, sosial, dan lingkungan yang mempengaruhi kesiapan belajar, bukan proses belajar itu sendiri.

Tabel 2. Faktor non-akademik

Age	*Gender*	*Internet_Access*	*Sleep_Hours*	*Sleep_Quality*	*Facility_Rating*
17	male	yes	7.4	poor	low
23	other	yes	4.6	average	medium
22	male	yes	8.5	poor	high
20	other	yes	5.8	average	low
20	female	yes	9.8	poor	low
23	male	yes	4.2	good	low
17	female	yes	7.4	average	high
22	male	yes	8.2	poor	low
18	other	yes	6.6	poor	low

Sehingga dapat dikatakan perbedaan dari faktor akademik dan non akademik terletak pada kebutuhan mahasiswa tersebut. Adapaun tabel terperincinya sebagai berikut:

Tabel 3. Rincian Dataset

Kategori	**Atribut**
Akademik	*study_hours, class_attendance, study_method, exam_difficulty, course*
Non-Akademik	*age, gender, internet_access, sleep_hours, sleep_quality, facility_rating*
Target	*exam_score*

B. Pre-processing Data

Tahapan pre-processing dilakukan untuk mempersiapkan data sebelum masuk ke proses klasifikasi. Berdasarkan alur proses pada RapidMiner, langkah-langkah yang dilakukan adalah sebagai berikut:

1. Retrieve Data

Tahap ini merupakan proses pengambilan dataset dari *repository* RapidMiner. Dataset yang digunakan adalah *Exam Score Prediction* yang terdiri dari 5.716 data dan 12 atribut. Data yang diambil masih dalam kondisi mentah sehingga perlu dilakukan pengolahan lebih lanjut sebelum digunakan dalam proses klasifikasi.

2. Select Attributes

Pada tahap ini dilakukan pemilihan atribut yang relevan untuk penelitian. Atribut yang digunakan terdiri dari:

a) Faktor Akademik: *study hours, class attendance, study method, exam difficulty, course*
b) Faktor Non-Akademik: *age, gender, internet_access, sleep_hours, sleep_quality, facility_rating*
c) Target: *exam score*

Tujuan dari tahap ini adalah untuk menghilangkan atribut yang tidak diperlukan sehingga proses analisis menjadi lebih fokus dan efisien.

3. Generate Attributes

Tahap ini digunakan untuk melakukan transformasi atau pembuatan atribut baru. Dalam penelitian ini, proses *generate attributes* dapat digunakan untuk:

a) Menyesuaikan format data
b) Melakukan normalisasi atau perhitungan tertentu
c) Mengubah tipe data agar sesuai dengan kebutuhan algoritma

Dalam penelitian ini Transformasi atribut *exam_score_prediction* menjadi kelas target kategorikal, yaitu :

a) Jelek, jika nilai exam_score < 60
b) Bagus, jika nilai exam_score ≥ 60

Transformasi ini bertujuan untuk menyesuaikan data dengan pendekatan klasifikasi *biner* yang digunakan dalam penelitian. Langkah ini membantu meningkatkan kualitas data sebelum dilakukan proses klasifikasi.

https://doi.org/10.30872/atasi.v5i2.4652

4. Set Role

Pada tahap ini dilakukan penentuan peran masing-masing atribut dalam dataset. Atribut:

a) *exam score* ditetapkan sebagai label (target)
b) Atribut lainnya sebagai *predictor* (input)

Penentuan ini sangat penting karena menentukan bagaimana algoritma memahami data dalam proses pembelajaran.

