

## Prediksi Ketuntasan Siswa Berbasis Data Multidimensi Menggunakan Metode *K-Nearest Neighbor* (KNN) di SMK NU Hasyim Asy'ari 2 Kudus

Muchammad Syafi'ul Huda<sup>1\*</sup>, Harminto Mulyo<sup>2</sup>, Gentur Wahyu Nyipto Wibowo<sup>3</sup>

<sup>1\*,2,3</sup> Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Nabdlatul Ulama Jepara, Kabupaten Jepara, Provinsi Jawa Tengah, Indonesia.

### article info

#### Article history:

Received 6 February 2026  
Received in revised form  
3 March 2026  
Accepted 25 April 2026  
Available online October  
2026.

#### Keywords:

Data Minings; K-Nearest  
Neighbors (KNN); Student  
Learning Mastery;  
Classification; Prediction.

#### Kata Kunci:

Data Mining; K-Nearest  
Neighbors (KNN); Ketuntasan  
Siswa; Klasifikasi; Prediksi.


### abstract

This research implements the K-Nearest Neighbors (KNN) algorithm to predict student learning mastery at SMK NU Hasyim Asy'ari 2 Kudus for the 2025/2026 academic year using multidimensional data. Following data preprocessing and labeling via median thresholding, the results indicate that the best performance is achieved at  $K$  values of 7, 9, and 10, with an accuracy of 58.62%. While the precision of 0.69 demonstrates reasonable accuracy in predicting students who achieve mastery, the recall of 0.50 highlights the model's limitations in identifying all students who actually pass. These results are primarily influenced by the limited sample size and imbalanced class distribution. Overall, KNN serves as an effective initial approach for objective academic prediction, though further optimization through parameter tuning or feature engineering is required to enhance future accuracy.

### abstrak

Penelitian ini mengimplementasikan algoritma K-Nearest Neighbors (KNN) untuk memprediksi ketuntasan belajar siswa di SMK NU Hasyim Asy'ari 2 Kudus TP 2025/2026 berbasis data multidimensi. Melalui tahapan preprocessing dan pelabelan menggunakan nilai median, hasil pengujian menunjukkan performa terbaik dicapai pada nilai  $k = 7, 9, \text{ dan } 10$  dengan akurasi sebesar 58,62%. Nilai precision sebesar 0,69 menunjukkan ketepatan yang cukup baik dalam memprediksi siswa tuntas, meskipun nilai recall (0,50) mengindikasikan keterbatasan model dalam mengenali seluruh siswa yang tuntas secara menyeluruh. Faktor jumlah data dan distribusi kelas yang kurang seimbang menjadi penyebab utama terbatasnya performa model. Secara keseluruhan, KNN efektif sebagai pendekatan awal prediksi akademik, namun memerlukan optimasi lebih lanjut untuk meningkatkan akurasi di masa mendatang.

\*Corresponding Author. Email: muchammadhuda6@gmail.com<sup>1\*</sup>.

Copyright 2026 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 

## 1. Pendahuluan

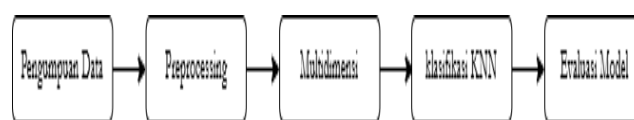
Perkembangan teknologi informasi, khususnya dalam bidang *machine learning* dan *data mining*, telah membawa transformasi signifikan dalam mendukung pengambilan keputusan berbasis data di sektor pendidikan. Analisis data saat ini tidak hanya berfungsi sebagai alat evaluasi hasil belajar, tetapi juga menjadi instrumen strategis untuk mengidentifikasi pola pembelajaran serta memprediksi ketuntasan belajar siswa secara dini. Urgensi pendekatan prediktif ini didasari oleh keterbatasan proses evaluasi manual yang cenderung lambat, subjektif, dan reaktif terhadap permasalahan akademik siswa. Pemanfaatan *machine learning* dalam pendidikan terbukti mampu menganalisis pola kompleks dalam data pembelajaran serta mendukung prediksi performa siswa secara lebih akurat dan efisien (Ahmed *et al.*, 2025). Data pendidikan bersifat kompleks dan multidimensi, mencakup aspek akademik dan non-akademik, serta sering menghadapi masalah kualitas seperti nilai yang hilang dan ketidakseimbangan kelas. Oleh karena itu, diperlukan preprocessing yang baik agar model prediksi dapat bekerja optimal dan menghasilkan keputusan yang akurat. Pemanfaatan *data mining* dan *machine learning* dalam pendidikan terbukti efektif untuk menganalisis pembelajaran dan memprediksi performa siswa. Salah satu metode yang banyak digunakan adalah *K-Nearest Neighbors* (KNN), yang mengklasifikasikan data berdasarkan kedekatan antar objek dalam ruang fitur.

Meskipun fleksibel dalam menangani data multidimensi, kinerja KNN sangat bergantung pada pemilihan parameter  $K$  dan kualitas skala data (Anggraini, 2025). Meskipun penelitian prediksi akademik telah banyak dilakukan, penerapannya pada Sekolah Menengah Kejuruan (SMK) masih relatif terbatas dibandingkan jenjang pendidikan umum. Kombinasi pembelajaran teori dan praktik di SMK menciptakan variabilitas data yang lebih kompleks. SMK NU Hasyim Asy'ari 2 Kudus memiliki aset data yang lengkap, namun pemanfaatannya masih terbatas pada pelaporan deskriptif administratif dan belum dioptimalkan untuk analisis prediktif. Akibatnya, intervensi terhadap siswa berisiko sering kali terlambat dilakukan. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun

model prediksi ketuntasan siswa menggunakan metode KNN berbasis data multidimensi. Model ini diharapkan mampu memberikan informasi prediktif yang akurat bagi pihak SMK NU Hasyim Asy'ari 2 Kudus dalam mengidentifikasi risiko ketidaktuntasan secara lebih cepat, sekaligus berkontribusi pada pengembangan literatur *machine learning* di jenjang pendidikan kejuruan.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data akademik siswa sebagai objek analisis karena mampu mengolah data numerik secara sistematis untuk menghasilkan model prediksi yang objektif dan terukur. Dalam konteks pendidikan, pendekatan berbasis data terbukti efektif dalam menganalisis performa siswa serta mendukung pengambilan keputusan yang lebih akurat. Tujuan penelitian ini adalah membangun model prediksi ketuntasan siswa menggunakan algoritma *K-Nearest Neighbors* (KNN) berbasis data multidimensi. Proses penelitian dilakukan secara sistematis melalui tahapan pengumpulan data, preprocessing, pembentukan data multidimensi, pembangunan model, dan evaluasi performa, yang merupakan tahapan umum dalam proses *data mining* untuk menghasilkan model prediktif yang andal (Aliefa Khana Kotjek *et al.*, 2025a).



Gambar 1. Skema Penelitian

### Dataset

Dataset yang digunakan dalam penelitian ini merupakan data akademik siswa yang diperoleh dari SMK NU Hasyim Asy'ari 2 Kudus dengan total sebanyak 137 data siswa. Data tersebut terdiri dari berbagai atribut yang mencerminkan aspek akademik dan non-akademik siswa, seperti jenis kelamin, kelas, jurusan, nilai teori, nilai keagamaan, nilai tugas, nilai kejuruan, kehadiran, jumlah alpha, keaktifan, kedisiplinan, serta kegiatan ekstrakurikuler. Atribut target berupa ketuntasan siswa tidak tersedia secara langsung dalam dataset, sehingga dilakukan pembentukan label berdasarkan nilai rata-rata siswa. Penentuan ketuntasan dilakukan menggunakan

pendekatan median sebagai threshold untuk menghasilkan distribusi kelas yang lebih seimbang antara siswa tuntas dan tidak tuntas.

### Preprocessing Data

Proses ini diawali dengan pembersihan data melalui identifikasi nilai yang hilang pada atribut akademik maupun non-akademik. Penanganan data yang tidak lengkap dilakukan secara selektif; rekam data yang tidak dapat diperbaiki melalui metode imputasi akan dihapus guna menghindari bias pada proses pembelajaran model. Selain itu, dilakukan pemeriksaan terhadap duplikasi dan ketidakkonsistenan format untuk menjamin keseragaman dataset. Validitas data pada tahap ini sangat penting karena data yang hilang bisa menurunkan akurasi model dan memengaruhi hasil prediksi. Oleh karena itu, data yang hilang perlu ditangani, misalnya dengan mengisi nilai yang kosong atau menghapus data yang tidak layak digunakan (Widiyanti & Pratama, 2024). Selanjutnya, dilakukan seleksi atribut untuk menentukan variabel yang relevan terhadap prediksi ketuntasan belajar. Atribut identitas seperti Nomor Induk Siswa (NIS), nama, dan alamat dieliminasi karena tidak memiliki kontribusi prediktif. Sebaliknya, variabel yang berkaitan dengan performa belajar seperti nilai teori, nilai kejuruan, tugas, kehadiran, keaktifan, kedisiplinan, serta partisipasi ekstrakurikuler dipertahankan. Proses seleksi atribut ini bertujuan untuk mengurangi dimensi data, meningkatkan efisiensi komputasi, serta meningkatkan akurasi model klasifikasi (Gunawan *et al.*, 2024). Mengingat algoritma KNN berbasis pada perhitungan jarak numerik dalam ruang multidimensi, atribut kategorikal seperti jenis kelamin, kelas, jurusan, dan ekstrakurikuler ditransformasikan melalui proses encoding. Transformasi ini memberikan representasi numerik pada setiap kategori tanpa mengubah makna semantiknya, sehingga atribut tersebut dapat diintegrasikan ke dalam perhitungan fitur. Bersamaan dengan itu, dilakukan pembentukan label target ketuntasan yang ditentukan berdasarkan nilai rata-rata dari komponen nilai teori, keagamaan, tugas, dan kejuruan. Penelitian ini menggunakan nilai median sebagai ambang batas (threshold) untuk membedakan kategori tuntas dan tidak tuntas. Pendekatan berbasis distribusi ini dipilih untuk menghasilkan proporsi kelas yang seimbang,

sehingga model dapat dilatih secara optimal tanpa mengalami bias terhadap salah satu kelas (tuntas atau tidak tuntas) (Silfana & Barata, 2024). Tahap akhir dari pra-pengolahan adalah normalisasi data menggunakan metode Min-Max Normalization untuk menyamakan skala nilai setiap atribut numerik ke dalam rentang 0 hingga 1. Langkah ini sangat krusial bagi algoritma KNN yang bersifat distance-based, karena perbedaan skala antar atribut dapat menyebabkan dominasi fitur tertentu yang memiliki rentang nilai lebih besar. Dengan normalisasi, setiap fitur memiliki bobot kontribusi yang seimbang dalam perhitungan jarak Euclidean, sehingga hasil klasifikasi menjadi lebih akurat dan stabil. Melalui integrasi seluruh rangkaian proses ini, data yang semula tidak terstruktur ditransformasikan menjadi dataset berkualitas tinggi yang siap digunakan untuk menghasilkan performa prediksi optimal (Palinggik Allorerung *et al.*, 2024).

### Pembentukan Data Multidimensi

Proses ini diawali dengan pembersihan data melalui identifikasi nilai yang hilang pada atribut akademik maupun non-akademik. Penanganan data yang tidak lengkap dilakukan secara selektif; rekam data yang tidak dapat diperbaiki melalui metode imputasi akan dihapus guna menghindari bias pada proses pembelajaran model. Selain itu, dilakukan pemeriksaan terhadap duplikasi dan ketidakkonsistenan format untuk menjamin keseragaman dataset. Validitas data yang tinggi pada tahap ini sangat krusial agar tidak terjadi distorsi saat algoritma K-Nearest Neighbors melakukan perhitungan jarak matematis antar objek. Tahapan pembersihan data tersebut merupakan bagian penting dalam proses data mining karena berpengaruh langsung terhadap kualitas model yang dihasilkan (Hadi *et al.*, 2024). Secara matematis, data multidimensi dapat dinyatakan sebagai  $x = (x_1, x_2, x_3, \dots, x_n)$ , di mana setiap komponen  $x_i$  merupakan nilai dari atribut yang telah melalui proses preprocessing, dan  $n$  menyatakan jumlah atribut yang digunakan dalam penelitian. Dengan representasi ini, hubungan antar data dapat dianalisis berdasarkan kedekatan posisi antar titik dalam ruang fitur. Pendekatan multidimensi menjadi sangat penting karena ketuntasan belajar siswa tidak hanya dipengaruhi oleh satu faktor, melainkan merupakan hasil interaksi dari berbagai faktor akademik dan non-akademik, sesuai dengan konsep representasi data

