

# Peningkatan Akurasi Model Untuk Prediksi KKM Siswa Sekolah Dasar Menggunakan *Supervised Machine Learning* dengan Integrasi Faktor Internal dan Eksternal

Arham Rahim dan Mustakim Mustakim



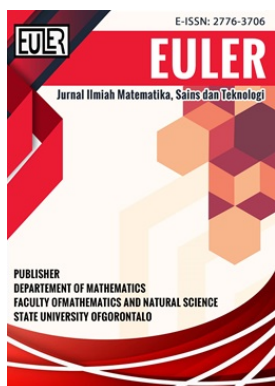
Volume 13, Issue 3, Pages 372–380, Dec. 2025













Diterima 26 September 2025, Direvisi 20 November 2025, Disetujui 26 November 2025, Diterbitkan 1 Desember 2025

To Cite this Article : A. Rahim dan M. Mustakim, “Peningkatan Akurasi Model Untuk Prediksi KKM Siswa Sekolah Dasar Menggunakan *Supervised Machine Learning* dengan Integrasi Faktor Internal dan Eksternal”, *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 3, pp. 372–380, 2025, <https://doi.org/10.37905/euler.v13i3.34577>

© 2025 by author(s)

## JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI

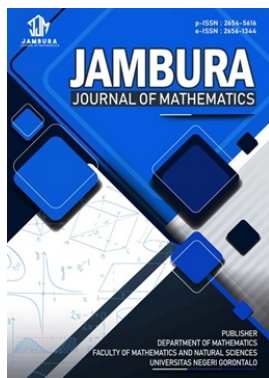


-  Homepage : <http://ejournal.ung.ac.id/index.php/euler/index>
-  Journal Abbreviation : Euler J. Ilm. Mat. Sains dan Teknol.
-  Frequency : Three times a year
-  Publication Language : English (preferable), Indonesia
-  DOI : <https://doi.org/10.37905/euler>
-  Online ISSN : 2776-3706
-  License : Creative Commons Attribution-NonCommercial 4.0 International License
-  Publisher : Department of Mathematics, Universitas Negeri Gorontalo
-  Country : Indonesia
-  OAI Address : <http://ejournal.ung.ac.id/index.php/euler/oai>
-  Google Scholar ID : QF\_r\_gAAAAJ
-  Email : [euler@ung.ac.id](mailto:euler@ung.ac.id)

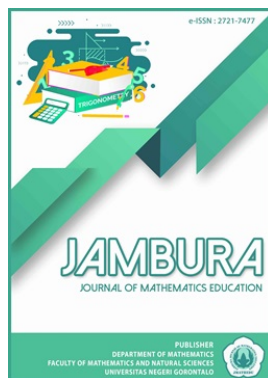
## JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

# Peningkatan Akurasi Model Untuk Prediksi KKM Siswa Sekolah Dasar Menggunakan *Supervised Machine Learning* dengan Integrasi Faktor Internal dan Eksternal

Arham Rahim<sup>1,\*</sup>, Mustakim Mustakim<sup>2</sup>

<sup>1</sup>Program Studi Teknologi Pendidikan, Universitas Madako Tolitoli, Tambun 94514, Indonesia

<sup>2</sup>Program Studi Pendidikan Guru Sekolah Dasar, Universitas Madako Tolitoli, Tambun 94514, Indonesia

## ARTICLE HISTORY

Diterima 26 September 2025

Direvisi 20 November 2025

Disetujui 26 November 2025

Diterbitkan 1 Desember 2025

## KATA KUNCI

Kriteria Ketuntasan Minimal (KKM)  
Pembelajaran Mesin Terawasi  
Faktor Internal Siswa  
Faktor Eksternal Siswa

## KEYWORDS

Minimum Competency Criteria  
(KKM)  
Supervised Machine Learning  
Student Internal Factors  
Student External Factors

**ABSTRAK.** Kriteria Ketuntasan Minimal (KKM) merupakan standar penilaian ketercapaian kompetensi siswa di sekolah dasar di Indonesia dan menjadi indikator penting keberhasilan belajar. Namun, masih banyak siswa yang mengalami kesulitan dalam memenuhi standar tersebut, sehingga diperlukan strategi deteksi dini berbasis data untuk mendukung intervensi yang tepat. Penelitian ini bertujuan membangun model prediksi capaian KKM siswa berdasarkan faktor internal dan eksternal menggunakan pendekatan supervised machine learning. Data internal meliputi nilai rapor dan kehadiran, sedangkan data eksternal diperoleh dari respons siswa dan informasi orang tua yang mencakup aspek lingkungan, ekonomi, motivasi, serta dukungan keluarga. Empat algoritma machine learning yang diuji adalah Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Neural Network, yang dievaluasi menggunakan confusion matrix. Eksperimen dilakukan melalui empat skenario preprocessing data, yaitu reverse scoring, feature selection, normalization, dan variable grouping. Hasil terbaik diperoleh pada skenario S3 dengan kombinasi normalization dan feature selection menggunakan algoritma SVM dengan akurasi 100%. Namun, untuk menghindari potensi overfitting, algoritma yang lebih stabil direkomendasikan, yaitu Naïve Bayes dengan akurasi 93%. Hasil ini menunjukkan bahwa penerapan machine learning dengan preprocessing yang tepat efektif dalam mengidentifikasi siswa yang berpotensi gagal mencapai KKM.