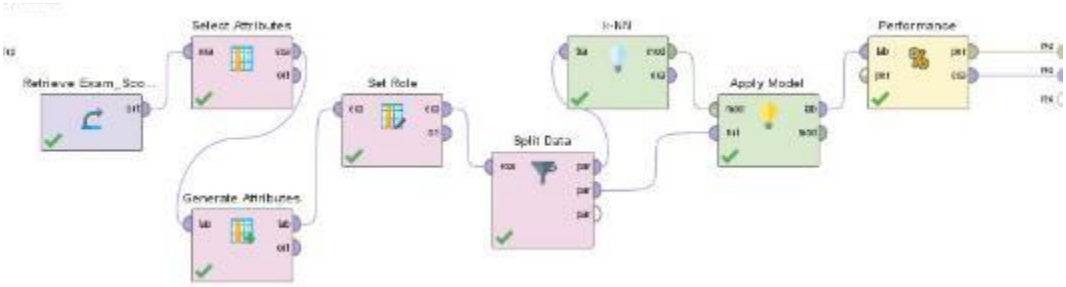
5. Split Data

data dibagi menjadi data latih dan data uji menggunakan *operator Split Data* dengan perbandingan *ratio* 70:30. Pembagian ini bertujuan untuk menguji performa model secara objektif terhadap data yang belum pernah dilatih sebelumnya agar hasil yang diperoleh lebih objektif dan tidak bias. Dengan melalui tahapan pre-processing ini, dataset menjadi lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi menggunakan algoritma *Naïve Bayes* maupun metode lainnya.

C. Hasil Data Mining

Penelitian ini memanfaatkan dua metode klasifikasi, yaitu KNN dan *Naive Bayes*. Proses pengolahan data mining dilakukan dengan menerapkan *operator Apply Model* guna menghasilkan prediksi dari model yang telah dibangun.

1. Penerapan Metode KNN



Gambar 3. Proses Klasifikasi KNN

accuracy: 78.94%

	true Jelek	true Bagus	class precision
pred. Jelek	579	173	76.99%
pred. Bagus	188	774	80.46%
class recall	75.49%	81.73%	

Gambar 4. *Confusion Matrix KNN*

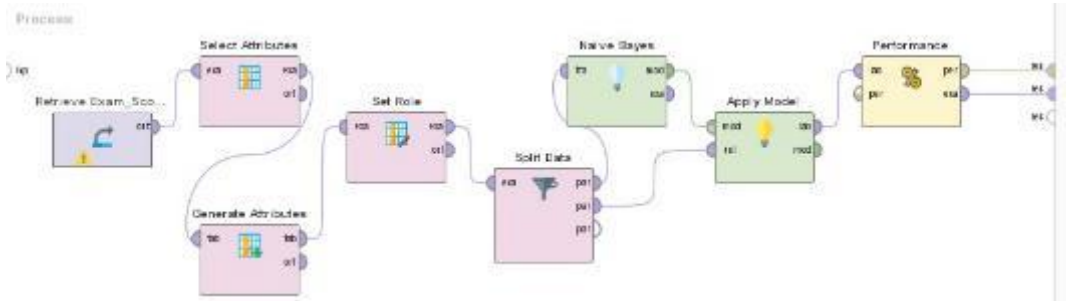
PerformanceVector

```
PerformanceVector:
accuracy: 78.94%
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  579     173
Bagus:  188     774
classification_error: 21.06%
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  579     173
Bagus:  188     774
weighted_mean_recall: 78.61%, weights: 1, 1
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  579     173
Bagus:  188     774
weighted_mean_precision: 78.73%, weights: 1, 1
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  579     173
Bagus:  188     774
```

Gambar 5. *Performance Vector KNN*

https://doi.org/10.30872/atasi.v5i2.4652

2. Penerapan Metode *Naïve Bayes Classifier*



Gambar 6. Proses *Klasifikasi Naive Bayes*

accuracy: 83.02%

	true Jelek	true Bagus	class precision
pred. Jelek	618	142	81.32%
pred. Bagus	149	805	84.38%
class recall	80.57%	85.01%	

Gambar 7. *Confusion Matrix Naïve Bayes*

PerformanceVector

```
PerformanceVector:
accuracy: 83.02%
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  618     142
Bagus:  149     805
classification_error: 16.98%
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  618     142
Bagus:  149     805
weighted_mean_recall: 82.79%, weights: 1, 1
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  618     142
Bagus:  149     805
weighted_mean_precision: 82.85%, weights: 1, 1
ConfusionMatrix:
True:   Jelek   Bagus
Jelek:  618     142
Bagus:  149     805
```

Gambar 8. *Performance Vector Naïve Bayes*

Berdasarkan hasil classification error, metode Naïve Bayes memiliki tingkat kesalahan klasifikasi sebesar 16,98%, sedangkan KNN sebesar 21,06%. Hal ini menunjukkan bahwa Naïve Bayes mampu mengurangi kesalahan prediksi dibandingkan KNN. Kesalahan klasifikasi pada kedua metode kemungkinan disebabkan oleh adanya kemiripan karakteristik antar kelas mahasiswa, terutama pada data dengan nilai skor ujian yang berada di sekitar batas kategori "Jelek" dan "Bagus". Selain itu, beberapa faktor non-akademik seperti kualitas tidur dan fasilitas belajar memiliki pola yang tidak selalu konsisten terhadap hasil skor ujian sehingga menyebabkan model mengalami kesulitan dalam membedakan kelas secara sempurna.

D. Pengukuran dan perbandingan Performa Metode KNN dan *Naïve Bayes Classifier*

Setelah tahap evaluasi kinerja selesai dilakukan, langkah berikutnya adalah membandingkan performa model klasifikasi yang menggunakan metode KNN dan *Naïve Bayes Classifier*. Perbandingan ini bertujuan untuk mengidentifikasi metode yang paling unggul dalam menangani klasifikasi skor ujian mahasiswa berdasarkan faktor akademik maupun non-akademik. Hasil dari perbandingan kedua metode tersebut disajikan pada tabel berikut.

Tabel 4. Perbandingan

Metode	*Accuracy*	*Precision*	*Recall*	*Classification Error*
Naïve Bayes	83.02%	82.85%	82.79%	16.98%
KNN	78.94%	78.73%	78.61%	21.06%

Berdasarkan hasil evaluasi kinerja yang telah dilakukan, keunggulan *Naïve Bayes* dalam penelitian ini disebabkan oleh kemampuannya dalam memodelkan hubungan probabilistik antar atribut secara efisien, sehingga lebih stabil dalam menangani data dengan variasi atribut akademik dan non-akademik. Dalam artian *Naïve Bayes* memperoleh performa lebih baik dibandingkan KNN karena mampu menangani kombinasi atribut kategorikal dan numerik secara lebih stabil.

Meskipun algoritma *Naïve Bayes* memperoleh performa terbaik dalam penelitian ini, metode tersebut masih memiliki beberapa keterbatasan. Naïve Bayes menggunakan asumsi independensi antar atribut, padahal pada

https://doi.org/10.30872/atasi.v5i2.4652

kondisi nyata beberapa variabel akademik dan non-akademik dapat saling berkaitan. Sementara itu, metode KNN memiliki kelemahan pada sensitivitas terhadap distribusi data dan pemilihan nilai K yang digunakan. Selain itu, penggunaan dataset sekunder juga menyebabkan penelitian bergantung pada kualitas data yang tersedia sehingga masih terdapat kemungkinan bias data dan ketidakseimbangan distribusi kelas.

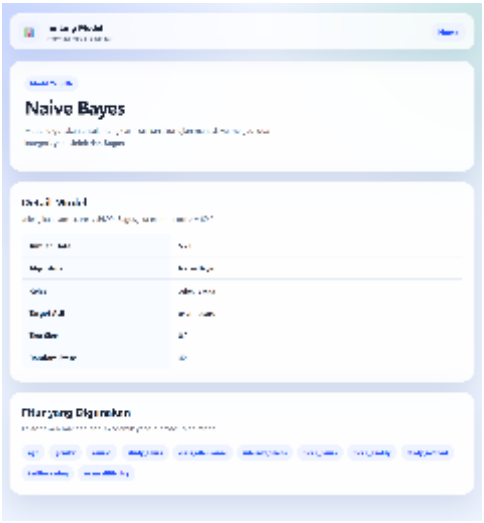
Disamping itu hasil ini menunjukkan bahwa keberhasilan akademik mahasiswa tidak hanya dipengaruhi oleh faktor pembelajaran formal seperti jam belajar dan kehadiran kelas, tetapi juga dipengaruhi oleh kondisi non-akademik yang mendukung kesiapan belajar mahasiswa secara menyeluruh.

E. Pembuatan Aplikasi

Aplikasi ini dibangun menggunakan bahasa pemrograman Python dengan framework Flask.

1. Model Aplikasi Klasifikasi Skor Ujian Mahasiswa

Pada halaman Tentang Model, aplikasi menampilkan informasi mengenai model yang digunakan. Informasi tersebut meliputi jumlah data, algoritma yang digunakan, kelas klasifikasi, target asli, ukuran data uji, *random state*, serta fitur-fitur yang digunakan dalam proses prediksi.

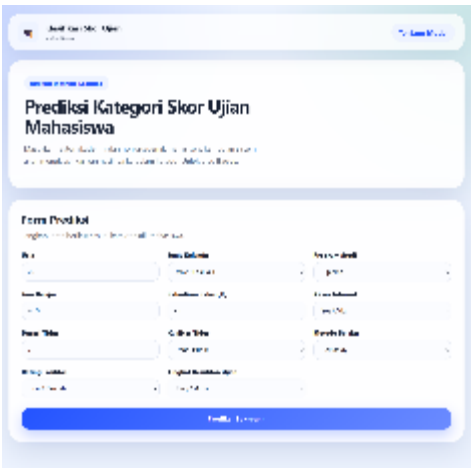


Gambar 9. Tampilan Informasi Model

Berdasarkan Gambar 9, halaman informasi model berfungsi untuk memberikan gambaran kepada pengguna mengenai model klasifikasi yang digunakan dalam aplikasi. Halaman ini juga menunjukkan bahwa aplikasi tidak hanya menampilkan hasil prediksi, tetapi juga memberikan informasi mengenai atribut yang diproses oleh model.

2. Tampilan Form Prediksi

Tampilan form prediksi merupakan halaman utama aplikasi yang digunakan oleh pengguna untuk memasukkan data mahasiswa. Pada halaman ini, pengguna diminta mengisi beberapa data yang berkaitan dengan faktor akademik dan non-akademik mahasiswa.

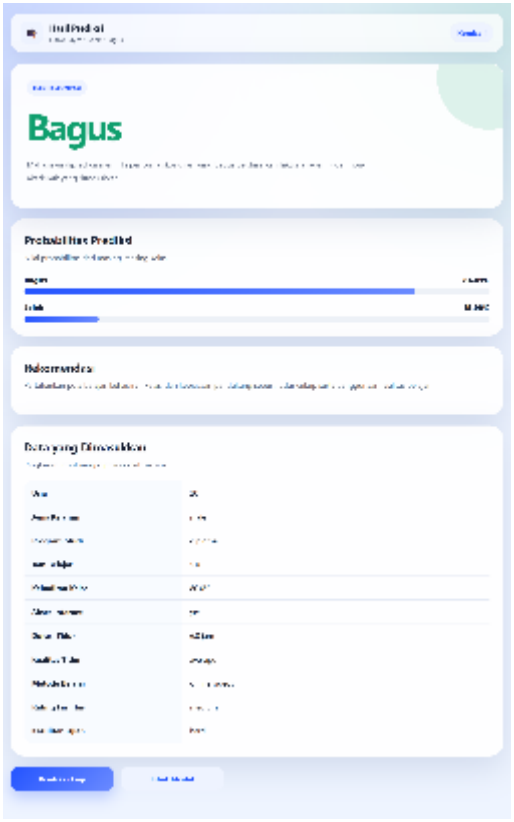


Gambar 10. Tampilan Form Prediksi

Data yang harus dimasukkan pada form prediksi meliputi usia, jenis kelamin, program studi, jam belajar, kehadiran kelas, akses internet, durasi tidur, kualitas tidur, metode belajar, rating fasilitas, dan tingkat kesulitan ujian. Setelah seluruh data dimasukkan, pengguna dapat menekan tombol Prediksi Sekarang untuk memproses data tersebut.

https://doi.org/10.30872/atasi.v5i2.4652

3. Tampilan Hasil Prediksi Kategori Bagus



Gambar 11. Tampilan Hasil Prediksi Kategori Bagus

Berdasarkan Gambar 11, hasil klasifikasi menunjukkan kategori Bagus dengan probabilitas sebesar 84,01%, sedangkan probabilitas kategori Jelek sebesar 15,99%. Hasil tersebut menunjukkan bahwa model klasifikasi yang diterapkan mampu memberikan prediksi dengan tingkat keyakinan yang lebih dominan pada kategori Bagus. Dengan demikian, Sistem mampu menerima input data pengguna dan menampilkan hasil prediksi sesuai model yang telah dibangun. Selain itu, fokus utama penelitian ini terletak pada perbandingan performa algoritma klasifikasi, sehingga evaluasi sistem aplikasi dilakukan secara fungsional untuk memastikan proses prediksi berjalan dengan baik.

5. KESIMPULAN

Berdasarkan hasil pengujian, algoritma *Naïve Bayes* menunjukkan performa lebih baik dibandingkan KNN dalam klasifikasi skor ujian mahasiswa. Naïve Bayes memperoleh accuracy 83,02%, precision 82,85%, recall 82,79%, dan classification error 16,98%, sedangkan KNN memperoleh accuracy 78,94%, precision 78,73%, recall 78,61%, dan classification error 21,06%. Hasil tersebut menunjukkan bahwa Naïve Bayes lebih efektif karena memiliki tingkat akurasi dan ketepatan yang lebih tinggi serta kesalahan klasifikasi yang lebih rendah. Model terbaik kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan Python dan Flask yang mampu menerima input faktor akademik dan non-akademik serta menampilkan hasil prediksi kategori Jelek atau Bagus dengan baik sesuai fungsi yang diharapkan. Penelitian selanjutnya disarankan untuk melakukan analisis *feature importance* atau seleksi fitur guna mengetahui variabel akademik maupun non-akademik yang paling dominan memengaruhi hasil klasifikasi. Selain itu, penelitian berikutnya juga dapat mengembangkan model menggunakan metode *ensemble* seperti *Random Forest* atau *XGBoost* agar diperoleh performa klasifikasi yang lebih optimal.

6. DAFTAR PUSTAKA

Amra, I. A. A., & Maghari, A. Y. A. (2017). Students performance prediction using KNN and naïve Bayesian. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 909–913).

Aram, M. (2024). *Strategi pengelolaan akademik dan penjaminan mutu di perguruan tinggi*. Penerbit Widina. https://repository.penerbitwidina.com/publications/585310/strategi-pengelolaan-akademik-dan-penjaminan-mutu-di-perguruan-tinggi

Arumdani, A. L. (2025). Strategi organisasi dalam meningkatkan prestasi (akademik dan non-akademik) mahasiswa. *Jurnal Penelitian Ilmu-Ilmu Sosial, 3*(2), 214–223.

Efendi, A., & Dahlia. (2025). Pengaruh aspek kognitif dan afektif terhadap prestasi akademik mahasiswa. *MAMEN: Jurnal Manajemen, 4*(3), 267–277. https://doi.org/10.55123/mamen.v4i3.5287

Feri, S. L. N. (2023). *Basic data mining from A to Z*. Penerbit

Widina. https://books.google.co.id/books/about/Basic_Data_Mining_from_A_to_Z.html?id=JcLhEAAAQBAJ

Gopinda, M. (2026). Studi perbandingan algoritma machine learning untuk prediksi penjualan retail. *Jurnal Ilmu Komputer dan Teknologi Informasi*, *1*(1), 22–28. https://ojs.feliciajurnal.com/index.php/jikti/article/view/9

Lumley, M., & Wilkinson, J. (2014). Teorema naïve Bayes. In *Developing employability for business*. Oxford University Press.

Munarsih, & Ningsi, B. A. (2023). Performance comparison of data mining classification algorithms on student academic achievement prediction. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM), 6*(1), 29–39.

Muthohharoh, N., Lidimilah, L. F., & Homaidi, A. (2025). Perbandingan algoritma naïve Bayes dan K-nearest neighbor untuk mengklasifikasikan status kesehatan. *Karsanusa: Jurnal Pengabdian dan Riset Kesehatan Masyarakat, 1*(1), 24–32. https://journalng.uwks.ac.id/karsanusa/article/view/512

Nooraeni, R., & Nurfalah, G. (2022). Kajian penerapan jarak Euclidean, Manhattan, Minkowski, dan Chebyshev pada algoritma clustering K-prototype. *Sains, Aplikasi, Komputasi dan Teknologi Informasi, 4*(2), 72–82.

Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi menggunakan metode naive Bayes untuk menentukan calon penerima PIP. *JATI (Jurnal Mahasiswa Teknik Informatika), 7*(1), 452–458. https://doi.org/10.36040/jati.v7i1.6303

Rohman, N., & Wibowo, A. (2024). Perbandingan metode K-medoids dan metode K-means dalam analisis segmentasi pelanggan mall. *SINTECH (Science and Information Technology) Journal, 7*(1), 49–58. https://doi.org/10.31598/sintechjournal.v7i1.1507

Runimeirati, Muis, A., & Muhammad, F. (2023). Pelatihan text mining menggunakan bahasa pemrograman Python. *ABDIMAS LANGKANAE: Jurnal Pengabdian kepada Masyarakat*, 36–46.

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analysis data mining Netflix data using the Rapid Miner application. *Journal of Business and Audit Information Systems, 4*(1), 13–21.

Warinda. (2025). *Keterlambatan penyelesaian studi mahasiswa hukum keluarga Islam fakultas syariah di IAIN Palopo* [Skripsi, IAIN Palopo]. Repository IAIN Palopo. https://repository.uinpalopo.ac.id/id/eprint/11231/

Watratan, A. F., Puspita B, A., & Moeis, D. (2020). Implementation of the naive Bayes algorithm to predict the spread of Covid-19 in Indonesia. *Journal of Applied Computer Science and Technology, 1*(1), 7–14.

Wijaya, A., & Bismi, W. (2025). Penerapan algoritma machine learning dalam mengklasifikasi data masa studi di Indonesia berdasarkan jenis kelamin. *Journal of Information Engineering and Educational Technology, 8*(2), 62–74. https://doi.org/10.26740/jieet.v8n2.p62-74

Yuliansyah, M. R., Muslimin, B., & Franz, A. (2022). Perbandingan metode K-nearest neighbors dan naïve Bayes classifier pada klasifikasi status gizi balita di Puskesmas Muara Jawa Kota Samarinda. *Adopsi Teknologi dan Sistem Informasi (ATASI), 1*(1), 8–20. https://doi.org/10.30872/atasi.v1i1.25

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine, 4*(11), 1–7. https://doi.org/10.21037/atm.2016.03.37

Zulkifli, R., Andryadi, A. A., Nurfadhilah, D. S., & Utami, L. L. (2024). Penerapan algoritma naive Bayes dalam memprediksi minat mahasiswa baru. *JOISM: Journal of Information System Management, 5*(1), 1–10. https://jurnal.amikom.ac.id/index.php/joism/article/view/2375

https://doi.org/10.30872/atasi.v5i2.4652