dalam ruang fitur pada data mining dan klasifikasi berbasis KNN. Selain itu, pembentukan data multidimensi juga memungkinkan penerapan metode berbasis jarak seperti K-Nearest Neighbors (KNN), yang memerlukan representasi numerik dalam ruang fitur untuk menghitung jarak antar data. Dengan struktur data yang telah terstandarisasi dan berada dalam skala yang seragam, proses perhitungan jarak menjadi lebih akurat dan tidak bias terhadap atribut tertentu. Oleh karena itu, tahap ini menjadi penghubung penting antara proses preprocessing dan tahap pemodelan, karena memastikan bahwa data telah siap digunakan dalam proses klasifikasi secara optimal (Pratiwi *et al.*, 2025).

### K-Nearest Neighbors (KNN)

Metode klasifikasi yang diimplementasikan dalam penelitian ini adalah K-Nearest Neighbors (KNN), sebuah algoritma berbasis instansi yang termasuk dalam kategori lazy learning. Prinsip kerja KNN didasarkan pada klasifikasi data baru melalui identifikasi tingkat kedekatan terhadap data latih yang telah memiliki label kelas. Kedekatan antar data diukur melalui perhitungan jarak dalam ruang fitur, di mana setiap data direpresentasikan sebagai vektor fitur hasil preprocessing. Pendekatan ini umum digunakan dalam klasifikasi data pendidikan karena berbasis pada konsep jarak dan representasi multidimensi (Aliefa Khana Kotjek *et al.*, 2025). Dalam mekanisme klasifikasinya, algoritma KNN menghitung jarak antara data uji dengan seluruh data latih menggunakan metrik jarak Euclidean. Penggunaan Euclidean distance dipilih karena efektivitasnya dalam menangani data numerik yang telah melalui proses normalisasi. Setelah seluruh jarak dihitung, algoritma akan melakukan pemeringkatan untuk menentukan sejumlah  $k$  data terdekat sebagai tetangga. Nilai  $k$  merupakan parameter krusial yang menentukan jumlah tetangga yang dilibatkan dalam proses pengambilan keputusan. Penentuan kelas akhir dari data uji dilakukan melalui mekanisme majority voting dari  $k$  tetangga terdekat tersebut, di mana data diklasifikasikan ke dalam kategori yang memiliki frekuensi kemunculan terbanyak. Keunggulan utama metode KNN terletak pada kemampuannya dalam menangani dataset dengan atribut yang beragam tanpa memerlukan proses pelatihan model yang kompleks. Hal ini menjadikannya algoritma yang relatif sederhana

namun tetap reliabel dalam menangani pola data lokal yang mungkin tidak terdeteksi oleh metode klasifikasi bersifat global. Meski demikian, performa KNN sangat bergantung pada sensitivitas pemilihan nilai  $k$  serta kualitas dataset, termasuk efektivitas tahap preprocessing seperti normalisasi dan seleksi atribut (Indah Wahyuni & Kurniawan R., 2025). Pada penelitian ini, optimalisasi nilai  $k$  dilakukan melalui serangkaian pengujian eksperimental terhadap beberapa nilai  $k$  guna memperoleh performa terbaik berdasarkan metrik evaluasi yang ditetapkan. Berdasarkan hasil pengujian yang dilakukan, nilai  $k = 1$  menghasilkan tingkat akurasi tertinggi dalam memprediksi ketuntasan siswa. Meskipun nilai  $k = 1$  yang rendah berisiko terhadap noise (overfitting), hasil klasifikasi dalam penelitian ini menunjukkan kemampuan generalisasi yang baik terhadap data baru. Dengan demikian, penerapan KNN berbasis data multidimensi ini diharapkan mampu menyediakan model klasifikasi yang objektif dan efektif dalam memprediksi ketuntasan belajar siswa di SMK NU Hasyim Asy'ari 2 Kudus.

### Evaluasi Model

Evaluasi model merupakan tahap krusial dalam penelitian ini untuk mengukur kinerja algoritma K-Nearest Neighbors dalam memprediksi ketuntasan belajar siswa menggunakan data uji (testing set) guna memberikan gambaran objektif terhadap kemampuan generalisasi model dalam menghadapi data baru (Fatah & Safira Muasolli, n.d.). Proses evaluasi performa ini dilakukan secara komprehensif menggunakan metrik klasifikasi yang meliputi Accuracy, Precision, Recall, dan F1-Score untuk memastikan analisis kinerja yang mendalam, baik dari segi ketepatan prediksi secara keseluruhan maupun kemampuan model dalam mengklasifikasikan masing-masing kelas secara akurat, terutama dalam mengantisipasi kondisi distribusi data yang tidak seimbang. Melalui integrasi metrik-metrik tersebut, efektivitas model dalam mendeteksi siswa yang tuntas maupun tidak tuntas dapat dinilai secara lebih valid, terukur, dan komprehensif.

#### 1) Accuracy

Metrik umum yang mengukur proporsi prediksi benar dari seluruh data (Helmiyah *et al.*, 2025)

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

2) Precision

Presisi mengukur seberapa akurat prediksi positif yang dibuat model (Helmiyah *et al.*, 2025)

$$precision = \frac{TP}{TP + FP}$$

3) Recall

Recall mengukur kemampuan model mendeteksi seluruh kasus positif yang sebenarnya (Helmiyah *et al.*, 2025).

$$Recall = \frac{TP}{TP + FN}$$

4) F1-score

Rata-rata harmonik dari Precision dan Recall, yang memberikan satu nilai untuk menilai keseimbangan keduanya (Helmiyah *et al.*, 2025)

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

### 3. Hasil dan Pembahasan

#### Hasil

##### Dataset

Dataset yang digunakan dalam penelitian ini merupakan data rekam akademik siswa SMK NU Hasyim Asy'ari 2 Kudus pada Tahun Pelajaran

2025/2026. Data ini mencakup berbagai dimensi informasi yang merepresentasikan aspek akademik maupun non-akademik siswa. Atribut-atribut yang dilibatkan telah melalui tahap seleksi fitur guna memastikan bahwa hanya variabel dengan relevansi prediktif tinggi yang digunakan dalam model. Atribut tersebut meliputi jenis kelamin, kelas, jurusan, nilai teori, nilai keagamaan, nilai tugas, nilai kejuruan, tingkat kehadiran, jumlah alpha, tingkat keaktifan, kedisiplinan, serta partisipasi dalam kegiatan ekstrakurikuler. Total data yang digunakan dalam penelitian ini adalah sebanyak 137 data yang diklasifikasikan ke dalam dua kelas target, yaitu "Tuntas" (1) dan "Tidak Tuntas" (0). Mengingat atribut ketuntasan tidak tersedia secara eksplisit dalam dataset mentah, dilakukan proses pelabelan (labeling) secara mandiri. Penentuan label tersebut didasarkan pada perhitungan nilai rata-rata kumulatif siswa yang kemudian divalidasi menggunakan nilai median sebagai ambang batas (threshold) klasifikasi. Pemilihan median sebagai batas klasifikasi merupakan strategi untuk menangani potensi ketidakseimbangan data. Hasil distribusi menunjukkan bahwa pendekatan ini mampu menghasilkan proporsi kelas yang relatif seimbang, sehingga algoritma K-Nearest Neighbor dapat melakukan proses pembelajaran secara optimal tanpa mengalami bias terhadap salah satu kelas mayoritas. Hal ini sangat penting untuk menjamin bahwa model memiliki kemampuan deteksi yang sama baiknya untuk kategori siswa tuntas maupun tidak tuntas.

Tabel 1. Atribut Dataset Siswa

No	Atribut	Deskripsi
1	NIS	Nomor Induk Siswa sebagai identitas unik.
2	Nama	Nama lengkap siswa
3	Jenis Kelamin	Laki-laki / Perempuan
4	Nilai Teori	Nilai akademik teori siswa.
5	Nilai Keagamaan	Nilai mata pelajaran keagamaan
6	Nilai Tugas	Nilai tugas siswa
7	Nilai UAS	Nilai ujian akhir semester
8	Kehadiran	Persentase kehadiran siswa
9	Keaktifan	Tingkat partisipasi siswa
10	Kedisiplinan	Tingkat kedisiplinan siswa
11	Status	Catatan Riwayat keluhan atau kondisi klinis pasien

## Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data yang digunakan dalam penelitian telah bersih, konsisten, dan siap digunakan dalam proses pemodelan menggunakan algoritma K-Nearest Neighbors. Proses ini meliputi pembersihan data, seleksi atribut, transformasi data kategorikal, pembentukan label ketuntasan, serta normalisasi data. Pada tahap pembersihan data (data cleaning), dilakukan pemeriksaan terhadap nilai yang hilang (missing value) dan inkonsistensi data. Data yang tidak lengkap dan tidak dapat diperbaiki dihapus untuk menghindari bias dalam model. Selanjutnya dilakukan seleksi atribut dengan menghapus atribut yang tidak relevan seperti NIS, nama, dan alamat, serta mempertahankan atribut akademik dan non-akademik yang berpengaruh terhadap ketuntasan siswa. Atribut kategorikal seperti jenis kelamin dan ekstrakurikuler kemudian ditransformasikan ke dalam bentuk numerik melalui proses encoding agar dapat diproses oleh algoritma KNN. Selain itu, dilakukan pembentukan label ketuntasan berdasarkan nilai rata-rata siswa dengan menggunakan nilai median sebagai batas klasifikasi, di mana siswa dengan nilai di atas atau sama dengan median dikategorikan sebagai tuntas (1) dan di bawah median sebagai tidak tuntas (0). Untuk melihat hasil dari proses labeling tersebut, distribusi kelas pada dataset ditampilkan pada Tabel berikut:

Tabel 2. Distribusi Kelas Ketuntasan Siswa

Kelas	Keterangan	Jumlah
0	Tidak tuntas	66
1	Tuntas	75

Tahap terakhir adalah normalisasi data menggunakan metode Min-Max untuk menyamakan skala seluruh atribut ke dalam rentang 0 sampai 1. Proses ini penting karena algoritma KNN menggunakan perhitungan jarak, sehingga perbedaan skala dapat memengaruhi hasil klasifikasi.

## Pembentukan Data Multidimensi

Setelah melalui preprocessing, data akademik siswa SMK NU Hasyim Asy'ari 2 Kudus Tahun Pelajaran 2025/2026 ditransformasikan ke dalam representasi multidimensi. Setiap siswa dinyatakan sebagai vektor fitur  $x = (x_1, x_2, x_3, \dots, x_n)$  yang menempati titik spesifik dalam ruang fitur (feature space).

Penggunaan 12 atribut numerik yang telah dinormalisasi ini memastikan bahwa setiap dimensi—mulai dari nilai akademik hingga kedisiplinan—memiliki bobot kontribusi yang seimbang dalam perhitungan jarak Euclidean tanpa adanya dominasi skala dari variabel tertentu. Selanjutnya, dataset sebanyak 137 data dibagi secara sistematis menjadi dua bagian dengan rasio 80:20, menghasilkan 110 data latih dan 27 data uji. Data latih digunakan sebagai basis pengetahuan bagi algoritma KNN untuk mengenali pola ketuntasan, sedangkan data uji berfungsi sebagai instrumen untuk mengukur kemampuan generalisasi model terhadap data baru. Pembagian ini menjadi dasar objektif bagi proses klasifikasi dan evaluasi performa model yang akan dibahas pada bagian selanjutnya.

## K-Nearest Neighbors (KNN)

Pada tahap ini dilakukan proses klasifikasi menggunakan algoritma K-Nearest Neighbors untuk memprediksi ketuntasan siswa berdasarkan data multidimensi yang telah dibentuk. Algoritma KNN bekerja dengan menghitung jarak antara data uji dan seluruh data latih, kemudian menentukan kelas berdasarkan mayoritas dari sejumlah K tetangga terdekat. Dalam penelitian ini, digunakan jarak Euclidean sebagai metode perhitungan jarak karena sesuai untuk data numerik yang telah dinormalisasi. Untuk memperoleh nilai K yang optimal, dilakukan pengujian terhadap beberapa nilai K dalam rentang 1 hingga 10. Hasil pengujian ditunjukkan pada tabel berikut.

Tabel 3. Hasil Uji

Nilai k	Akurasi
1	0.55
2	0.58
3	0.58
4	0.48
5	0.51
6	0.48
7	0.58
8	0.55
9	0.58
10	0.58

## Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja algoritma K-Nearest Neighbors dalam memprediksi ketuntasan siswa menggunakan data uji. Evaluasi

dilakukan menggunakan confusion matrix dan metrik klasifikasi untuk memberikan gambaran yang lebih detail terhadap performa model.

Tabel 4. Confusion Matrix

	Prediksi tidak tuntas	Prediksi tuntas
Aktual tidak tuntas	7	4
Aktual Tuntas	9	9

Berdasarkan confusion matrix tersebut, model mampu mengklasifikasikan sebagian besar data dengan benar, meskipun masih terdapat kesalahan

prediksi baik pada kelas tidak tuntas maupun tuntas.

Tabel 5. Classification Report

Kelas	Precision	Recall	F1-Score
Tidak Tuntas (0)	0.44	0.64	0.52
Tuntas (1)	0.69	0.50	0.58

## Pembahasan

Penelitian ini menunjukkan bahwa penggunaan algoritma *K-Nearest Neighbors* (KNN) sebagai metode untuk memprediksi ketuntasan belajar siswa di SMK NU Hasyim Asy'ari 2 Kudus dapat memberikan hasil yang cukup menjanjikan, dengan akurasi tertinggi mencapai 58,62% pada nilai  $k$  tertentu. Hasil ini sejalan dengan temuan yang diungkapkan oleh Aliefa Khana Kotjek *et al.* (2025), yang juga menemukan bahwa KNN efektif dalam mengklasifikasikan data akademik siswa. Penelitian tersebut menekankan pentingnya pemilihan parameter  $k$  yang tepat, di mana nilai  $k = 1$  dalam penelitian ini menunjukkan potensi untuk mencapai akurasi yang lebih tinggi, meskipun juga membawa risiko *overfitting*. Selain itu, proses preprocessing yang dilakukan, termasuk pembersihan data dan seleksi atribut, sangat berpengaruh terhadap kualitas model yang dihasilkan. Hal ini sejalan dengan penelitian oleh Gunawan *et al.* (2024), yang menyoroti bahwa pemilihan atribut yang relevan dapat meningkatkan efisiensi komputasi dan akurasi model klasifikasi. Dengan demikian, hasil penelitian ini menegaskan bahwa penerapan KNN dalam konteks pendidikan, khususnya di SMK, dapat menjadi langkah awal yang baik untuk pengembangan sistem prediksi akademik yang lebih canggih, meskipun tetap memerlukan optimasi lebih lanjut untuk meningkatkan akurasi dan efektivitasnya.

## 4. Kesimpulan

Berdasarkan hasil analisis, algoritma KNN dapat diimplementasikan untuk prediksi ketuntasan belajar siswa di SMK NU Hasyim Asy'ari 2 Kudus TP 2025/2026 dengan akurasi tertinggi sebesar 58,62% pada nilai  $K = 7, 9, \text{ dan } 10$ . Meskipun nilai precision (0,69) menunjukkan ketepatan yang cukup baik dalam memprediksi siswa tuntas, nilai recall yang rendah (0,50) mengindikasikan bahwa model masih kesulitan mengenali keseluruhan siswa yang benar-benar tuntas akibat keterbatasan jumlah data dan distribusi kelas yang tidak seimbang. Secara keseluruhan, KNN efektif sebagai pendekatan awal, namun memerlukan pengembangan lebih lanjut melalui optimasi parameter atau feature engineering untuk meningkatkan stabilitas dan akurasi prediksi di masa mendatang.

## 5. Ucapan Terima Kasih

Penulis menyampaikan apresiasi kepada Universitas Islam Nahdlatul Ulama Jepara, dosen pembimbing, serta rekan-rekan yang telah memberikan dukungan, arahan, dan motivasi sehingga penelitian ini dapat diselesaikan dengan baik.

## 6. Daftar Pustaka

- Ahmed, W., Wani, M. A., Plawiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-12353-4>.
- Aliefa Khana Kotjek, R., Shabrina Wijaya, F., Wahyudi, M., & Surya Budiman, A. (2025a). Klasifikasi nilai ujian siswa berdasarkan kebiasaan belajar menggunakan *K-Nearest Neighbor* dan *Support Vector Machine*. *Jurnal Sains Informatika Terapan (JSIT)*, E-ISSN (Number 2).
- Aliefa Khana Kotjek, R., Shabrina Wijaya, F., Wahyudi, M., & Surya Budiman, A. (2025b). Klasifikasi nilai ujian siswa berdasarkan kebiasaan belajar menggunakan *K-Nearest Neighbor* dan *Support Vector Machine*. *Jurnal Sains Informatika Terapan (JSIT)*, E-ISSN (Number 2).
- Anggraini, D. (2025). Data mining pendidikan: Prediksi gaya belajar mahasiswa teknik menggunakan *machine learning*. 11(3), 563–572.
- Fatah, Z., & Safira Muasolli, H. (n.d.). Penerapan algoritma *K-Nearest Neighbor* (K-NN) untuk memprediksi kelulusan mahasiswa menggunakan *RapidMiner*. *JAMASTIKA*, 4.
- Florent Amor, A., Rizal Ihsanudin, M., Oka Yulistianto, S., & Direvisi, D. (2025). Prediksi kinerja siswa SMP berdasarkan data akademik dan perilaku menggunakan *machine learning*. 18(2), 67–76.
- Gunawan, M. T., Tine, J. Y., & Murwaningtyas, C. E. (2024). Model decision tree untuk prediksi prestasi akademik matematika siswa kelas VIII SMP Frater Don Bosco Manado. *Jurnal Pendidikan Informatika dan Sains*, 13(2), 141–153. <https://doi.org/10.31571/saintek.v13i2.7696>.
- Hadi, R., Luh, N., Suwirmayanti, G. P., Ngurah, G., Kusuma, A., Ayu, G., Saryanti, D., Novayanti, D., Stikom Bali, I., & Raya Puputan, J. (2024). Implementasi metode normalisasi dan seleksi fitur dalam optimasi algoritma *K-Nearest Neighbor* (KNN) untuk klasifikasi data bank. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 7(5).
- Helmiyah, S., Pramestiawan, R., Studi Pendidikan Informatika, P., & Tinggi Keguruan dan Ilmu Pendidikan Rosalia Lampung, S. (2025). Analisis komparatif algoritma *machine learning* dengan metrik akurasi, presisi, recall, dan F1-score pada dataset kacang kering. 6(3), 152–159. <https://doi.org/10.35960/ikomti.v6i3.2031>.
- Indah Wahyuni, U., & Kurniawan, R. (2025). Algoritma *K-Nearest Neighbor Classification* sebagai sistem pengelompokan kemampuan akademik siswa berbasis web. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(2), 589–600. <https://doi.org/10.51454/decode.v5i2.1246>.
- Palinggik Allorerung, P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis performa normalisasi data untuk klasifikasi *K-Nearest Neighbor* pada dataset penyakit. *Jurnal Informatika Sunan Kalijaga*, 9(3).
- Pratiwi, A., Manurung, H., Pita Uli Sitompul, M., Informasi, S., & Kaputama, S. (2025). Implementasi metode *K-Means Clustering* untuk mengklasifikasi status gizi balita berdasarkan wilayah kerja Puskesmas Karang Rejo.
- Silfana, F. I., & Barata, M. A. (2024). Using *K-NN* algorithm for evaluating feature selection on high dimensional datasets. *JURNAL TEKNIK INFORMATIKA*, 17(2), 190–202. <https://doi.org/10.15408/jti.v17i2.40866>.
- Widianti, A., & Pratama, I. (2024). Penanganan missing values dan prediksi data timbunan sampah berbasis *machine learning*. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 9(2), 242–251. <https://doi.org/10.36341/rabit.v9i2.4789>.