**ABSTRACT.** The Minimum Mastery Criteria (KKM) is a standard used to assess students' competency achievement in elementary schools in Indonesia and serves as an important indicator of learning success. However, many students still have difficulties meeting this standard, thus requiring a data-driven early detection strategy to support timely intervention. This study aims to develop a prediction model for students' KKM achievement based on internal and external factors using a supervised machine learning approach. Internal data include report card scores and attendance, while external data are obtained from student responses and parental information covering environmental, economic, motivational, and family support aspects. Four machine learning algorithms were evaluated, namely Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Neural Network, using a confusion matrix. Experiments were conducted under four data preprocessing scenarios: reverse scoring, feature selection, normalization, and variable grouping. The best result was obtained in Scenario S3, which combines normalization and feature selection, using the SVM algorithm with 100% accuracy. However, to avoid potential overfitting, a more stable algorithm is recommended, namely Naïve Bayes, which achieved 93% accuracy. These results indicate that the application of machine learning with appropriate preprocessing is effective for identifying students at risk of not achieving the KKM.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. **Editorial of EULER:** Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

## 1. Pendahuluan

Pendidikan dasar merupakan fondasi penting dalam pembentukan karakter dan kompetensi siswa. Pada tahap ini, peserta didik dituntut menguasai aspek kognitif, afektif, dan psiko-motor agar dapat melanjutkan ke jenjang pendidikan berikutnya dengan baik [1, 2]. Salah satu instrumen penilaian capaian pembelajaran yang berlaku di Indonesia adalah Kriteria Ketuntasan Minimal (KKM). KKM digunakan sebagai standar untuk menen-

tukan apakah siswa telah mencapai kompetensi tertentu, serta menjadi acuan keberhasilan suatu program pendidikan [3]. Pencapaian KKM juga berimplikasi pada kenaikan kelas dan kelulusan, sehingga kegagalannya berdampak serius bagi perkembangan akademik siswa [4, 5].

Dalam praktiknya, pencapaian KKM di sekolah dasar masih menghadapi tantangan. Hasil penelitian menunjukkan bahwa banyak siswa mengalami kesulitan dalam memenuhi standar minimal, khususnya pada mata pelajaran inti. Rendahnya pencapai-

\*Penulis Korespondensi.

ian KKM tidak hanya menggambarkan keterbatasan siswa secara individual [5], tetapi juga menjadi indikator bahwa strategi pembelajaran maupun dukungan eksternal belum optimal [6, 7]. Oleh karena itu, deteksi dini terhadap potensi penurunan capaian KKM sangat diperlukan agar guru dapat memberikan intervensi tepat waktu.

Seiring perkembangan teknologi, muncul pendekatan baru dalam memprediksi capaian akademik, yaitu dengan *machine learning*. *Machine learning* adalah cabang kecerdasan buatan yang memungkinkan komputer belajar dari data historis dan meningkatkan kinerjanya dalam memprediksi hasil [8]. Salah satu metodenya, yaitu *supervised machine learning*, melatih model berdasarkan data berlabel untuk memprediksi capaian masa depan [9]. Keunggulan pendekatan ini terletak pada kemampuannya menemukan pola kompleks dalam data yang sulit diidentifikasi secara manual [9].

Beberapa penelitian telah menerapkan *supervised machine learning* dalam konteks pendidikan [10–12]. Penelitian terdahulu yang khusus untuk memprediksi capaian KKM menggunakan algoritma *Neural Network* memberikan akurasi terbaik sebesar 89% dengan *recall* sempurna menggunakan data internal berupa nilai kognitif, afektif, dan psikomotor [13]. Hasil ini memperlihatkan potensi besar *machine learning* sebagai sistem deteksi dini bagi guru. Namun, penelitian tersebut masih memiliki keterbatasan karena hanya menyoroti faktor internal siswa. Padahal, capaian akademik juga dipengaruhi oleh faktor eksternal.

Faktor eksternal seperti kondisi keluarga, dukungan lingkungan belajar, status sosial-ekonomi, dan kehadiran siswa terbukti berpengaruh signifikan terhadap hasil belajar [14, 15]. Performa siswa tidak hanya dipengaruhi karakteristik internal, melainkan juga interaksi kompleks dengan lingkungannya [16]. Misalnya, motivasi belajar yang rendah akibat keterbatasan dukungan keluarga dapat menjadi penyebab siswa gagal mencapai KKM meskipun nilai kognitifnya cukup baik. Dengan demikian, pengabaian faktor eksternal dalam pemodelan prediksi berpotensi menurunkan akurasi sistem.

Literatur lain juga mendukung perlunya integrasi faktor eksternal dan internal [17, 18]. Studi sebelumnya mencatat bahwa capaian akademik siswa dipengaruhi oleh lingkungan keluarga dan masyarakat, yang dalam jangka panjang berdampak pada kualitas pendidikan [15, 19, 20]. Dengan kata lain, capaian KKM adalah hasil dari gabungan faktor internal dan eksternal yang saling berinteraksi. Oleh karena itu, membangun model prediksi yang hanya mengandalkan data internal tidak cukup untuk memberikan gambaran menyeluruh.

Penelitian sebelumnya hanya menggunakan faktor internal berupa nilai kognitif, afektif, dan psikomotor [10, 11, 13]. Penelitian ini menambahkan faktor eksternal seperti kondisi keluarga, kehadiran, dan sosial-ekonomi. Pendekatan ini menawarkan analisis yang lebih holistik dan realistis. Jika penelitian lain banyak menggunakan *dataset* sekunder atau sumber terbuka, penelitian ini menggunakan data primer langsung dari sekolah dasar di Kabupaten Tolitoli. Hal ini meningkatkan validitas hasil karena sesuai dengan konteks lokal pendidikan Indonesia.

Penelitian ini tidak hanya menguji satu algoritma, tetapi membandingkan performa empat algoritma, yaitu *Naïve Bayes*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, dan *Neural Network*, menggunakan *confusion matrix* untuk mengevaluasi aku-

rasi, presisi, dan *recall*. Dengan demikian, hasil yang diperoleh lebih komprehensif dibandingkan penelitian sebelumnya. Model yang dibangun diarahkan untuk menjadi sistem deteksi dini (*early warning system*). Dengan sistem ini, guru dapat lebih cepat mengidentifikasi siswa berisiko gagal mencapai KKM dan segera merancang intervensi pedagogis sesuai penyebabnya.

Kebaruan dari penelitian ini dibandingkan dengan penelitian terdahulu yang telah memanfaatkan *machine learning* untuk menganalisis data siswa adalah bahwa penelitian yang dilakukan oleh Hartanti dkk. [10] dan Saputra dkk. [12] tidak menjelaskan secara spesifik tingkatan sekolah yang diteliti, sedangkan penelitian ini berfokus pada tingkat sekolah dasar. Penelitian yang dilakukan oleh Purwaningsih dan Nurelasari [11] juga memanfaatkan *machine learning* pada tingkat sekolah dasar, namun hanya berfokus pada kelulusan siswa berdasarkan data kelas IV, sedangkan penelitian ini berfokus pada nilai KKM siswa di kelas atas berdasarkan data internal dan eksternal di kelas bawah.

Berdasarkan kajian tersebut, penting dilakukan penelitian lanjutan yang tidak hanya melibatkan faktor internal siswa, tetapi juga memasukkan faktor eksternal agar model prediksi yang dihasilkan lebih akurat dan aplikatif di lapangan. Integrasi kedua faktor ini diharapkan mampu meningkatkan performa sistem deteksi dini capaian KKM. Menggabungkan faktor internal dan faktor eksternal siswa menjadikan model prediksi memiliki cakupan variabel yang lebih komprehensif, sehingga mampu menangkap pola yang lebih kompleks dan mengurangi bias. Oleh karena itu, penelitian ini difokuskan pada dua permasalahan utama, yaitu bagaimana menentukan faktor eksternal siswa yang dapat meningkatkan akurasi model prediksi nilai KKM siswa dan bagaimana membangun sebuah model yang dapat melakukan prediksi nilai KKM siswa berdasarkan data faktor internal dan eksternal.

## 2. Metode

Penelitian ini menerapkan pendekatan kuantitatif dengan desain *experimental design*. Fokus desain ini adalah melakukan serangkaian skenario pengujian sistematis terhadap *dataset* untuk memvalidasi *predictive model*. Sumber data penelitian diperoleh dari Sekolah Dasar Negeri Lantapan. Pengambilan sampel data menggunakan teknik *purposive sampling* guna memastikan kelengkapan dan relevansi data. Kriteria khusus utama yang ditetapkan mencakup dua komponen data yang diintegrasikan: (1) faktor internal, yang terdiri dari data arsip berupa rekam jejak historis KKM siswa dari kelas 1, 2, dan 3; serta (2) faktor eksternal, yang datanya diperoleh melalui *questionnaire* tanggapan dari wali siswa dan siswa terkait.

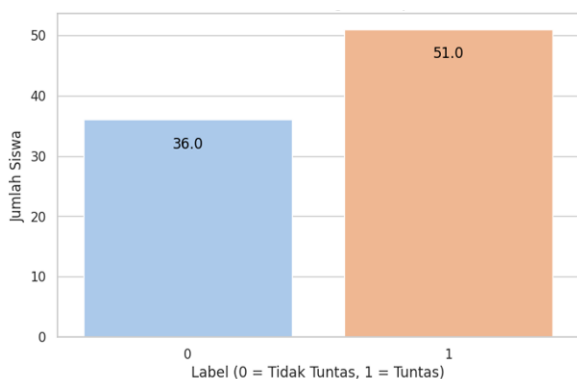
Setelah *dataset* representatif ini terkonsolidasi, proses dilanjutkan ke tahap eksperimen. Tahapan ini mencakup penerapan berbagai teknik *data preprocessing* untuk optimasi. *Dataset* yang telah diproses kemudian dilatih menggunakan beberapa algoritma *supervised machine learning*, meliputi *Naïve Bayes*, *K-Nearest Neighbors*, *Support Vector Machines*, serta *Neural Networks*. Tujuan akhir dari pengujian komparatif ini adalah untuk mengevaluasi dan menganalisis algoritma dan teknik *preprocessing*, sehingga dapat menghasilkan model yang paling akurat untuk memprediksi capaian KKM siswa.

**Tabel 1.** Skenario pengolahan data

No.	Skenario	Reverse scoring	Gabungan data	Seleksi fitur	Normalisasi	Pengelompokan variabel
1.	S1	✓	✓			
2.	S2	✓	✓	✓		
3.	S3		✓	✓	✓	
4.	S4	✓	✓		✓	✓

**2.1. Metode Pengumpulan Data**

Metode pengumpulan data dalam penelitian ini dilakukan melalui dua cara utama, yaitu angket dan dokumentasi. Angket digunakan untuk mengumpulkan data eksternal yang diberikan kepada siswa dan wali siswa. Angket siswa berisi 42 butir pertanyaan terkait lingkungan sosial sekolah, fasilitas sekolah, kondisi rumah tangga, ekonomi, pemanfaatan teknologi, motivasi belajar, hingga kegiatan ekstrakurikuler. Angket orang tua memuat 36 butir pertanyaan mengenai pendampingan belajar, pengawasan, dukungan emosional, kondisi ekonomi keluarga, dan ketersediaan sarana belajar di rumah. Kedua instrumen ini dimaksudkan untuk menggambarkan faktor eksternal yang berasal dari lingkungan personal maupun keluarga. Kedua, dokumentasi digunakan untuk mengumpulkan data internal berupa nilai rapor pada mata pelajaran, catatan kehadiran serta label capaian siswa. Sebagai informasi pendukung, wawancara dilakukan dengan guru untuk memperoleh informasi mengenai faktor eksternal yang dipandang berpengaruh terhadap proses pembelajaran siswa. Semua data yang terkumpul akan diberi label sesuai dengan nilai KKM siswa yang terbaru. Label 0 untuk nilai KKM yang tidak tuntas dan 1 untuk nilai KKM yang tuntas. Distribusi frekuensi label dari data yang diperoleh dapat dilihat pada Gambar 1.



**Gambar 1.** Distribusi frekuensi label

Seluruh data dari kedua metode tersebut digabungkan untuk membentuk dataset yang mencakup dimensi internal dan eksternal, untuk mendukung analisis prediksi capaian Kriteria Ketuntasan Minimal (KKM) siswa. Hasil pengumpulan data menunjukkan bahwa dari 87 siswa yang menjadi sampel, terdapat 36 siswa (41,38%) yang tidak mencapai KKM dan 51 siswa (58,62%) yang berhasil mencapainya. Distribusi ini relatif seimbang, sehingga dataset dinilai cukup representatif untuk digunakan dalam pelatihan dan pengujian model prediksi. Kondisi ini juga memperlihatkan adanya variasi capaian belajar antar siswa yang penting untuk dianalisis lebih lanjut.

**2.2. Metode Pengolahan Data**

Dalam penelitian ini, proses pengolahan data disusun melalui beberapa skenario yang bertujuan mengevaluasi pengaruh teknik *preprocessing* terhadap performa model. Skenario pengujian pada penelitian ini dapat dilihat pada Tabel 1.

Pada skenario S1, melibatkan *reverse scoring* dan gabungan data. *Reverse scoring* adalah butir angket yang dirumuskan dalam bentuk negatif atau berlawanan dengan konstruk yang diukur yang bertujuan untuk mengurangi *acquiescence bias*, yakni kecenderungan responden menyetujui pernyataan tanpa pertimbangan. Dengan adanya item *reversed*, responden terdorong lebih berhati-hati membaca setiap pernyataan sehingga jawaban yang diberikan lebih mencerminkan kondisi sebenarnya [22]. Pada item angket siswa yang memiliki arah interpretasi terbalik, kemudian data dari tiga sumber digabung tanpa melalui tahap normalisasi ataupun seleksi fitur, sehingga skenario ini berfungsi sebagai *baseline* dengan tingkat pengolahan minimal.

Pada skenario S2, *reverse scoring* dan penggabungan data tetap diterapkan tetapi ditambahkan prosedur seleksi fitur. *Feature selection* adalah proses memilih *subset* dari keseluruhan variabel (fitur) dalam *dataset* yang dianggap paling relevan untuk memprediksi variabel target, sambil menghapus fitur yang tidak relevan, redundan, atau *noise* [23]. Pada tahap ini, variabel dengan korelasi absolut terhadap label kurang dari 0,3 dieliminasi sehingga jumlah fitur menyusut dari 122 menjadi sekitar 81. Pendekatan ini bertujuan menekan keberadaan *noise* sekaligus mengurangi risiko *overfitting* akibat fitur yang tidak relevan.

Pada skenario S3, mencakup langkah *preprocessing* yang lebih komprehensif meliputi *reverse scoring*, gabungan data, seleksi fitur dan normalisasi. Normalisasi adalah prosedur *data preprocessing* dimana fitur numerik diubah agar memiliki skala yang sama atau rentang yang sebanding [24]. Proses normalisasi dilakukan pada nilai akademik dari skala 0–100 diubah ke dalam rentang 0–1, normalisasi kehadiran, serta transformasi jawaban angket orang tua dari skala 1–5 ke 0–1. Selanjutnya, *feature selection* dilakukan kembali dengan ambang korelasi < 0,3 terhadap label, sehingga diperoleh *dataset* yang lebih bersih.

Pada skenario S4, proses dirancang dengan pendekatan yang berbeda. Selain *reverse scoring*, gabungan data dan normalisasi, skenario ini menerapkan proses pengelompokan variabel. *Variable grouping* adalah metode untuk menyusun fitur-fitur yang memiliki kemiripan atau keterkaitan tinggi ke dalam kelompok atau *clusters*. Tujuannya adalah mengurangi dimensi data, mengurangi redundansi, dan meningkatkan *interpretability* model [25]. Pada proses ini, *variable grouping* diterapkan pada angket siswa dan angket orang tua. Pada angket siswa, item-item pertanyaan dikelompokkan ke beberapa kategori, antara lain Sosial di Sekolah, Fasilitas Sekolah, Kondisi Rumah, Ekonomi, Teknologi, Motivasi, serta Kegiatan Ekstrakurikuler. Nilai dari setiap kategori kemudian dihitung rata-ratanya untuk menghasilkan *composite score* yang merepresentasikan aspek tersebut. Prosedur serupa dite-

```

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, stratify=y, random_state=16)

print("Distribusi label di y_train:\n", y_train.value_counts())
print("Distribusi label di y_test:\n", y_test.value_counts())

Distribusi label di y_train:
LABEL
1    35
0    25
Name: count, dtype: int64
Distribusi label di y_test:
LABEL
1    16
0    11
Name: count, dtype: int64
    
```

Gambar 2. Pembagian data

rapkan pada angket orang tua, dengan pembentukan kelompok seperti Pendampingan, Monitoring, Fasilitas Rumah, Lingkungan Rumah, Sosial, Ekonomi, Teknologi, dan Dukungan Emosional. Melalui langkah ini, jumlah variabel berhasil direduksi menjadi 59 fitur sehingga *dataset* menjadi lebih ringkas.

Setelah melakukan pengolahan data sesuai dengan skenario, selanjutnya dilakukan pembagian *training data* dan *testing data* dengan skema 70:30 secara *stratified sampling*, sehingga distribusi kelas “Mencapai KKM” dan “Tidak Mencapai KKM” tetap seimbang dalam kedua *subsets*. Dengan cara ini, model dapat dilatih menggunakan mayoritas data sambil menyisihkan sebagian data untuk evaluasi performa. Proses pembagian dilakukan secara otomatis dan hasilnya dapat dilihat pada Gambar 2.

### 2.3. Metode Analisis Data

Penelitian ini menggunakan pendekatan kuantitatif dengan penerapan beberapa algoritma *machine learning*, yaitu *Naïve Bayes*, *Support Vector Machine*, *Neural Network*, dan *K-Nearest Neighbor* yang diimplementasikan menggunakan bahasa pemrograman *Python*. Eksperimen dilakukan untuk menentukan algoritma dengan tingkat akurasi terbaik dalam memprediksi capaian KKM siswa tingkat atas berdasarkan data dari tingkat bawah. Proses *training model* dalam *machine learning* merupakan tahap ketika algoritma mempelajari pola dari *training data* guna menghasilkan prediksi atau keputusan pada data baru. Pada tahap ini, parameter model disesuaikan agar kesalahan prediksi dapat diminimalkan dibandingkan dengan nilai aktual.

*Naïve Bayes* (NB) merupakan metode klasifikasi berbasis probabilitas yang mengacu pada teori inferensi Bayes dengan asumsi independensi antar fitur. Algoritma ini digunakan untuk menghitung peluang suatu data termasuk ke dalam kelas tertentu [26]. NB banyak diterapkan dalam proses pengelompokan atau kategorisasi data yang diperoleh dari *dataset* hasil pelatihan. Prinsip dasarnya mengikuti Teorema Bayes sebagai berikut:

$$P(C | A) = \frac{P(A | C) P(C)}{P(A)} \tag{1}$$

di mana:

- $P(A | C)$  adalah probabilitas hipotesis  $A$  diberikan bukti  $C$  (*posterior*);
- $P(C | A)$  adalah probabilitas bukti  $C$  diberikan bahwa hipotesis  $A$  benar (*likelihood*);

- $P(A)$  adalah probabilitas awal dari hipotesis  $A$  (*prior*);
- $P(C)$  adalah probabilitas dari bukti  $C$  (*evidence*).

*Support Vector Machine* (SVM) merupakan algoritma *supervised learning* yang efektif digunakan untuk tugas klasifikasi maupun regresi. Prinsip kerja SVM adalah mencari *hyperplane* terbaik yang dapat memisahkan data ke dalam dua kelas dengan *maximum margin* [27]. Dalam pendekatan linier, model SVM membangun *optimal hyperplane* melalui persamaan tertentu.

$$f(w, x) = w \cdot x + b. \tag{2}$$

Keterangan:

- $w$  adalah vektor bobot (*weight vector*);
- $x$  adalah vektor fitur (*feature vector*);
- $b$  adalah *bias* atau *intercept*.

SVM mempelajari parameter dengan memecahkan masalah optimisasi dengan menggunakan persamaan berikut:

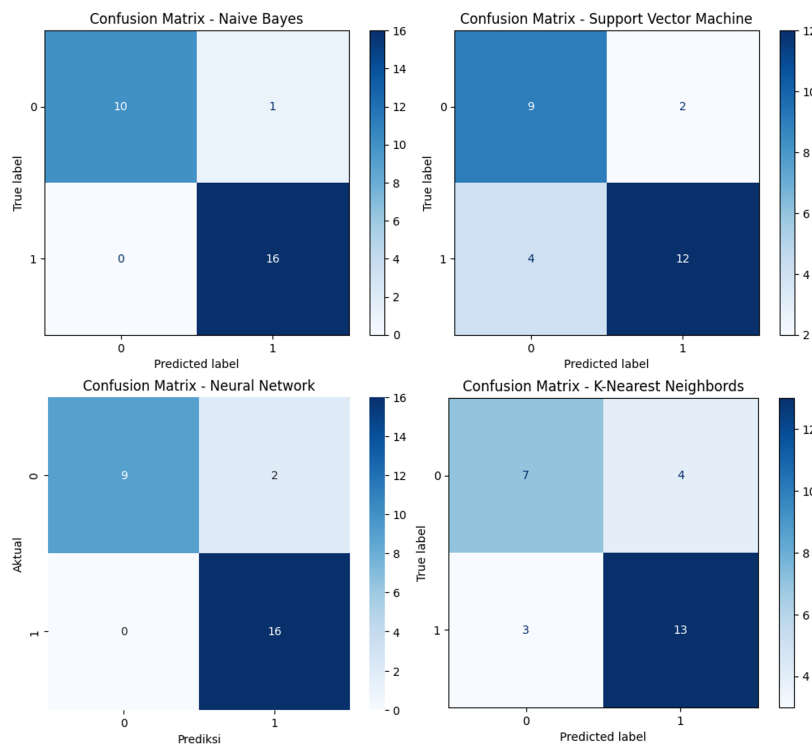
$$\min \left( \frac{1}{2} w^T w + C \sum_{i=1}^p \max(0, 1 - y'_i (w^T x_i + b)) \right). \tag{3}$$

Keterangan:

- $w^T w$  adalah norma Euclidean (*L2 norm*), yang digunakan untuk regularisasi model;
- $C$  adalah parameter penalti (*penalty parameter*), yang dapat ditentukan secara arbitrer atau melalui *hyperparameter tuning*;
- $y'_i$  adalah label aktual (*true label*);
- $w^T x_i + b$  adalah fungsi prediktor (*predictor function*).

*Neural Network* (NN) adalah algoritma yang terinspirasi dari cara kerja neuron dalam otak manusia, di mana setiap neuron saling terhubung dan mengalirkan informasi ke neuron lainnya. Secara umum, NN terdiri dari tiga lapisan utama, yaitu lapisan input, lapisan tersembunyi (*hidden layer*), dan lapisan output. Jumlah lapisan tersembunyi dapat bervariasi, bisa hanya satu atau lebih, dan hingga kini belum ada ketentuan pasti mengenai jumlah yang ideal. Pada dasarnya, sebuah neuron dalam NN berfungsi melakukan pemetaan dari input ke output, sehingga menjadi elemen paling mendasar dalam membangun jaringan tersebut [28].

*K-Nearest Neighbors* (KNN) merupakan algoritma *machine learning* yang dapat digunakan baik untuk klasifikasi maupun regresi. KNN termasuk algoritma non-parametrik karena tidak mensyaratkan asumsi tertentu terhadap distribusi data. Sebagai al-



Gambar 3. Confusion matrix hasil testing model skenario S1

goritma berbasis *instance*, KNN menyimpan seluruh data pelatihan dan menentukan kelas dari data baru dengan membandingkan tingkat kesamaan terhadap data yang sudah ada [29]. Proses klasifikasi KNN meliputi beberapa tahapan: (1) menentukan nilai *k* sebagai jumlah tetangga terdekat, biasanya bilangan ganjil untuk menghindari hasil imbang; (2) menghitung *Euclidean distance* antara data baru dan seluruh data latih; (3) mengurutkan jarak tersebut mulai dari yang terkecil; (4) menetapkan kategori berdasarkan mayoritas label tetangga terdekat; dan (5) menghasilkan prediksi akhir sesuai dengan kategori yang paling dominan. Persamaan matematis yang digunakan untuk menghitung jarak dalam metode ini adalah sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}. \tag{4}$$

Keterangan:

- $d_{ij}$  : jarak antara objek  $i$  dengan objek  $j$ ;
- $x_{ik}$  : nilai objek  $i$  pada variabel ke- $k$ ;
- $x_{jk}$  : nilai objek  $j$  pada variabel ke- $k$ ;
- $p$  : banyaknya variabel.

Untuk mendapatkan model dengan akurasi terbaik, skenario percobaan kemudian dievaluasi dengan menggunakan metode *confusion matrix* untuk menilai keakuratan dalam mengklasifikasikan nilai KKM siswa, sehingga dapat diketahui algoritma mana yang paling sesuai untuk permasalahan tersebut.

### 3. Hasil dan Pembahasan

Dalam upaya memprediksi apakah seorang siswa dapat mencapai nilai KKM atau tidak, penelitian ini menguji kinerja empat algoritma *machine learning*, yaitu *Naive Bayes*, *Support Vector Machine*, *Neural Network*, dan *K-Nearest Neighbor*, pada berbagai

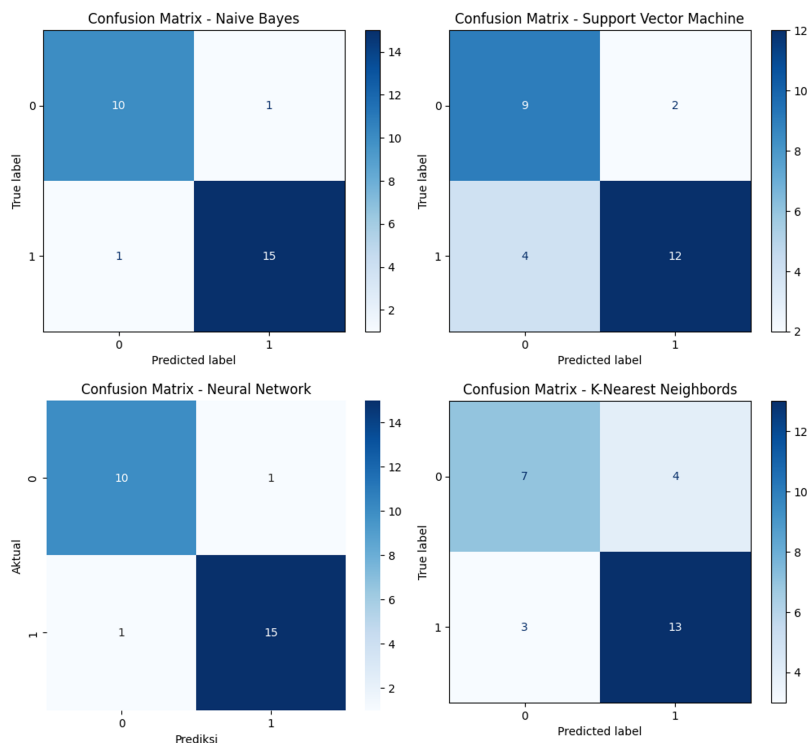
skenario *preprocessing* data.

Algoritma *Naive Bayes* mengasumsikan bahwa semua fitur bersifat independen. Algoritma NB menghitung probabilitas kemunculan setiap variabel, seperti probabilitas siswa memiliki motivasi tinggi jika siswa mencapai nilai KKM, dan probabilitas jika siswa tidak mencapai nilai KKM. Untuk data uji, model mengalikan probabilitas-probabilitas tersebut dan memprediksi kelas yang memiliki nilai probabilitas akhir tertinggi, sebagaimana diilustrasikan oleh pers. (1).

Algoritma *Support Vector Machine* (SVM) berupaya memaksimalkan jarak (*margin*) antara *hyperplane* dan titik data terdekat (*support vectors*). Posisi *linear hyperplane* ini didefinisikan oleh pers. (2), sedangkan fungsi optimasi dan penggunaan *kernel* yang membentuk *hyperplane* tersebut ditunjukkan oleh pers. (3). Untuk data uji, model memprediksi kelas berdasarkan sisi *hyperplane* tempat data tersebut berada.

Algoritma *Neural Network* (NN) menerima fitur input berupa faktor internal dan eksternal, kemudian memprosesnya melalui *hidden layers* yang menerapkan berbagai transformasi menggunakan fungsi aktivasi. Selama proses *training*, bobot koneksi disesuaikan secara berulang (*epoch*) untuk meminimalkan *prediction error*. Untuk data uji, input diproses melalui bobot yang telah dilatih, dan lapisan output menghasilkan probabilitas prediksi.

Algoritma *K-Nearest Neighbor* (KNN) memprediksi kelas dengan menghitung jarak antara siswa dalam data uji dan seluruh siswa dalam data latih, menggunakan metrik jarak yang diilustrasikan oleh pers. (4). Algoritma kemudian memilih  $K$  tetangga terdekat (*K-Nearest Neighbors*). Prediksi kelas KKM siswa dalam data uji ditentukan berdasarkan mayoritas label dari  $K$  tetangga terdekat tersebut. Hasil evaluasi dari setiap algoritma ditampilkan sebagai berikut.



Gambar 4. Confusion matrix hasil testing model skenario S2

3.1. Skenario S1

Pada skenario S1, data yang digunakan masih dalam bentuk mentah setelah dilakukan *reverse scoring* terhadap butir angket yang memiliki arah interpretasi negatif. Semua variabel asli dari angket siswa, angket orang tua, nilai akademik, dan kehadiran tetap dipertahankan, sehingga dimensi data tidak mengalami reduksi. Hal ini menjadikan *dataset* tetap kompleks dan penuh dengan keragaman informasi. Hasil evaluasi untuk masing-masing algoritma dapat dilihat pada Gambar 3.

Berdasarkan Gambar 3, dengan menggunakan *confusion matrix* diperoleh nilai *accuracy*, *precision*, dan *recall*. Hasil pengujian masing-masing model dapat dilihat pada Tabel 2.

Tabel 2. Rekap hasil testing model skenario S1

Model	Accuracy	Kelas	Precision	Recall
Naive Bayes	96%	0	100%	91%
		1	94%	100%
Support Vector Machine	78%	0	69%	82%
		1	86%	75%
Neural Network	93%	0	100%	82%
		1	89%	100%
K-Nearest Neighbor	74%	0	70%	64%
		1	76%	81%

Hasil pengujian melalui *confusion matrix* menunjukkan bahwa algoritma *Naive Bayes* memberikan kinerja terbaik pada skenario ini. Sementara itu, *Neural Network* dan SVM juga mencapai akurasi yang cukup tinggi, sedangkan KNN belum menunjukkan hasil yang maksimal karena sensitif terhadap skala data yang masih bervariasi.

3.2. Skenario S2

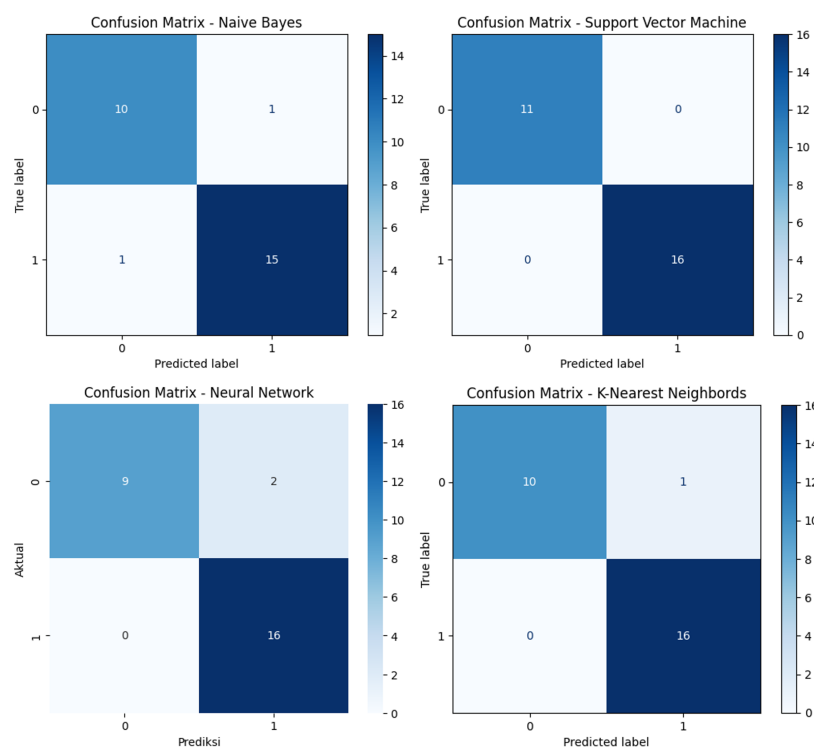
Pada skenario S2, *preprocessing* ditingkatkan dengan seleksi fitur. Variabel dengan korelasi rendah terhadap label dihapus, sehingga jumlah fitur berkurang cukup signifikan dari 122 menjadi 81. Perubahan ini membuat data lebih ringkas dan fokus pada variabel yang paling relevan untuk prediksi capaian KKM. Hasil evaluasi untuk masing-masing algoritma dapat dilihat pada Gambar 4.

Berdasarkan Gambar 4 dengan menggunakan *confusion matrix* diketahui nilai *accuracy*, *precision* dan *recall*. Hasil pengujian masing-masing model dapat dilihat pada Tabel 3.

Tabel 3. Rekap hasil testing model skenario S2

Model	Accuracy	Kelas	Precision	Recall
Naive Bayes	93%	0	91%	91%
		1	94%	94%
Support Vector Machine	78%	0	69%	82%
		1	86%	75%
Neural Network	93%	0	91%	91%
		1	94%	94%
K-Nearest Neighbor	74%	0	70%	64%
		1	76%	81%

*Confusion matrix* pada skenario ini memperlihatkan stabilitas performa model. *Neural Network* dan *Naive Bayes* masih konsisten dengan akurasi tinggi, sementara KNN dan SVM menunjukkan sedikit peningkatan meski belum optimal. *Feature selection* membantu mengurangi *noise*, namun skala data yang belum diselaraskan masih menjadi tantangan.



Gambar 5. Confusion matrix hasil testing model skenario S3

### 3.3. Skenario S3

Pada skenario S3, *preprocessing* lebih ditingkatkan dengan normalisasi dan seleksi fitur. Variabel dengan korelasi rendah terhadap label dihapus, sehingga jumlah fitur berkurang cukup signifikan dari 122 menjadi 81. Perubahan ini menghasilkan dataset yang lebih seimbang dan seragam dalam skala nilai. Hasil evaluasi untuk masing-masing algoritma dapat dilihat pada Gambar 5.

Berdasarkan Gambar 5 dengan menggunakan *confusion matrix* diketahui nilai *accuracy*, *precision* dan *recall*. Hasil pengujian masing-masing model dapat dilihat pada Tabel 4.

Tabel 4. Rekap hasil testing model skenario S3

Model	Accuracy	Kelas	Precision	Recall
Naive Bayes	93%	0	91%	91%
		1	94%	94%
Support Vector Machine	100%	0	100%	100%
		1	100%	100%
Neural Network	93%	0	100%	82%
		1	89%	100%
K-Nearest Neighbor	96%	0	100%	91%
		1	94%	100%

Visualisasi *confusion matrix* memperlihatkan hasil paling optimal pada skenario ini. Algoritma *Support Vector Machine* (SVM) mencapai akurasi sempurna sebesar 100%, sementara KNN juga menunjukkan peningkatan signifikan setelah data dinormalisasi. *Neural Network* dan *Naive Bayes* tetap memberikan hasil yang baik, meskipun tidak melampaui performa SVM.

### 3.4. Skenario S4

Skenario S4 menerapkan pendekatan pengelompokan variabel. Skor variabel dihitung dengan rata-rata tiap kelompok, sehingga dimensi data berkurang drastis menjadi 59 fitur. Dataset menjadi lebih sederhana, namun detail informasi antar-item ikut berkurang. Hasil evaluasi untuk masing-masing algoritma dapat dilihat pada Gambar 6.

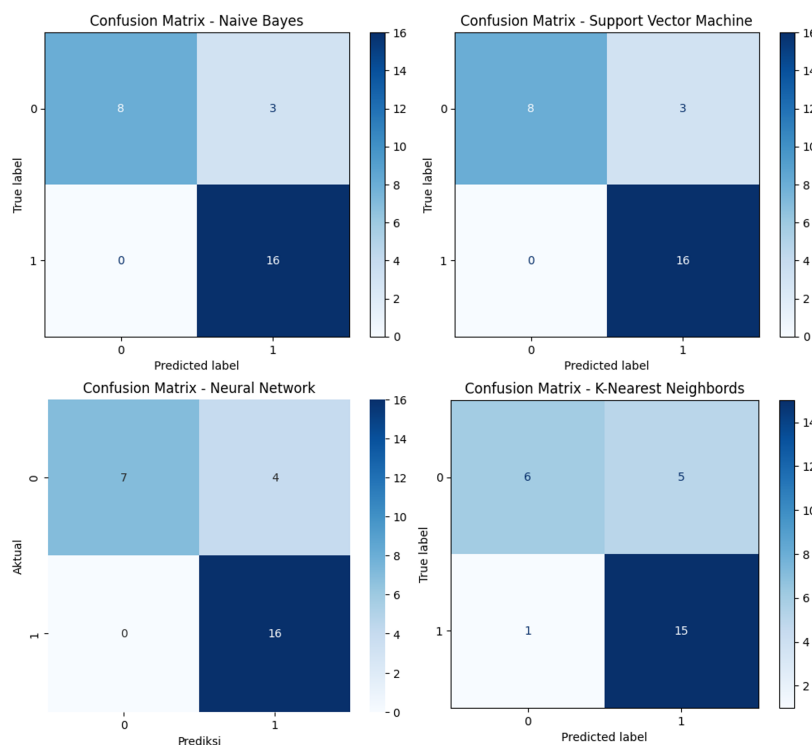
Berdasarkan Gambar 6 dengan menggunakan *confusion matrix* diketahui nilai *accuracy*, *precision* dan *recall*. Hasil pengujian masing-masing model dapat dilihat pada Tabel 5.

Tabel 5. Rekap hasil testing model skenario S4

Model	Accuracy	Kelas	Precision	Recall
Naive Bayes	89%	0	100%	73%
		1	84%	100%
Support Vector Machine	89%	0	100%	73%
		1	84%	100%
Neural Network	85%	0	100%	64%
		1	80%	100%
K-Nearest Neighbor	78%	0	86%	55%
		1	75%	94%

Hasil *confusion matrix* menunjukkan adanya penurunan kinerja pada sebagian algoritma. *Neural Network* mengalami penurunan akurasi paling signifikan akibat hilangnya variasi detail data, sementara *Naive Bayes* dan SVM masih cukup stabil meskipun tidak setinggi skenario ketiga. Hal ini menegaskan bahwa reduksi variabel melalui *variable grouping* memang membuat *dataset* lebih ringkas, tetapi berisiko menurunkan akurasi model prediksi.

Skenario S1 dan S2 memperlihatkan performa rendah pada algoritma berbasis jarak seperti SVM dengan akurasi 78% karena sensitivitasnya terhadap skala data yang belum seragam. Seba-



Gambar 6. Confusion matrix hasil testing model skenario S4

liknya, algoritma seperti *Naive Bayes* dan *Neural Network* terbukti lebih stabil dengan akurasi tinggi yang konsisten sekitar 93–96% di hampir semua skenario. Peningkatan drastis performa SVM menjadi 100% pada skenario S3 setelah penerapan *normalization* mengonfirmasi bahwa keseragaman skala merupakan faktor kunci. Namun, akurasi sempurna ini—terutama dengan *dataset* yang relatif kecil—mengindikasikan risiko *overfitting* yang sangat tinggi. Oleh karena itu, meskipun skenario S3 adalah yang terbaik, model yang lebih stabil seperti *Naive Bayes* mungkin menawarkan solusi yang lebih kokoh dan realistis dalam penerapan dunia nyata. