

## Prediksi Tingkat Pemahaman Materi Peserta Didik Dalam Pembelajaran Matematika Menggunakan Algoritma *K-Nearest Neighbors* (KNN)

Muammar Ashari Abuspin<sup>1)\*</sup>, Andi Sri Yulianti<sup>2)</sup>

<sup>1)</sup>Prodi Statistika, Universitas Patempo, Makassar, [muammarashari96@gmail.com](mailto:muammarashari96@gmail.com)

<sup>2)</sup>Prodi Statistika Universitas Patempo, Makassar, [andisriyulianti@gmail.com](mailto:andisriyulianti@gmail.com)

### Abstract

*This research is motivated by the importance of understanding mathematics material as a primary factor influencing the success of students' learning processes. Differences in understanding levels between individuals often pose a challenge for educators in accurately and quickly identifying learning needs. Therefore, a data-driven approach capable of objectively predicting students' understanding levels is needed. This study aims to develop a predictive model for understanding levels in mathematics learning using the K-Nearest Neighbors algorithm. Research data were collected from learning outcomes, learning activities, and other indicators that represent the process of conceptual understanding. The research stages include data preprocessing, feature selection, model training, and performance evaluation using specific testing methods. The resulting model is used to classify students' understanding levels based on the data patterns formed. The analysis results show that the K-Nearest Neighbors algorithm has a good ability to identify differences in understanding levels with high accuracy. These findings indicate that a machine learning-based approach can be an effective tool for educators in designing more adaptive and targeted learning strategies. Furthermore, the developed prediction model can support data-driven decision-making to continuously improve the quality of the mathematics learning process. Thus, the application of this model is expected to provide a real contribution in the development of a learning evaluation system that is more objective, measurable, and responsive to the learning needs of students in various formal and non-formal educational contexts at various levels of different educational units and encourage technology-based learning innovation.*

*Keywords: K-Nearest Neighbors ; understanding of material; learning mathematics three*

### Abstrak

Penelitian ini dilatarbelakangi oleh pentingnya pemahaman materi matematika sebagai faktor utama yang mempengaruhi keberhasilan proses pembelajaran peserta didik. Perbedaan tingkat pemahaman antarindividu seringkali menjadi tantangan bagi pendidik dalam mengidentifikasi kebutuhan belajar secara tepat dan cepat. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu memprediksi tingkat pemahaman peserta didik secara objektif. Penelitian ini bertujuan untuk mengembangkan model prediksi tingkat pemahaman materi dalam pembelajaran matematika menggunakan algoritma K-Nearest Neighbors. Data penelitian dikumpulkan dari hasil belajar, aktivitas pembelajaran, serta indikator lain yang merepresentasikan proses pemahaman konsep. Tahapan penelitian meliputi praproses data, pemilihan fitur, pelatihan model, dan evaluasi kinerja menggunakan metode pengujian tertentu. Model yang dihasilkan digunakan untuk mengklasifikasikan tingkat pemahaman peserta didik berdasarkan pola data yang terbentuk. Hasil analisis menunjukkan bahwa algoritma K-Nearest Neighbors memiliki kemampuan yang baik dalam mengidentifikasi perbedaan tingkat pemahaman dengan akurasi yang tinggi. Temuan ini menunjukkan bahwa pendekatan berbasis machine learning dapat menjadi alat bantu yang efektif bagi pendidik dalam merancang strategi pembelajaran yang lebih adaptif dan tepat sasaran. Selain itu, model prediksi yang dikembangkan dapat mendukung pengambilan keputusan berbasis data untuk meningkatkan kualitas proses pembelajaran matematika secara berkelanjutan. Dengan demikian, penerapan model ini diharapkan mampu memberikan kontribusi nyata dalam pengembangan sistem evaluasi pembelajaran yang lebih objektif, terukur, dan responsif terhadap kebutuhan belajar peserta didik di berbagai konteks pendidikan formal maupun nonformal pada berbagai tingkat satuan pendidikan yang berbeda serta mendorong inovasi pembelajaran berbasis teknologi.

