

# Classification of Post-School Tendencies of Madrasah Aliyah Students Using the K-Nearest Neighbor Algorithm

Eka Asfira Atha Yanuha Fiftani<sup>1\*</sup>, Ifnu Wisma Dwi Prastya<sup>2\*</sup>, Ita Aristia Sa'ida<sup>3\*</sup>

\* Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro  
[ekaasfira592@gmail.com](mailto:ekaasfira592@gmail.com)<sup>1</sup>, [ifnuprastya@unugiri.ac.id](mailto:ifnuprastya@unugiri.ac.id)<sup>2</sup>, [itaaristia@unugiri.ac.id](mailto:itaaristia@unugiri.ac.id)<sup>3</sup>

## Article Info

### Article history:

Received 2026-01-26

Revised 2026-03-31

Accepted 2026-04-10

### Keyword:

Data Mining,  
K-Nearest Neighbor,  
Classification,  
Madrasah Aliyah,  
Graduation Prediction.

## ABSTRACT

This research develops a classification model to identify post-graduation tendencies of Madrasah Aliyah students using the K-Nearest Neighbor (KNN) algorithm with academic report card scores as input features. The dataset includes 76 students, using average scores from semesters 1 to 5 as predictors and students' post-school tendencies as the target variable. Data preprocessing involved normalization and splitting the dataset into training and testing subsets, while similarity between instances was measured using Euclidean Distance with  $k = 5$ . The experimental results achieved an accuracy of 87.50%, indicating that KNN performs well on small-scale academic datasets. This study contributes by specifically applying KNN to classify post-school tendencies in a Madrasah Aliyah context using limited academic features, an area that has not been extensively explored in previous educational data mining studies. The proposed model can assist schools in providing data-driven academic counseling and decision support within the Madrasah Aliyah environment.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Pesatnya kemajuan teknologi informasi telah memengaruhi berbagai dimensi kehidupan manusia, salah satunya pada bidang pendidikan[1]. Pemanfaatan teknologi tidak hanya mempermudah aktivitas manusia, tetapi juga berperan penting dalam mendukung proses pengambilan keputusan yang dapat dilakukan dengan lebih efisien dan tepat[2]. Melalui pengolahan data yang tepat, berbagai permasalahan dapat dianalisis dan diprediksi guna meningkatkan efektivitas suatu sistem[3]. Pendidikan merupakan bidang yang memiliki peran penting dalam pengembangan mutu sumber daya manusia Indonesia memiliki kewajiban untuk menempuh pendidikan formal sebagai upaya memperoleh pengetahuan dan keterampilan[4]. Selama ini, evaluasi kelulusan murid umumnya didasarkan pada nilai rapor tanpa didukung oleh analisis pola data secara menyeluruh[5].

Selain itu, belum semua sekolah memiliki aplikasi yang mampu memberikan gambaran prediksi kelulusan murid secara objektif[6]. Kondisi ini menyebabkan proses evaluasi belum sepenuhnya optimal dalam membantu guru dan pihak sekolah mengambil keputusan yang tepat terkait kesiapan

akademik murid[7]. Salah satu pendekatan yang dapat diterapkan untuk menangani permasalahan tersebut adalah penerapan metode klasifikasi berbasis data mining[8]. Melalui penerapan teknik klasifikasi, data akademik murid dapat dianalisis untuk memprediksi kecenderungan kelulusan berdasarkan variabel-variabel tertentu[9]. Data Mining juga bertujuan untuk menemukan pola, hubungan, dan informasi penting yang perlu dicapai agar dapat dipakai untuk meramalkan keputusan ke depannya[10]. Penerapan data mining salah satunya dilakukan melalui metode klasifikasi yang memanfaatkan algoritma tertentu guna mengelompokkan data sesuai dengan pola yang terbentuk [11].

Untuk membantu mengurangi kesalahan dalam proses penentuan kelulusan murid maka penulis memberikan saran melalui tugas akhir ini dengan mengusulkan penerapan data mining untuk memprediksi kelulusan murid menggunakan Algoritma *K-Nearest Neighbor* (KNN). Dalam penelitian ini penulis melakukan pengembangan sistem yang dilakukan dengan mengimplementasikan suatu algoritma *K-Nearest Neighbour* (KNN) guna untuk mengetahui tingkat kelulusan setiap murid berdasarkan hasil rata-rata perolehan nilai murid. Algoritma *K-Nearest Neighbor* (KNN) merupakan

salah satu metode klasifikasi yang banyak digunakan karena konsepnya yang sederhana dan kemampuannya dalam mengelompokkan data berdasarkan kedekatan jarak antar data [12],[13].

Berdasarkan beberapa penelitian, algoritma KNN mampu memberikan hasil klasifikasi yang baik dengan tingkat akurasi, *presisi*, dan *recall* yang tinggi[14]. Pada penelitian sebelumnya yang memprediksi dengan menerapkan algoritma *Naive Bayes* pada data mining yang bertujuan untuk memprediksi kemampuan lulusan SMK untuk bersaing di pasar kerja [15]. Dengan nilai akurasi 98% dan nilai *Area Under Curve* (AUC) sebesar 0,980, penelitian ini menunjukkan hasil yang baik[16]. Pada penelitian sebelumnya yang memprediksi dengan menerapkan algoritma KNN mendukung institusi pendidikan meningkatkan angka kelulusan melalui upaya perbaikan yang tepat sasaran berdasarkan data akademik dan non-akademi[17].

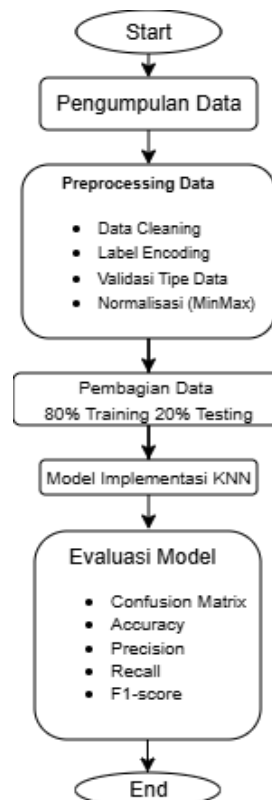
Penelitian ini bertujuan untuk mengklasifikasikan murid MA Miftahul Ulum setelah lulus sekolah menggunakan algoritma KNN[18], model klasifikasi ini diharapkan membantu pihak sekolah terutama waka Kesiswaan, waka Akademik, dan guru Bimbingan Konseling dalam memberikan arahan bimbingan karier berdasarkan data seluruh murid[19]. Ruang lingkup penelitian ini mencakup data murid kelas XII di MA Miftahul Ulum selama 4 tahun terakhir [20]. Penelitian ini diharapkan bisa mendukung sekolah dalam mengenali pola kecenderungan murid pasca lulus, agar dapat dijadikan acuan dalam merancang program bimbingan karier dan pendampingan bagi murid [21]. Oleh karena itu, penelitian ini menerapkan algoritma KNN untuk mengklasifikasikan kecenderungan murid Madrasah Aliyah pasca sekolah berdasarkan data nilai rapor, sehingga diharapkan dapat membantu pihak sekolah dalam mendukung proses evaluasi akademik dan pengambilan keputusan berbasis data.

## II. METODE

Pada penelitian ini digunakan algoritma *K-Nearest Neighbor* (KNN) sebagai metode untuk melakukan proses klasifikasi kelulusan murid pasca lulus sekolah, setelah dilakukan pengujian untuk mengetahui tingkat akurasi dengan beberapa algoritma yaitu *Artificial Neural Network* (ANN), *C4.5*, *Random Forest*, *K-Nearest Neighbor* (KNN), *XGBoost*, dan *Logistic Regression*.

Selanjutnya, dilakukan proses perbandingan performa antar algoritma untuk menentukan model yang paling optimal dalam melakukan klasifikasi. Setiap algoritma diuji menggunakan dataset yang sama dengan skema pembagian data latih dan data uji yang seragam, sehingga hasil evaluasi dapat dibandingkan secara objektif. Parameter evaluasi yang digunakan meliputi akurasi sebagai indikator utama, serta didukung oleh metrik lain seperti *presisi*, *recall*, dan *F1-score* untuk memberikan gambaran yang lebih komprehensif terhadap kinerja model. Hasil dari proses pengujian ini kemudian digunakan sebagai dasar dalam menentukan

algoritma terbaik yang akan diimplementasikan pada tahap klasifikasi kecenderungan murid pasca lulus sekolah.



Gambar 1. Flowchart Metode KNN

### A. Pengambilan Data Murid

Pada tahap ini dilakukan proses awal pengambilan data murid MA Miftahul Ulum, data yang digunakan berasal dari nilai rata-rata semester 1 sampai dengan semester 5, seperti gambar diatas. Selain data nilai, juga ada data tambahan seperti pekerjaan orang tua dan kecenderungan murid setelah lulus, misalnya apakah mereka menentukan untuk kuliah, kerja, mondok, ataupun menikah.

TABEL I  
ATRIBUT DATA

Atribut	Tipe	Deskripsi
SEM 1	Numeric	Nilai rapor semester 1
SEM 2	Numeric	Nilai rapor semester 2
SEM 3	Numeric	Nilai rapor semester 3
SEM 4	Numeric	Nilai rapor semester 4
SEM 5	Numeric	Nilai rapor semester 5
Target	Categorical (Label)	Kecenderungan murid: Kuliah/Kerja/Nikah.dll

### B. Preprocessing Data

Setelah proses pengumpulan data selesai, langkah berikutnya adalah prosesing data atau pengolahan awal. Pada tahap ini, data awal akan diseleksi dan dipersiapkan sehingga dapat digunakan oleh model algoritma. Pada tahap ini dilakukan proses pra-pemrosesan data yang meliputi *data cleaning*, *label encoding*, validasi tipe data, dan normalisasi menggunakan metode *Min-Max Normalization*.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan :

$X$  : nilai asli

$X_{min}$  : nilai minimum atribut

$X_{max}$  : nilai maksimum atribut

$X'$  : nilai hasil normalisasi

### C. Data Cleaning

Pada fase pembersihan data, dilakukan upaya untuk membersihkan informasi, seperti menghapus entri yang kosong, memperbaiki informasi yang ganda, serta memastikan setiap kolom memiliki nilai yang terisi sepenuhnya. Proses ini bertujuan untuk menghindari kesalahan ketika proses pemodelan dilakukan.

### D. Data Latih dan Data Uji

Data Latih (Training Data) digunakan untuk melakukan pemodelan KNN untuk mengidentifikasi pola dari data dengan jumlah Data Latih 80%. Data Uji (Testing Data) digunakan untuk mengukur seberapa akurat model dalam melakukan klasifikasi terhadap data baru dengan jumlah Data Uji 20%.

### E. Pemilihan Algoritma

Tahap pemodelan dalam penelitian ini melibatkan penggunaan algoritma pembelajaran mesin untuk menganalisis pola data dan menghasilkan prediksi kecenderungan murid setelah lulus sekolah. Beberapa algoritma klasifikasi diterapkan sebagai pembanding, antara lain *Random Forest*, *K-Nearest Neighbor* (KNN), *Extreme Gradient Boosting* (XGBoost), dan *Regresi Logistik*. Seluruh algoritma diuji menggunakan dataset dan skema pembagian data yang sama agar hasil evaluasi dapat dilakukan secara objektif. Berdasarkan hasil evaluasi, algoritma KNN memiliki kinerja paling optimal dengan tingkat akurasi tertinggi, sehingga dipilih sebagai model utama dalam penelitian ini. Selain memiliki akurasi yang baik, KNN juga memiliki keunggulan dalam kesederhanaan implementasi dan tidak memerlukan asumsi khusus terhadap distribusi data[22]. KNN dipilih sebagai model utama karena dataset penelitian relatif kecil dan seluruh fitur berupa nilai numerik. Pendekatan berbasis jarak pada KNN lebih sesuai dibanding algoritma ensemble atau boosting yang biasanya memerlukan dataset lebih besar. Selain itu, hasil pengujian menunjukkan KNN memiliki akurasi tertinggi dibanding *Random Forest*, *XGBoost*, dan *Regresi Logistik*.

### F. Algoritma Klasifikasi

Pada fase pemodelan, penelitian ini menggunakan berbagai algoritma klasifikasi untuk menilai keberhasilan model dalam mengidentifikasi kecenderungan murid pasca sekolah. Algoritma yang diterapkan mencakup *K-Nearest Neighbor* (KNN), *Random Forest*, *Extreme Gradient Boosting* (XGBoost), dan *Regresi Logistik*. Tujuan dari perbandingan ini adalah untuk menemukan model dengan

kinerja unggul yang diukur berdasarkan akurasi, *presisi*, *recall*, dan *F1-score*. Dataset yang sama diterapkan pada semua algoritma agar evaluasi hasilnya objektif dan adil.

#### 1) Algoritma Random Forest

Random Forest merupakan algoritma klasifikasi berbasis teknik *ensemble learning* yang menggabungkan beberapa pohon keputusan yang dibangun secara acak. Setiap pohon menghasilkan prediksi kelas, kemudian hasil akhir ditentukan berdasarkan mayoritas prediksi dari seluruh pohon yang terbentuk. Pendekatan ini memungkinkan *Random Forest* untuk mengurangi risiko overfitting serta meningkatkan stabilitas model dalam melakukan klasifikasi, terutama pada data yang memiliki variasi tinggi[23].

$$\hat{y} = \text{mode} \{h_1(x), h_2(x), \dots, h_n(x)\}$$

Keterangan:

$h_n(x)$  : hasil prediksi dari pohon keputusan ke- $n$

$\hat{y}$  : kelas hasil prediksi akhir

#### 2) Algoritma Extreme Gradient Boosting (XGBoost)

*Extreme Gradient Boosting* atau *XGBoost* adalah pengembangan dari metode *gradient boosting* yang dirancang untuk meningkatkan performa dan efisiensi komputasi. Algoritma ini membangun model secara bertahap dengan memperbaiki kesalahan yang dihasilkan oleh model sebelumnya. *XGBoost* menggunakan regularisasi untuk mengontrol kompleksitas model sehingga performa prediksinya menjadi lebih optimal, terutama pada dataset berukuran menengah[24].

Fungsi Objektif *XGBoost*:

$$Obj = \sum_{i=1}^n \iota(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Dengan fungsi regularisasi

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|_2$$

Keterangan:

$\iota(y_i, \hat{y}_i)$  : fungsi loss

$\Omega(f_k)$  : fungsi regularisasi

$T$  : jumlah daun

$w$  : bobot daun

#### 3) Algoritma Regresi Logistik

*Regresi Logistik* merupakan metode klasifikasi statistik yang digunakan untuk memodelkan probabilitas suatu data termasuk ke dalam kelas tertentu. Algoritma ini memanfaatkan fungsi sigmoid untuk memetakan hasil perhitungan linier ke dalam rentang nilai antara 0 dan 1. *Regresi Logistik* banyak digunakan karena proses perhitungannya yang relatif sederhana serta kemampuannya dalam memberikan interpretasi hasil klasifikasi secara jelas[25].

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

Keterangan:

$\beta_0$ : konstanta

$\beta_n$ : koefisien variabel

$x_n$ : atribut ke- $n$

#### 4) Algoritma K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan metode pengelompokan yang digunakan untuk menetapkan kategori pada suatu data berdasarkan tingkat kemiripannya dengan data lain yang memiliki karakteristik serupa. Proses klasifikasi dilakukan dengan mengukur kedekatan antara data uji dan data latih menggunakan ukuran jarak tertentu, seperti *Euclidean Distance*, kemudian menetapkan kelas berdasarkan mayoritas tetangga terdekat. Pemilihan nilai  $K$  berperan penting dalam menentukan kinerja model, karena nilai yang tidak sesuai dapat mempengaruhi tingkat akurasi hasil klasifikasi [26].

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

Keterangan:

$d(x, y)$ : jarak antara data latih dan data uji

$x_i$  : nilai atribut ke- $i$  pada data latih

$y_i$  : nilai atribut ke- $i$  pada data uji

$n$  : jumlah atribut

#### G. Evaluasi

Setelah proses pelatihan model selesai dilakukan, tahap evaluasi diterapkan untuk mengetahui kemampuan algoritma dalam mengklasifikasikan kecenderungan murid secara tepat berdasarkan data yang digunakan. Penilaian kinerja model dilakukan dengan mengacu pada perbandingan hasil prediksi model terhadap data sebenarnya yang dianalisis menggunakan *Confusion Matrix*, dapat dihitung metrik evaluasi seperti akurasi, *presisi*, dan *recall* untuk mengetahui sejauh mana model mampu melakukan klasifikasi secara tepat. Evaluasi ini digunakan sebagai dasar dalam menentukan efektivitas algoritma KNN sebagai model utama dalam penelitian klasifikasi. Mengingat ukuran dataset yang relatif kecil (76 data), potensi *overfitting* menjadi perhatian dalam penelitian ini. Oleh karena itu, interpretasi hasil dilakukan secara hati-hati dan direkomendasikan penggunaan *cross-validation* pada penelitian lanjutan.

#### H. Aspek Etika Penelitian

Pada penelitian ini, pemanfaatan data akademik siswa dilakukan dengan memperhatikan prinsip etika penelitian serta perlindungan kerahasiaan informasi. Data yang dianalisis telah melalui proses penyamaran identitas dengan menghapus unsur-unsur yang dapat mengungkap identitas pribadi, seperti nama maupun informasi sensitif lainnya.

Seluruh data digunakan secara terbatas hanya untuk keperluan analisis dan pengembangan model penelitian,

tanpa menampilkan atau mengaitkan hasil dengan individu tertentu. Dengan pendekatan tersebut, penelitian ini tetap berpegang pada prinsip tanggung jawab dan keamanan dalam pengelolaan data pendidikan.

### III. HASIL DAN PEMBAHASAN

#### A. Pengambilan Data

Dataset penelitian ini berasal dari data nilai raport murid MA Miftahul Ulum 5 semester lulusan 4 tahun terakhir, *sampel* yang digunakan berjumlah 76 data, dengan atribut berupa nilai raport setiap semester. Seleksi atribut bertujuan untuk menjamin bahwa informasi yang dipakai tidak mencakup data identitas pribadi murid, seperti nama, alamat, atau nomor identifikasi murid. Maka dari itu, fitur yang digunakan hanya berupa nilai akademik dari Semester 1 hingga Semester 5. Sementara itu, atribut kelas (label) yang diterapkan adalah kecenderungan murid setelah menyelesaikan studinya, yakni Kuliah, Kerja, atau Nikah dll. Variabel target "kecenderungan pasca sekolah" didefinisikan sebagai pilihan utama murid setelah lulus, yang meliputi kategori melanjutkan ke perguruan tinggi (Kuliah), memasuki dunia kerja (Kerja), melanjutkan ke pondok pesantren (Mondok), atau menikah. Kategori ini diperoleh berdasarkan data aktual dari pihak sekolah.

TABEL II  
DATASET

Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Target
77	77	74	79	79	Kerja
77	78	78	79	79	Kerja
77	77	77	78	80	Kerja
...	...	...	...	...	...
80	80	79	79	79	Kerja
84	84	83	83	84	Kuliah
88	88	86	87	85	Kerja

#### B. Preprocessing

Persiapan merupakan langkah untuk menyiapkan data agar dapat diproses dan digunakan dalam algoritma KNN. Proses persiapan data yang dilakukan meliputi pembentukan atribut baru dan juga modifikasi isi dari atribut sesuai dengan algoritma yang hendak diterapkan. Pra-pemrosesan data dilakukan untuk mempersiapkan kumpulan data agar siap untuk tahap modeling. Kumpulan data yang awal mencakup nama murid, nilai dari semester 1 sampai semester 5, pekerjaan orang tua, dan kecenderungan setelah sekolah. Atribut berupa nama murid dan pekerjaan orang tua dihapus karena mengandung informasi sensitif dan tidak berpengaruh pada klasifikasi. Kumpulan data tersebut kemudian difokuskan pada lima nilai semester sebagai fitur dan kecenderungan setelah sekolah sebagai variabel target. Variabel target yang semula berbentuk kategorikal dikonversi ke dalam bentuk numerik menggunakan label encoding agar dapat diproses oleh algoritma KNN. Selanjutnya, dilakukan validasi tipe data untuk memastikan semua nilai semester berada dalam format numerik. Semua tahapan pra-

pemrosesan dilakukan dengan menggunakan Microsoft Excel sebelum data tersebut digunakan dalam tahap modeling.

TABEL III  
DATA MINMAX

Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Target
0.1538	0.0833	0.0000	0.1818	0.1818	Kerja
0.1538	0.1667	0.3077	0.1818	0.1818	Kerja
0.1538	0.0833	0.2308	0.0909	0.2727	Kerja
...	...	...	...	...	...
0.3846	0.3333	0.3846	0.1818	0.1818	Kerja
0.6923	0.6667	0.6923	0.5455	0.6364	Kuliah
10000	10000	0.9231	0.9091	0.7273	Kerja

Pada tahap *preprocessing* juga dilakukan proses seleksi fitur untuk mengetahui tingkat kontribusi masing-masing atribut terhadap variabel target. Metode yang digunakan adalah uji *Chi-Square* dengan memanfaatkan fungsi *SelectKBest* dari pustaka *scikit-learn*. Teknik ini digunakan untuk mengukur hubungan antara setiap fitur dengan kelas target sehingga dapat diketahui fitur mana yang paling berpengaruh dalam proses klasifikasi. Berdasarkan hasil perhitungan skor *Chi-Square*, fitur SEM 2 memperoleh nilai tertinggi sebesar 1,340989, diikuti oleh SEM 3 sebesar 1,292288, SEM 4 sebesar 1,279839, SEM 5 sebesar 1,242823, dan SEM 1 sebesar 1,008302. Hasil tersebut menunjukkan bahwa seluruh fitur memiliki kontribusi terhadap variabel target, sehingga seluruh atribut tetap digunakan dalam proses pemodelan klasifikasi menggunakan metode KNN.

### C. Pembagian Data Latih dan Data Uji

Data penelitian terbagi menjadi dua kategori, yakni data pelatihan dan data pengujian. Data pelatihan berfungsi untuk membangun model klasifikasi, sedangkan data pengujian dimanfaatkan untuk menilai kinerja model yang telah dibuat. Dalam penelitian ini digunakan skema pembagian data sebesar 80% untuk data latih dan 20% untuk data uji, sehingga dari total 76 data diperoleh 60 data latih dan 16 data uji. Setiap skenario menghasilkan jumlah data pelatihan dan data pengujian yang bervariasi berdasarkan keseluruhan dataset yang digunakan, yaitu sebanyak 76 data.

### D. Tahap Pengujian

TABEL IV  
AKURASI DATA MINMAX

No	Algoritma	Akurasi MinMax
1	KNN	0,81
2	<i>Random Forest</i>	0,75
3	<i>XGBoost</i>	0,75
4	<i>Naïve Bayes</i>	0,43
5	<i>Decision Tree</i>	0,62
6	<i>Logistic Regression</i>	0,81

Pada fase pengujian model klasifikasi, dilakukan evaluasi beberapa algoritma dengan memanfaatkan data yang telah melalui proses normalisasi Min-Max. Proses normalisasi ini

bertujuan untuk menyetarakan rentang nilai setiap atribut sehingga perbedaan skala tidak mempengaruhi efektivitas algoritma. Hasil dari pengujian menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) dan *Logistic Regression* mencapai akurasi tertinggi yaitu 0,81. Hal ini menandakan bahwa kedua algoritma tersebut dapat mengklasifikasikan data dengan keakuratan yang lebih baik dibandingkan algoritma lainnya pada dataset yang telah dinormalisasi dengan metode Min-Max.

Algoritma KNN beroperasi dengan cara menentukan kelas dari sebuah data berdasarkan kedekatannya dengan sejumlah tetangga terdekat di data pelatihan, sehingga normalisasi Min-Max sangat berperan dalam menyetarakan skala atribut untuk mengoptimalkan perhitungan jarak. Di sisi lain, *Logistic Regression* mengklasifikasikan data dengan memodelkan hubungan antara variabel independen dan probabilitas dari suatu kelas melalui fungsi logistik. Selain itu, algoritma *Random Forest* dan *XGBoost* mencatat nilai akurasi sebesar 0,75, yang menunjukkan performa yang cukup baik dalam klasifikasi menggunakan pendekatan ensemble dengan menggabungkan beberapa model pohon keputusan.

Selanjutnya, algoritma *Decision Tree* mendapatkan akurasi sebesar 0,62, yang menunjukkan bahwa model ini masih mampu melakukan klasifikasi meski dengan tingkat ketepatan yang lebih rendah dibandingkan algoritma yang telah disebutkan sebelumnya. Sedangkan algoritma *Naïve Bayes* memperoleh akurasi terendah yaitu 0,43, yang menunjukkan bahwa asumsi independensi antar fitur dalam algoritma ini kurang sesuai dengan karakteristik dataset yang dipakai. Dengan demikian, dapat disimpulkan dari hasil pengujian bahwa penggunaan algoritma KNN pada data yang telah dinormalisasi dengan metode Min-Max memberikan kinerja terbaik dalam studi ini.

Berdasarkan hasil pengujian menggunakan beberapa algoritma klasifikasi, algoritma *K-Nearest Neighbors* menunjukkan performa terbaik dengan tingkat akurasi tertinggi dibandingkan *Random Forest*, *XGBoost*, dan *Logistic Regression* pada dataset penelitian ini. Oleh karena itu, algoritma KNN dipilih sebagai model utama dalam proses klasifikasi kecenderungan murid pasca sekolah.

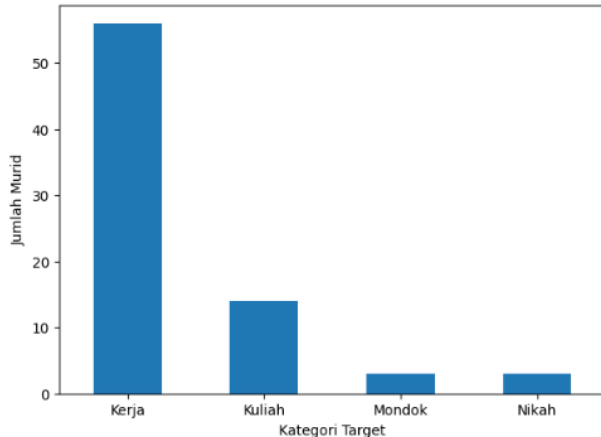
Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma KNN menunjukkan performa paling optimal dalam penelitian ini ketika diterapkan pada data yang telah melalui proses normalisasi menggunakan metode Min-Max.

### E. Distribusi Kelas

Gambar 2 menunjukkan sebuah diagram batang yang menggambarkan penyebaran kecenderungan siswa setelah menyelesaikan pendidikan dasar dalam kategori target yang diperoleh dari data penelitian. Visualisasi ini bertujuan untuk memberikan gambaran awal terkait distribusi data di setiap kelompok yang akan digunakan dalam tahap klasifikasi.

Grafik tersebut, dapat dilihat bahwa mayoritas siswa cenderung memilih untuk bekerja, dengan jumlah sekitar 56 siswa. Di posisi kedua, kategori kuliah mencatat jumlah sekitar 14 siswa. Sementara itu, kategori mondok dan

menikah masing-masing memiliki jumlah yang lebih sedikit, yaitu sekitar 3 siswa.



Gambar 2. Grafik Batang

Penyebaran data ini menunjukkan bahwa sebagian besar lulusan lebih memilih untuk segera bekerja daripada melanjutkan studi atau memilih opsi lain. Variasi jumlah di setiap kategori juga menunjukkan adanya ketidakmerataan kelas dalam dataset. Oleh karena itu, pemahaman tentang distribusi data ini sangat penting sebelum melakukan klasifikasi dengan algoritma KNN, karena distribusi kelas dapat berpengaruh terhadap performa model dalam memberikan prediksi yang tepat.

TABEL V  
SMOTE AKURASI

No	Algoritma	Akurasi Smote
1	KNN	0,77
2	Random Forest	0,82
3	XGBoost	0,75
4	Naive Bayes	0,6
5	Decision Tree	0,75

Berdasarkan hasil visualisasi distribusi kelas pada dataset, terlihat bahwa jumlah data pada setiap kategori target tidak seimbang. Kategori kerja memiliki jumlah data yang jauh lebih banyak dibandingkan kategori lainnya seperti kuliah, mondok, dan nikah. Ketidakseimbangan distribusi kelas ini berpotensi memengaruhi proses pembelajaran model klasifikasi karena algoritma cenderung lebih mudah mengenali kelas dengan jumlah data yang lebih dominan. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*) yang bertujuan untuk menyeimbangkan jumlah data pada kelas minoritas dengan cara menghasilkan sampel sintetis berdasarkan data yang sudah ada.

Namun, hasil pengujian menunjukkan bahwa penerapan SMOTE tidak selalu memberikan peningkatan performa pada seluruh algoritma klasifikasi. Pada algoritma KNN, nilai akurasi yang diperoleh setelah penerapan SMOTE adalah sebesar 0,77, yang lebih rendah dibandingkan dengan akurasi sebelumnya yaitu 0,81 pada data yang hanya melalui proses normalisasi Min-Max. Hal ini menunjukkan bahwa penambahan data sintetis melalui SMOTE dapat

memengaruhi pola kedekatan antar data yang menjadi dasar perhitungan pada algoritma KNN. Karena KNN mengandalkan jarak antar sampel untuk menentukan kelas suatu data, perubahan distribusi akibat penambahan data sintetis dapat menyebabkan proses identifikasi tetangga terdekat menjadi kurang optimal.

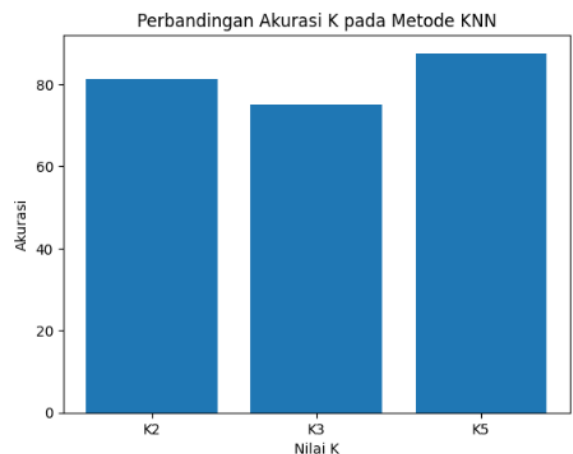
Dengan demikian, meskipun teknik SMOTE efektif dalam mengatasi permasalahan ketidakseimbangan kelas, penerapannya tidak selalu meningkatkan kinerja seluruh algoritma klasifikasi. Hasil penelitian ini menunjukkan bahwa pada algoritma KNN, penggunaan data yang telah melalui proses normalisasi Min-Max tanpa penyeimbangan data menggunakan SMOTE justru memberikan performa yang lebih baik dalam proses klasifikasi kecenderungan murid pasca sekolah.

F. Pemilihan Nilai K

Nilai K adalah elemen yang krusial dalam algoritma KNN karena mengatur jumlah tetangga terdekat yang digunakan untuk penentuan kategori data pengujian. Hasil dari uji coba menunjukkan bahwa tingkat akurasi model bervariasi pada setiap nilai K. Melalui pengujian yang telah dilakukan, akurasi tertinggi didapati pada K = 5, yaitu mencapai 87.50%.

TABEL VI  
AKURASI KNN

K	Akurasi (%)
1	75.00
2	81.25
3	75.00
4	81.25
5	87.50
6	81.25
...	...
15	81.25
16	81.25
17	81.25



Gambar 3. Perbandingan Akurasi KNN Pada Nilai K

Hasil akurasi pada beberapa nilai K dalam metode KNN divisualisasikan menggunakan grafik batang untuk mempermudah analisis performa model. Berdasarkan grafik

tersebut terlihat bahwa nilai K tertentu menghasilkan tingkat akurasi yang lebih tinggi dibandingkan nilai K lainnya.

Berdasarkan Tabel VI, terlihat bahwa nilai  $K = 5$  menghasilkan akurasi tertinggi sebesar 87,50%. Peningkatan akurasi pada K kecil menunjukkan bahwa struktur data relatif homogen, sehingga pendekatan berbasis jarak efektif dalam membedakan kelas. Namun setelah  $K > 5$ , akurasi cenderung stabil, yang mengindikasikan bahwa penambahan jumlah tetangga tidak lagi meningkatkan kemampuan generalisasi model secara signifikan. Nilai ini menunjukkan bahwa lima tetangga terdekat memberikan indikasi yang paling akurat dalam menetapkan kecenderungan murid pada kumpulan data penelitian.

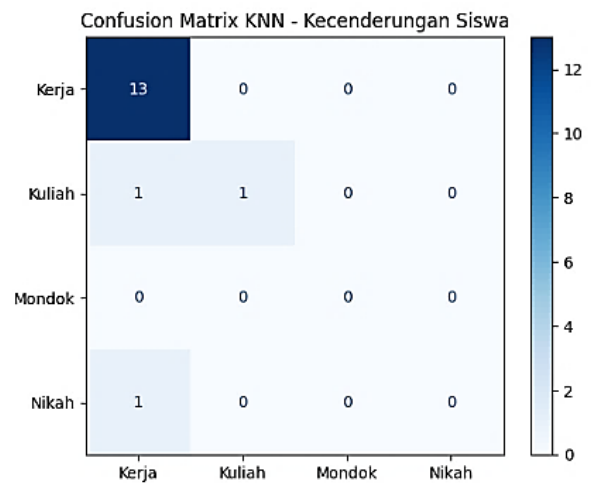
Setelah model KNN dipilih, dilakukan proses penentuan parameter K yang digunakan dalam proses klasifikasi. Pada algoritma KNN, proses klasifikasi dilakukan dengan menghitung jarak antara data baru dengan data latih untuk menemukan sejumlah tetangga terdekat yang memiliki kemiripan paling tinggi berdasarkan nilai rapor. Nilai K menunjukkan jumlah tetangga terdekat yang digunakan sebagai dasar dalam menentukan kelas suatu data. Pada penelitian ini, nilai K dipilih berdasarkan hasil pengujian performa model. Pengujian terhadap metrik jarak lain seperti Manhattan Distance belum dilakukan dan dapat menjadi peluang pengembangan pada penelitian selanjutnya.

#### G. Evaluasi Kinerja dengan Confusion Matrix

TABEL VII  
PRECISION RECALL

	precision	recall	F1-score	support
0	0.87	1.00	0.93	13
1	1.00	0.50	0.67	2
2	0.00	0.00	0.00	1
<b>accuracy</b>			0.88	16
<b>macro avg</b>	0.62	0.50	0.53	16
<b>weighted avg</b>	0.83	0.88	0.84	16

*Confusion Matrix* dimanfaatkan untuk melihat kesesuaian antara hasil prediksi model dan data aktual, sehingga performa klasifikasi dapat dianalisis secara lebih rinci. Setelah mendapatkan nilai K yang terbaik, langkah selanjutnya adalah mengevaluasi kinerja model lebih jauh dengan menggunakan matriks kebingungan untuk menganalisis penyebaran prediksi dari setiap kategori. Hasil dari matriks kebingungan untuk  $K = 5$  dapat ditemukan pada Gambar 4.



Gambar 4. Confusion Matrix

Berdasarkan *Confusion Matrix* tersebut, nilai akurasi, *presisi*, dan *recall* dihitung guna memberikan gambaran performa model klasifikasi yang dibangun dalam penelitian ini:

- 1) Kelas 0 diklasifikasikan dengan sangat baik, menunjukkan tingkat prediksi yang benar mencapai 100%.
- 2) Kelas 1 memperlihatkan akurasi yang sedang, sebab 1 dari 2 data uji diklasifikasikan dengan benar.
- 3) Kelas 3 tidak terdeteksi dengan tepat, hal ini disebabkan oleh jumlah data pada kelas tersebut yang sangat terbatas.

Performa yang rendah pada kelas minoritas menunjukkan bahwa distribusi data yang tidak seimbang berpengaruh signifikan terhadap kemampuan model dalam mengenali kelas dengan jumlah sampel terbatas. Kondisi ini menyebabkan model lebih cenderung memprediksi kelas mayoritas, sehingga meningkatkan akurasi keseluruhan namun menurunkan sensitivitas terhadap kelas minoritas. Fenomena ini umum terjadi pada dataset pendidikan dengan distribusi tidak merata.

Nilai rata-rata tertimbang yang tinggi menunjukkan bahwa model memiliki kinerja yang baik secara keseluruhan, khususnya pada kategori dengan jumlah data yang lebih besar. Sebaliknya, nilai rata-rata makro yang lebih rendah mengindikasikan bahwa performa model pada kategori minoritas masih belum optimal. Distribusi kelas yang tidak seimbang, dengan dominasi kategori Kerja, turut memengaruhi performa model. Ketidakseimbangan tersebut menyebabkan model cenderung bias terhadap kelas mayoritas, yang tercermin pada nilai recall yang rendah pada kelas minoritas.

Hasil *classification report* menunjukkan adanya ketimpangan distribusi data pada variabel target, di mana jumlah sampel pada kelas tertentu jauh lebih banyak dibandingkan kelas lainnya (misalnya 13, 2, dan 1 data). Kondisi ini membuat model cenderung memiliki kinerja lebih baik pada kelas mayoritas dibandingkan kelas minoritas. Dalam penelitian ini, ketidakseimbangan tersebut belum ditangani secara khusus karena keterbatasan jumlah data.

Oleh sebab itu, penelitian selanjutnya disarankan menerapkan teknik penyeimbangan data seperti *oversampling*, *undersampling*, maupun penyesuaian bobot kelas.

#### H. Akurasi Model

Akurasi dari model ditentukan melalui perbandingan antara jumlah prediksi yang akurat dan total data yang diuji. Model KNN dengan  $K = 5$  menghasilkan tingkat akurasi KNN: 87.50%. Meskipun nilai akurasi menunjukkan performa yang baik, ukuran dataset yang relatif kecil yaitu 76 data berpotensi menimbulkan bias evaluasi. Oleh karena itu, penerapan teknik *k-fold cross-validation* direkomendasikan pada penelitian selanjutnya untuk meningkatkan reliabilitas dan stabilitas hasil model.

Ukuran dataset yang relatif kecil juga berpotensi meningkatkan risiko overfitting, di mana model dapat terlalu menyesuaikan pola pada data pelatihan. Temuan ini menunjukkan bahwa algoritma KNN dapat mengklasifikasikan kecenderungan murid setelah sekolah dengan tingkat keakuratan yang cukup tinggi berdasarkan data nilai rapor. Nilai K terbaik dapat dilihat pada Tabel VIII.

TABEL VIII  
K TERBAIK

K	Akurasi	Precision	Recall
1	0.75	0.80	0.92
3	0.75	0.80	0.92
5	0.875	0.87	1.00
7	0.8125	0.81	1.00
9	0.8125	0.81	1.00
11	0.8125	0.81	1.00
13	0.8125	0.81	1.00
15	0.8125	0.81	1.00
17	0.8125	0.81	1.00
19	0.8125	0.81	1.00
21	0.8125	0.81	1.00

#### I. Analisis Pengaruh Fitur Semester

Pada tahap ini dilakukan analisis terhadap pengaruh fitur berdasarkan data nilai pada setiap semester terhadap hasil klasifikasi kecenderungan murid pasca lulus sekolah. Analisis ini bertujuan untuk mengetahui sejauh mana kontribusi masing-masing semester dalam membentuk pola data yang digunakan oleh model klasifikasi. Dengan memahami pengaruh tiap fitur semester, diharapkan dapat diperoleh insight mengenai semester mana yang paling berperan dalam menentukan kecenderungan murid, sehingga dapat menjadi dasar dalam proses seleksi fitur maupun peningkatan performa model.

TABEL IX  
KORELASI FITUR SEMESTER

Semester	Nilai Korelasi
Semester 1	0,111
Semester 2	0,095
Semester 3	0,092
Semester 4	0,095
Semester 5	0,134

Untuk mengetahui semester yang paling berkontribusi terhadap performa model klasifikasi, dilakukan pengujian menggunakan koefisien korelasi *Pearson* antara nilai tiap semester dan variabel target yang sebelumnya telah diubah ke bentuk numerik melalui proses label *encoding*. Berdasarkan hasil pengolahan data, seluruh nilai semester menunjukkan hubungan yang tergolong lemah terhadap variabel target ( $r < 0,2$ ). Walaupun demikian, Semester 5 memiliki nilai korelasi paling tinggi ( $r = 0,134$ ) dibandingkan semester lainnya. Hal ini menunjukkan bahwa nilai pada semester akhir cenderung memiliki peran yang relatif lebih besar dalam membantu proses penentuan kecenderungan murid setelah menyelesaikan pendidikan.

Hasil perbandingan kinerja masing-masing algoritma klasifikasi ditunjukkan pada Tabel X yang memuat nilai akurasi, *precision*, *recall*, dan *F1-score*. Seluruh algoritma diuji menggunakan dataset dan skema pembagian data yang sama sehingga hasil evaluasi yang diperoleh dapat dibandingkan secara objektif.

TABEL X  
PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI

Algoritma	Akurasi (%)	Precision	Recall	F1-Score
<i>K-Nearest Neighbor</i>	87.50%	0.87	1.00	0.93
<i>Random Forest</i>	56,52%	0.65	0.87	0.74
<i>XGBoost</i>	75.00%	0.80	0.92	0.86
<i>Logistic Regression</i>	81.25%	0.81	1.00	0.90

Berdasarkan hasil evaluasi yang ditampilkan pada Tabel X, algoritma *K-Nearest Neighbors* memperoleh nilai akurasi tertinggi dibandingkan metode lain yang diuji, yaitu sebesar 87,50%. Keunggulan KNN pada penelitian ini kemungkinan dipengaruhi oleh karakteristik dataset yang relatif kecil dan homogen, sehingga pendekatan berbasis jarak lebih efektif dalam mengidentifikasi kemiripan antar data dibandingkan metode lain yang membutuhkan variasi data yang lebih kompleks untuk menghasilkan performa optimal. Selain itu, nilai *precision*, *recall*, dan *F1-score* yang tinggi menunjukkan bahwa model mampu mengklasifikasikan data dengan baik serta memiliki tingkat kesalahan klasifikasi yang relatif rendah.

Sementara itu, algoritma *Random Forest* dan *XGBoost* menghasilkan performa yang relatif stabil, namun belum mampu melampaui kinerja KNN pada dataset yang digunakan. Hal ini dapat disebabkan oleh karakteristik kedua algoritma tersebut yang umumnya lebih optimal pada dataset dengan ukuran yang lebih besar dan variasi fitur yang lebih kompleks. Adapun *Logistic Regression* juga menunjukkan performa yang cukup baik dengan nilai akurasi sebesar 81,25%, meskipun masih berada di bawah KNN.

Meskipun model menunjukkan performa yang baik pada dataset yang digunakan, hasil penelitian ini memiliki keterbatasan dalam hal generalisasi. Data yang dianalisis berasal dari satu Madrasah Aliyah dengan karakteristik akademik dan distribusi kelas tertentu, sehingga belum dapat dipastikan bahwa model akan memberikan tingkat akurasi

yang sama apabila diterapkan pada madrasah lain dengan kondisi yang berbeda. Oleh karena itu, pengujian menggunakan dataset dari institusi pendidikan lain diperlukan untuk menguji konsistensi dan daya generalisasi model.

Hasil penelitian ini sejalan dengan beberapa penelitian terdahulu yang menyatakan bahwa algoritma KNN mampu memberikan performa yang kompetitif dalam tugas klasifikasi data numerik. Penelitian tersebut menegaskan bahwa pemilihan parameter yang tepat serta proses prapemrosesan data, seperti normalisasi, memiliki pengaruh signifikan terhadap kinerja KNN meskipun diterapkan pada konteks permasalahan yang berbeda [27].

Jika dibandingkan dengan penelitian tersebut, temuan dalam studi ini menunjukkan bahwa KNN lebih optimal ketika diterapkan pada dataset berukuran relatif kecil dengan fitur numerik yang homogen, seperti data nilai rapor murid. Hal ini juga sejalan dengan penelitian yang membahas perbandingan *Regresi Logistik*, *Random Forest*, dan *XGBoost*, di mana dijelaskan bahwa algoritma yang lebih sederhana dapat menghasilkan performa yang kompetitif apabila karakteristik data sesuai dan proses prapemrosesan dilakukan dengan baik [28].

Dalam konteks *Educational Data Mining*, penelitian ini memperkuat temuan bahwa data akademik seperti nilai rapor dapat dimanfaatkan sebagai indikator prediktif dalam proses klasifikasi dan pengambilan keputusan pendidikan. Sejumlah penelitian dalam bidang ini menunjukkan bahwa algoritma klasifikasi berbasis data numerik sering digunakan untuk memprediksi kelulusan, performa akademik, maupun risiko putus sekolah. Hasil penelitian ini konsisten dengan temuan tersebut, khususnya terkait efektivitas KNN pada dataset berukuran kecil dengan fitur yang relatif homogen.

Namun demikian, berbeda dengan beberapa penelitian *educational data mining* yang mengintegrasikan variabel non-akademik seperti tingkat kehadiran, latar belakang sosial ekonomi, atau aktivitas ekstrakurikuler, penelitian ini masih berfokus pada indikator akademik saja. Perbedaan ini menunjukkan bahwa meskipun model mampu menghasilkan akurasi yang tinggi, potensi pengembangan model berbasis *educational data mining* masih terbuka melalui integrasi variabel yang lebih komprehensif.

Dengan demikian, hasil penelitian ini tidak bertolak belakang dengan penelitian sebelumnya, melainkan memperkuat kesimpulan bahwa pemilihan algoritma klasifikasi perlu disesuaikan dengan karakteristik data dan konteks permasalahan. Penelitian ini memberikan kontribusi tambahan dengan menunjukkan bahwa algoritma KNN mampu menunjukkan performa yang sangat baik dalam proses klasifikasi kecenderungan murid pasca sekolah pada lingkungan Madrasah Aliyah, sehingga dapat mendukung pengambilan keputusan berbasis data di bidang pendidikan.

Penelitian ini memanfaatkan nilai rapor semester 1–5 sebagai indikator utama kinerja akademik murid. Walaupun data tersebut mampu mencerminkan pencapaian belajar, penggunaannya saja belum cukup untuk menggambarkan kecenderungan murid setelah lulus secara menyeluruh.

Berbagai aspek lain seperti minat individu, latar belakang ekonomi keluarga, dukungan orang tua, serta keterlibatan dalam kegiatan ekstrakurikuler turut berpengaruh terhadap pilihan murid pasca sekolah. Oleh karena itu, penelitian selanjutnya disarankan menambahkan variabel non-akademik agar hasil klasifikasi menjadi lebih representatif.

Secara praktis, keluaran dari model klasifikasi ini dapat dimanfaatkan sebagai bahan pertimbangan dalam pelaksanaan layanan bimbingan dan konseling di sekolah. Informasi mengenai kecenderungan pilihan murid setelah lulus dapat memberikan gambaran awal bagi guru Bimbingan Konseling dalam menyusun strategi pendampingan yang lebih spesifik sesuai dengan capaian akademik masing-masing murid. Misalnya, murid yang teridentifikasi memiliki kecenderungan melanjutkan studi ke perguruan tinggi dapat diarahkan pada konsultasi pemilihan jurusan serta persiapan menghadapi seleksi masuk. Sebaliknya, murid yang diproyeksikan memasuki dunia kerja dapat diberikan pembekalan keterampilan praktis maupun informasi terkait peluang kerja yang sesuai.

Selain itu, model klasifikasi yang dikembangkan dalam penelitian ini juga berpotensi untuk diimplementasikan dalam bentuk sistem pendukung keputusan berbasis web atau dashboard interaktif. Integrasi model ke dalam sistem digital memungkinkan pihak sekolah, seperti guru Bimbingan Konseling dan manajemen madrasah, untuk memantau kecenderungan murid secara lebih sistematis serta memperoleh visualisasi hasil prediksi secara informatif.

Dari sisi metodologis, penelitian ini masih memiliki ruang penguatan terutama pada aspek evaluasi model. Meskipun telah dilakukan pembagian data latih dan data uji serta perhitungan metrik evaluasi utama, ukuran dataset yang relatif kecil dapat memengaruhi stabilitas hasil. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menerapkan teknik validasi silang secara menyeluruh, serta mempertimbangkan metode penanganan ketidakseimbangan kelas untuk memperoleh evaluasi yang lebih komprehensif.

Meskipun nilai akurasi yang diperoleh tergolong tinggi, perlu diperhatikan bahwa ukuran dataset yang terbatas dapat mempengaruhi stabilitas model. Pada dataset kecil, perubahan kecil dalam pembagian data latih dan data uji dapat berdampak signifikan terhadap hasil evaluasi. Oleh karena itu, interpretasi hasil penelitian perlu dilakukan secara hati-hati dan tidak digeneralisasikan secara luas tanpa pengujian tambahan.

#### IV. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, algoritma *K-Nearest Neighbor* (KNN) menunjukkan kemampuan yang baik dalam mengelompokkan kecenderungan murid Madrasah Aliyah setelah lulus sekolah. Pengujian terhadap beberapa nilai parameter  $K$  menunjukkan bahwa penggunaan  $K = 5$  memberikan hasil paling optimal dengan tingkat akurasi sebesar 87,50%, disertai nilai *precision* dan *recall* yang tinggi. Hasil tersebut menandakan bahwa model mampu

memanfaatkan pola nilai rapor murid secara efektif dalam proses klasifikasi.

Penerapan algoritma KNN pada penelitian ini mengindikasikan bahwa metode klasifikasi berbasis data mining dapat digunakan sebagai sistem pendukung keputusan di lingkungan pendidikan. Meskipun algoritma ini memiliki keunggulan dalam hal kemudahan implementasi dan fleksibilitas terhadap karakteristik data, keterbatasan tetap ditemukan pada aspek efisiensi komputasi ketika diterapkan pada dataset berskala besar. Oleh karena itu, studi mendatang disarankan untuk memanfaatkan data dalam jumlah yang lebih banyak guna meningkatkan kualitas hasil penelitian serta mengkaji penggunaan algoritma klasifikasi lain guna meningkatkan performa dan efisiensi sistem secara keseluruhan. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar serta menerapkan teknik validasi silang *cross-validation* untuk meningkatkan keandalan model klasifikasi.

#### DAFTAR PUSTAKA

- [1] T. Setiyorini and R. T. Asmono, "Implementation of Gain Ratio and K-Nearest Neighbor for Classification of Student Performance," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 19–24, 2020, doi: 10.33480/pilar.v16i1.813.
- [2] "View of Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students.pdf."
- [3] "View of Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama.pdf."
- [4] 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.
- [5] D. R. Sari, V. Julianto, and H. Rhomadona, "Prediction Of Student Graduation Using The K-Nearest Neighbor Method Case Study in Politeknik Negeri Tanah Laut," *J. Ilm. Inform.*, vol. 8, no. 1, pp. 74–88, 2023, doi: 10.35316/jimi.v8i1.74-88.
- [6] "View of Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa Sma Negeri 11 Kota Bengkulu.pdf."
- [7] L. H. Jufri and D. Dasari, "Evaluation of the implementation of machine learning algorithm K-Nearest Neighbors (KNN) using rapid miner on junior high school student learning outcomes," *J. Gantang*, vol. 8, no. 2, pp. 193–197, 2023, doi: 10.31629/jg.v8i2.6590.
- [8] A. Qurotul, E. Tasia, N. Nazira, P. F. Pratama, M. R. Anugrah, and J. Adhiva, "Implementasi Algoritma Naïve Bayes Classifier ( NBC ) untuk Klasifikasi Penyakit Ginjal Kronik," vol. 4, no. September, pp. 72–76, 2022, doi: 10.30865/json.v4i1.4781.
- [9] Z. Fatah *et al.*, "Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa," vol. 2, no. 1, pp. 29–37, 2025.
- [10] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing Comparative Analysis Of Data Mining Classification Algorithm In Phishing Website Classification," vol. 11, no. 28, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [11] P. Studi *et al.*, "Data Mining Klasifikasi Penduduk Miskin Menggunakan Metode Support Vektor Machine," vol. 8, no. April, 2025.
- [12] I. M. Karo Karo, R. Romia, S. Dewi, and P. M. Fadilah, "Hoax Detection on Indonesian Tweets using Naïve Bayes Classifier with TF-IDF," *J. Inf. Syst. Res.*, vol. 4, no. 3, pp. 914–919, 2023, doi: 10.47065/josh.v4i3.3317.
- [13] "View of Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa.pdf."
- [14] R. I. Sulthoni, L. A. Muharom, and M. Rahman, "Jurnal Smart Teknologi Analisis Tingkat Kepuasan Siswa Dalam Pembelajaran Hybrid Menggunakan Algoritma K – Nearest Neighbor ( KNN ) Student Satisfaction Level Analysis In Hybrid Learning Using K – Nearest Neighbor ( KNN ) Algorithm Jurnal Smart Teknologi," vol. 4, no. 4, pp. 406–411, 2023.
- [15] "View of Analisis Perbandingan Algoritma ID3 dan KNN Pada Klasifikasi Emosi Teks Berita Berbahasa Indonesia.pdf."
- [16] "Tampilan Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Kemampuan Lulusan Siswa Dalam Bersaing untuk Mendapatkan Pekerjaan (Studi Kasus\_ SMK 'SORE' Tulungagung).pdf," 2017, *Fakultas Ilmu Komputer, Universitas Brawijaya, Malang*. doi: e-ISSN: 2548-964X.
- [17] R. Dirantara, F. Sugandi, and U. D. Wacana, "Total data yang diprediksi dengan benar Akurasi = Total pengujian yang dilaksanakan x 100 %," vol. 4307, no. 1, pp. 552–556, 2025.
- [18] N. F. Munazhif, G. J. Yanris, and M. N. S. Hasibuan, "Implementation of the K-Nearest Neighbor (kNN) Method to Determine Outstanding Student Classes," *Sinkron*, vol. 8, no. 2, pp. 719–732, 2023, doi: 10.33395/sinkron.v8i2.12227.
- [19] D. Zapata-Medina, A. Espinosa-Bedoya, and J. A. Jiménez-Builes, "Improving the Automatic Detection of Dropout Risk in Middle and High School Students: A Comparative Study of Feature Selection Techniques," *Mathematics*, vol. 12, no. 12, 2024, doi: 10.3390/math12121776.
- [20] "View of Analisis Klasifikasi Kelulusan Mahasiswa menggunakan Algoritma K-Nearest Neighbor di Fakultas Teknik Universitas Negeri Manado.pdf."
- [21] Roysid HA, Maulana A, Pujiyanto U. Can K-nearest neighbor method be used to predict success in Indonesia state university student selection. *Jurnal Ilmiah Kursor*. 2018 Dec 10;9(4).
- [22] "Tampilan Optimasi Penyaluran Bantuan Langsung Tunai Menggunakan Algoritma K-Nearest Neighbor di Desa Penedagador.pdf."
- [23] W. Apriliah *et al.*, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," vol. 10, pp. 163–171, 2021.
- [24] I. G. Ayu, N. Lestari, N. Made, R. Masita, and K. G. Meiliana, "Effectiveness of AdaBoost and XGBoost Algorithms in Sentiment Analysis of Movie Reviews," vol. 9, no. 2, pp. 258–264, 2025.
- [25] A. Perbandingan *et al.*, "Jurnal Teknologi Terpadu," vol. 8, no. 2, pp. 121–126, 2022.
- [26] M. Salsabil, N. Lutvi, and A. Eviyanti, "Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost," *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 51–58, 2024, doi: 10.32409/jikstik.23.1.3507.
- [27] A. Alhafiz and R. K. R., "Implementation of the K-Nearest Neighbor Algorithm for Birth Rate Prediction," vol. 9, no. 4, pp. 1441–1450, 2025.
- [28] H. Putra, "Comparative Study of Logistic Regression , Random Forest , and XGBoost for Bank Loan Approval Classification," vol. 9, no. 5, pp. 2822–2835, 2025.