Kata kunci: K-Nearest Neighbors, pemahaman materi, pembelajaran matematika.

### PENDAHULUAN

Pembelajaran matematika merupakan salah satu komponen penting dalam sistem pendidikan karena berperan dalam membentuk kemampuan berpikir logis, analitis, sistematis, dan kritis peserta didik (Sulistiani & Masrukan, 2016). Namun demikian, dalam praktiknya, pemahaman materi

matematika di kalangan peserta didik masih menunjukkan variasi yang cukup signifikan. Perbedaan kemampuan awal, gaya belajar, motivasi, serta faktor lingkungan belajar seringkali menyebabkan tingkat pemahaman materi yang tidak merata (Aprilia, n.d.). Kondisi ini menuntut adanya pendekatan yang lebih adaptif dan berbasis data untuk mengidentifikasi serta memprediksi tingkat pemahaman peserta didik secara lebih akurat.

Selama ini, evaluasi pemahaman materi umumnya dilakukan melalui tes formatif maupun sumatif (Taqiyuddin et al., 2024). Meskipun metode tersebut mampu memberikan gambaran capaian belajar, pendekatan ini bersifat reaktif dan belum sepenuhnya mampu memberikan prediksi dini terhadap kemungkinan kesulitan belajar yang akan dialami peserta didik. Oleh karena itu, diperlukan suatu metode prediktif yang dapat membantu pendidik dalam mengambil keputusan pembelajaran yang lebih tepat dan cepat.

Perkembangan teknologi informasi dan analisis data telah mendorong pemanfaatan teknik *machine learning* dalam bidang pendidikan, yang dikenal sebagai *Educational Data Mining* (Sulistiani & Masrukan, 2016). Salah satu algoritma yang banyak digunakan untuk klasifikasi dan prediksi adalah K-Nearest Neighbors (KNN). KNN memiliki keunggulan dalam menangani data berdimensi tinggi, mampu bekerja pada data berukuran kecil hingga menengah, serta memiliki tingkat akurasi yang baik dalam proses klasifikasi (Dwitama, 2025).

Dalam konteks pembelajaran matematika, penerapan KNN berpotensi digunakan untuk mengklasifikasikan dan memprediksi tingkat pemahaman peserta didik berdasarkan berbagai variabel, seperti nilai tugas, keaktifan belajar, hasil kuis, kehadiran, serta faktor pendukung lainnya (Setyawan et al., 2025). Dengan model prediksi yang akurat, pendidik dapat melakukan intervensi lebih dini terhadap peserta didik yang berpotensi mengalami kesulitan belajar, sehingga proses pembelajaran menjadi lebih efektif dan terarah (Andika et al., 2024).

Meskipun berbagai penelitian telah mengkaji penerapan *machine learning* dalam pendidikan, kajian yang secara spesifik memprediksi tingkat pemahaman materi matematika menggunakan algoritma KNN masih memerlukan penguatan, khususnya pada konteks pembelajaran di tingkat pendidikan menengah atau tinggi. Oleh karena itu, penelitian ini menjadi relevan untuk dilakukan guna mengembangkan model prediksi tingkat pemahaman materi peserta didik dalam pembelajaran matematika menggunakan algoritma KNN yang diharapkan dapat memberikan kontribusi dalam peningkatan kualitas pembelajaran berbasis data. Dengan demikian, penelitian ini diharapkan mampu memberikan solusi inovatif dalam mendukung pengambilan keputusan pembelajaran yang lebih efektif, objektif, dan berbasis teknologi.

## METODE PENELITIAN

### Jenis Sumber data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu berupa data daftar hadir dan nilai peserta didik semester ganjil 2025/2026 yang diperoleh dari SMAN 1 Gowa. Data tersebut merupakan data tentang absensi kehadiran siswa dan faktor-faktor yang mempengaruhi kelulusan siswa(i) dalam pembelajaran matematika.

Daftar hadir dan nilai peserta didik dibagi menjadi data *training* dan *testing*.

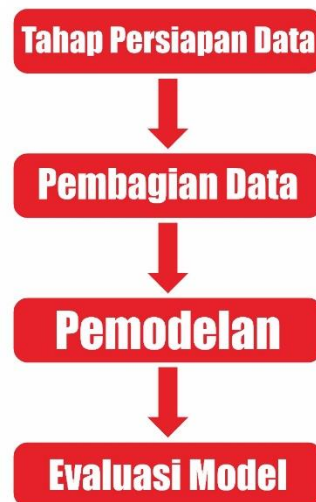
### Variabel Penelitian

Variabel yang akan digunakan pada penelitian ini adalah :

- y = Keterangan (Lulus dan Tidak Lulus)
- x1 = Jenis Kelamin x2 = Nilai Semester
- x3 = Nilai Ujian Harian
- x4 = Persentase Kehadiran x5 = Kelas
- x6 = Semester

## Tahapan Penelitian

Penelitian ini berfokus pada pengembangan model *data mining* dengan klasifikasi *K-Nearest Neighbor*. Adapun metode analisis disusun sebagai berikut:



Gambar 1. Tahapan Penelitian

Tahap Persiapan Data (Anggraini & Alita, 2024) :

1. Identifikasi data yang terkumpul dalam basis data yang tersedia di SMAN 1 Gowa
2. Melakukan seleksi atau pemilihan data yang relevan dengan analisis yang dilakukan
3. *Preparation* data dan transformasi data yang terdiri dari :
  - a. Pembersihan data diterapkan untuk menghilangkan *noise* dan memperbaiki data yang tidak konsisten
  - b. Reduksi data digunakan untuk mengurangi ukuran data. Misalnya menggabungkan, menghilangkan fitur yang berlebihan atau mengelompokkan transformasi.
4. Transformasi data yang telah dipilih menjadi *coding* agar data tersebut sesuai dengan proses *data mining*

Tahap Pembagian Data

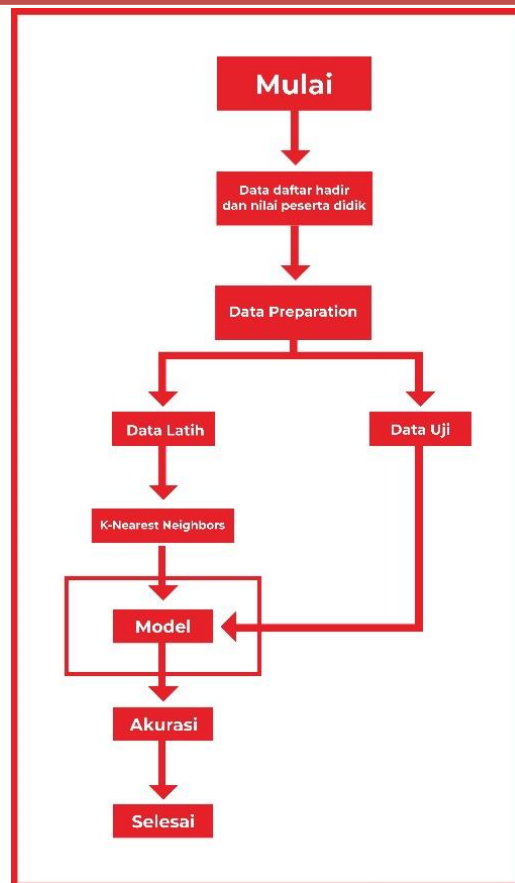
Membagi data ke dalam data latih dan data uji. Pada penelitian ini pembagian data menggunakan sampel acak sederhana

Tahap Pemodelan

Pada bagian ini dijelaskan tentang langkah-langkah eksperimen meliputi cara pemilihan arsitektur yang tepat dari model atau metode yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat. Menggunakan algoritma K-Nearest Neighbors.

Evaluasi Model

Pada bagian ini dijelaskan tentang evaluasi dan validasi hasil penerapan metode pada penelitian yang dilakukan prediksi dengan data uji dengan model klasifikasi yang diperoleh dari tahap pemodelan K-Nearest Neighbors.



Gambar 2. Diagram Alir

## HASIL PENELITIAN DAN PEMBAHASAN

### Deskripsi Data

Dalam penelitian ini data yang diperoleh dari SMAN 1 Gowa diuji menggunakan algoritma *K-Nearest Neighbors*. Data yang dikumpulkan sebanyak 216 peserta didik. Kemudian dilakukan Pre-processing dan cleaning data sehingga diperoleh 6 atribut. Atribut pertama X1 yaitu jenis kelamin, jenis kelamin dari peserta didik dengan skala pengukuran laki laki dan perempuan, selanjutnya X2 yaitu nilai ujian Semester, kemudian untuk X3 yaitu Nilai Ujian Harian, Selanjutnya X4 yaitu Persentase kehadiran dari peserta didik, Selanjutnya X5 yaitu Kelas dari Peserta didik dan yang terakhir X6 yaitu Semester dari Peserta didik.

### Pre-processing dan Cleaning Data

Pada proses cleaning data dilakukan penghapusan data yang *noise*, dan duplikasi data (Perwitasari et al., 2023). Fungsi yang digunakan dalam penelitian ini yaitu *filtering*, fungsi ini menyediakan opsi berbagai penyaringan data. Pada data daftar hadir dan nilai peserta didik semester ganjil 2025 2026, fungsi filter diaplikasikan pada variabel Y (Keterangan) sebanyak 216 data, tidak diperoleh dengan simbol "NA", sehingga data yang akan masuk ke dalam proses analisis yaitu 216 data. Selain fungsi filter, proses transformasi data juga dilakukan untuk menyesuaikan data yang dianalisis dengan proses data mining. Pada variabel Y, Status Lulus dikategorikan menjadi "0" dan Status Tidak Lulus dikategorikan menjadi "1".

### Split Data

Split Data adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian lainnya sebagai data testing (Wijiyanto et al., 2024). Dengan menggunakan Split Data akan dilakukan percobaan training berdasarkan split ratio yang telah ditentukan sebelumnya, untuk kemudian sisa dari split ratio data training akan dianggap sebagai data testing (Aftha Harianto et al., 2025). Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran.

**Tabel 1.** Skenario Split Data

| Pengujian | Training % | Testing % |
|-----------|------------|-----------|
| 1         | 80         | 20        |
| 2         | 70         | 30        |
| 3         | 60         | 40        |

## Pengujian Model *K-Nearest Neighbors*

Pada model klasifikasi *K-Nearest Neighbors* memasukkan data training untuk memprediksi peserta didik dalam pembelajaran matematika masuk ke kategori Lulus atau Tidak Lulus. Adapun tahapan dalam membangun model di antaranya Setelah data melalui beberapa proses di antara *Pre-processing* data, *Cleaning* data, kemudian data dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Untuk menghasilkan model maka diperlukan data *training*, sedangkan untuk data *testing* dimanfaatkan dalam melakukan evaluasi model yang telah dibuat. Pengujian ini dilakukan untuk mengetahui seberapa pengaruh banyaknya data training dan testing terhadap klasifikasi yang dilakukan oleh *K-Nearest Neighbors*. Evaluasi akurasi model klasifikasi menggunakan metode *confusion matrix* untuk mengetahui apakah model yang digunakan baik atau tidak.

Pada Tabel 1 akan dilakukan pengujian terhadap model *K-Nearest Neighbors* yang telah dibuat dengan dengan presentase pembagian data training dan testing pada waktu proses splitting data. Selanjutnya untuk mengetahui performa dari model *K-Nearest Neighbors* yaitu nilai *Accuracy*, *Precision*, *Recall*. Nilai *Accuracy* menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar dengan keseluruhan data. Nilai *Precision* menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. *Recall* menunjukkan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Nilai AUC memiliki rentang antara 0.5 sampai dengan 1. Interpretasi nilai AUC dapat diklasifikasikan menjadi lima bagian yang berbeda yaitu Akurasi salah, tingkat akurasi lemah, tingkat akurasi sedang, tingkat akurasi tinggi, dan tingkat akurasi tinggi. Proses pengujiannya akan dijabarkan satu persatu di bawah ini yaitu:

### Pengujian 1 KNN

Pada Pengujian 1 pembagian data training 80% dan data testing 20% dengan data training sebanyak 173 serta data testing sebanyak 43. Adapun hasil *Confusion Matrix* seperti dijelaskan pada Tabel di bawah ini :

**Tabel 2.** Confusion Matrix Pengujian 1

| Confusion Matrix |       |           |             |
|------------------|-------|-----------|-------------|
|                  |       | Predicted |             |
|                  |       | Lulus     | Tidak Lulus |
| Observed         | Lulus | 41        | 0           |

| Confusion Matrix |           |             |
|------------------|-----------|-------------|
|                  | Predicted |             |
|                  | Lulus     | Tidak Lulus |
| Tidak Lulus      | 2         | 0           |

Tabel 2 di atas adalah perhitungan berdasarkan data testing, diketahui dari 216 data, 41 diklasifikasikan Lulus sesuai dengan prediksi yang dilakukan dengan metode *K-Nearest Neighbors*, lalu 2 data diprediksi Lulus tetapi ternyata Tidak Lulus, 0 data diklasifikasikan Tidak Lulus diprediksi sesuai, dan 0 data diprediksi Tidak Lulus ternyata Lulus.

Nilai accuracy, precision, dan recall dari data training serta Nilai AUC yang hasil pengujian dengan menggunakan model *K-Nearest Neighbors* didapatkan hasil *accuracy* = 95.30%, *precision* = 90.90%, *recall* = 95.30%, Nilai Accuracy dan AUC pada table 3 dan tabel 4 di bawah ini:

**Tabel 3.** Accuracy dan AUC Pengujian 1

| K-Nearest Neighbors (KNN) |        |
|---------------------------|--------|
| <b>Accuracy</b>           | 95.30% |
| <b>AUC</b>                | 0.500  |

**Tabel 4.** Model Performance Metrics Pengujian

| Model Performance Metrics             |        |             |                 |
|---------------------------------------|--------|-------------|-----------------|
|                                       | Lulus  | Tidak Lulus | Average / Total |
| Support                               | 41     | 2           | 43              |
| Accuracy                              | 0.953  | 0.953       | 0.953           |
| Precision (Positive Predictive Value) | 0.953  | NaN         | 0.909           |
| Recall (True Positive Rate)           | 1.000  | 0.000       | 0.953           |
| False Positive Rate                   | 1.000  | 0.000       | 0.500           |
| False Discovery Rate                  | 0.047  | NaN         | 0.047           |
| F1 Score                              | 0.976  | NaN         | 0.931           |
| Matthews Correlation Coefficient      | NaN    | NaN         | NaN             |
| Area Under Curve (AUC)                | 0.500  | 0.500       | 0.500           |
| Negative Predictive Value             | NaN    | 0.953       | 0.953           |
| True Negative Rate                    | 0.000  | 1.000       | 0.500           |
| False Negative Rate                   | 0.000  | 1.000       | 0.500           |
| False Omission Rate                   | NaN    | 0.047       | 0.047           |
| Threat Score                          | 10.250 | 0.000       | 5.125           |
| Statistical Parity                    | 1.000  | 0.000       | 1.000           |

*Note.* All metrics are calculated for every class against all other classes.

## Pengujian 2 KNN

Pada Pengujian 2 pembagian data training 70% dan data testing 30% dengan data training sebanyak 152 serta data testing sebanyak 64. Adapun hasil *Confusion Matrix* seperti dijelaskan pada Tabel 5 di bawah ini :

**Tabel 5. Confusion Matrix Pengujian 2**

| <i>Confusion Matrix</i> |             |           |             |
|-------------------------|-------------|-----------|-------------|
|                         |             | Predicted |             |
|                         |             | Lulus     | Tidak Lulus |
| Observed                | Lulus       | 63        | 0           |
|                         | Tidak Lulus | 1         | 0           |

Tabel 5 di atas adalah perhitungan berdasarkan data testing, diketahui dari 216 data, 63 diklasifikasikan Lulus sesuai dengan prediksi yang dilakukan dengan metode *K-Nearest Neighbors*, lalu 1 data diprediksi Lulus tetapi ternyata Tidak Lulus, 0 data diklasifikasikan Tidak Lulus diprediksi sesuai, dan 0 data diprediksi Tidak Lulus ternyata Lulus.

Nilai accuracy, precision, dan recall dari data training serta Nilai AUC yang hasil pengujian dengan menggunakan model *K-Nearest Neighbors* didapatkan hasil *accuracy* = 98.40%, *precision* = 96.90%, *recall* = 98.40%, Nilai Accuracy dan AUC pada Tabel 6 dan tabel 7 di bawah ini:

**Tabel 6. Accuracy dan AUC Pengujian 2**

| <i>K-Nearest Neighbors (KNN)</i> |        |
|----------------------------------|--------|
| <i>Accuracy</i>                  | 98.40% |
| <i>AUC</i>                       | 0.984  |

**Tabel 7. Model Performance Metrics Pengujian 2**

| <i>Model Performance Metrics</i>                                                   |        |             |                 |
|------------------------------------------------------------------------------------|--------|-------------|-----------------|
|                                                                                    | Lulus  | Tidak Lulus | Average / Total |
| Support                                                                            | 63     | 1           | 64              |
| Accuracy                                                                           | 0.984  | 0.984       | 0.984           |
| Precision (Positive Predictive Value)                                              | 0.984  | NaN         | 0.969           |
| Recall (True Positive Rate)                                                        | 1.000  | 0.000       | 0.984           |
| False Positive Rate                                                                | 1.000  | 0.000       | 0.500           |
| False Discovery Rate                                                               | 0.016  | NaN         | 0.016           |
| F1 Score                                                                           | 0.992  | NaN         | 0.977           |
| Matthews Correlation Coefficient                                                   | NaN    | NaN         | NaN             |
| Area Under Curve (AUC)                                                             | 0.984  | 0.984       | 0.984           |
| Negative Predictive Value                                                          | NaN    | 0.984       | 0.984           |
| True Negative Rate                                                                 | 0.000  | 1.000       | 0.500           |
| False Negative Rate                                                                | 0.000  | 1.000       | 0.500           |
| False Omission Rate                                                                | NaN    | 0.016       | 0.016           |
| Threat Score                                                                       | 31.500 | 0.000       | 15.750          |
| Statistical Parity                                                                 | 1.000  | 0.000       | 1.000           |
| <i>Note.</i> All metrics are calculated for every class against all other classes. |        |             |                 |

## Pengujian 3 KNN

Pada Pengujian 3 pembagian data training 60% dan data testing 40% dengan data training sebanyak 130 serta data testing sebanyak 86. Adapun hasil *Confusion Matrix* seperti dijelaskan pada Tabel 8 di bawah ini :

**Tabel 8.** *Confusion Matrix* Pengujian 3

| <i>Confusion Matrix</i> |             |           |             |
|-------------------------|-------------|-----------|-------------|
|                         |             | Predicted |             |
|                         |             | Lulus     | Tidak Lulus |
| Observed                | Lulus       | 82        | 2           |
|                         | Tidak Lulus | 2         | 0           |

Tabel di atas adalah perhitungan berdasarkan data testing, diketahui dari 216 data, 82 diklasifikasikan Lulus sesuai dengan prediksi yang dilakukan dengan metode *K-Nearest Neighbors*, lalu 2 data diprediksi Lulus tetapi ternyata Tidak Lulus, 0 data diklasifikasikan Tidak Lulus diprediksi sesuai, dan 2 data diprediksi Tidak Lulus ternyata Lulus.

Nilai accuracy, precision, dan recall dari data training serta Nilai AUC yang hasil pengujian dengan menggunakan model *K-Nearest Neighbors* didapatkan hasil *accuracy* = 95.30%, *precision* = 95.30%, *recall* = 95.30%, Nilai Accuracy dan AUC pada tabel 9 dan tabel 10 di bawah ini:

**Tabel 9.** Accuracy dan AUC Pengujian 3

| <i>K-Nearest Neighbors (KNN)</i> |        |
|----------------------------------|--------|
| <i>Accuracy</i>                  | 95.30% |
| <i>AUC</i>                       | 0.500  |

**Tabel 10.** *Model Performance Metrics* Pengujian 3

| <i>Model Performance Metrics</i>      |        |             |                 |
|---------------------------------------|--------|-------------|-----------------|
|                                       | Lulus  | Tidak Lulus | Average / Total |
| Support                               | 84     | 2           | 86              |
| Accuracy                              | 0.953  | 0.953       | 0.953           |
| Precision (Positive Predictive Value) | 0.976  | 0.000       | 0.953           |
| Recall (True Positive Rate)           | 0.976  | 0.000       | 0.953           |
| False Positive Rate                   | 1.000  | 0.024       | 0.512           |
| False Discovery Rate                  | 0.024  | 1.000       | 0.512           |
| F1 Score                              | 0.976  | NaN         | 0.953           |
| Matthews Correlation Coefficient      | NaN    | NaN         | NaN             |
| Area Under Curve (AUC)                | 0.500  | 0.500       | 0.500           |
| Negative Predictive Value             | NaN    | 0.976       | 0.488           |
| True Negative Rate                    | 0.000  | 0.976       | 0.488           |
| False Negative Rate                   | 0.024  | 1.000       | 0.512           |
| False Omission Rate                   | NaN    | 0.024       | 0.512           |
| Threat Score                          | 13.667 | 0.000       | 6.833           |
| Statistical Parity                    | 0.977  | 0.023       | 1.000           |



*Note.* All metrics are calculated for every class against all other classes.

Dari hasil pengujian yang telah dilakukan sebanyak 3 kali pengujian didapatkan nilai rata-rata Accuracy, Precision, Recall seperti ditampilkan pada Tabel 11 sebagai berikut:

**Tabel 11.** Nilai Rata-Rata Accuracy, Precision, Recall dan AUC pada KNN

| Penguji | Training | Testin | Accuracy | Precision | Recall% | Nilai |
|---------|----------|--------|----------|-----------|---------|-------|
| 1       | 8        | 2      | 95,3     | 90,90     | 95,30   | 0,50  |
| 2       | 7        | 3      | 98,4     | 96,90     | 98,40   | 0,98  |
| 3       | 6        | 4      | 95,3     | 95,30     | 95,30   | 0,50  |
| Rata-   |          |        | 96,3     | 94,37     | 96,33   | 0,66  |

Pada Tabel 11 di atas terlihat bahwa model *K-Nearest Neighbors* mampu melakukan prediksi. Dengan 3 kali percobaan melakukan perubahan pada data training dan testing menunjukkan nilai rata-rata Accuracy sebesar 96,33%, Precision sebesar 94,37%, Recall sebesar 96,33% dan AUC 0,661.

## Hasil Pengujian K-Nearest Neighbors

Prediksi yang dilakukan dengan *Algortima K-Nearest Neighbors* mendapatkan hasil dari uji coba menyatakan bahwa presentase pembagian data training dan data testing sangat mempengaruhi hasil Accuracy, Precision, Recall dan Nilai AUC dalam proses pengujian. Uji coba dilakukan sebanyak 3 kali percobaan dengan skenario pembagian data training dan data testing secara berbeda dalam setiap tahap percobaannya. Pada hasil pengujian yang dilakukan menunjukkan bahwa nilai akurasi tertinggi didapat pada pengujian 3 yaitu data training 60% dan data testing 40% dengan data training sebanyak 130 serta data testing sebanyak 86 dengan menghasilkan Accuracy 98,80%, Precision 97,70%, Recall 98,80% dan Nilai AUC 0,500

## PENUTUP

Dalam penelitian ini dilakukan pembuatan dengan menggunakan model algoritma K-Nearest Neighbor menggunakan data daftar hadir dan nilai peserta didik semester ganjil 2025 2026 yang diperoleh dari SMAN 1 Gowa. Model yang dihasilkan baik dalam penentuan tingkat akurasi *K-Neares Neighbors* terhadap pemahaman materi peserta didik dalam pembelajaran matematika.. Untuk mengukur kinerja algoritma tersebut digunakan metode pengujian Cross Validation dan Confusion Matrix, diketahui bahwa algoritma *K-Nearest Neighbors* memiliki nilai AUC tinggi. Dari hasil analisis tersebut dapat disimpulkan bahwa metode *K-Nearest Neighbors* merupakan metode yang cukup baik dalam menentukan tingkat akurasi terhadap pemahaman materi peserta didik dalam pembelajaran matematika.

Meskipun *K-Nearest Neighbors* merupakan metode yang cukup baik dalam menentukan tingkat akurasi terhadap pemahaman materi peserta didik dalam pembelajaran matematika, namun ada beberapa hal yang dapat ditambahkan agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Bagi peneliti selanjutnya dapat menggunakan lebih banyak jumlah sampel dan variabel prediktor.
2. Pada penelitian berikutnya dapat dikembangkan dengan objek yang lain.

---

## DAFTAR PUSTAKA

- Aftha Harianto, F. R., Alawi, Z., & Sa'ida, I. A. (2025). Pengaruh Komposisi Split Data Pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning. *Jurnal Sistem Informasi Dan Informatika (Simika)*, 8(1), 36–44. <https://doi.org/10.47080/simika.v8i1.3663>
- Andika, A. Wira, Nurhakim, L., & Andas, Netty Huzniati. (2024). Penggunaan Deep Learning Untuk Memprediksi Using Deep Learning To Predict Academic Performance and. *Jurnal Ilmiah Bidang Sosial Ekonomi Budaya Teknologi Dan Pendidikan*, 4(7), 1647–1664.
- Anggraini, J., & Alita, D. (2024). Implementasi Metode KNN Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter. *Jurnal Informatika: Jurnal Pengembangan IT*, 9(2), 102–111. <https://doi.org/10.30591/jpit.v9i2.6560>
- Aprilia, P. (n.d.). Cara Penanganan Siswa Berkemampuan di Atas Rata-Rata. *Journal of Knowledge and Collaboration*.
- Dwitama, R. P. (2025). Analisis Komparatif Metode Naive Bayes Dan K-Nearest Neighbors (KNN) untuk Klasifikasi Gender Berdasarkan Komentar Pada Platform X (Twitter). <https://dspace.uui.ac.id/handle/123456789/58988%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/58988/21523151.pdf?sequence=1>
- Fatimah, N. S., & Agustin, S. (2025). Klasifikasi Citra Batik Menggunakan Local Binary Pattern (LBP) dan Support Vector Machine (KNN). *Jurnal Algoritma*, 22(1), 185–196. <https://doi.org/10.33364/algoritma/v.22-1.2208>
- Perwitasari, A., Septiriana, R., & Tursina, T. (2023). Data preparation Structure untuk Pemodelan Prediktif Jumlah Peserta Ajar Matakuliah. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 9(1), 7. <https://doi.org/10.26418/jp.v8i3.57321>
- Setyawan, A., Kiswanto, R. H., & Sutejo, H. (2025). Penerapan Metode K-Nearest Neighbors Untuk Memprediksi Kelulusan Tepat Waktu. *Progresif: Jurnal Ilmiah Komputer*, 21(2), 768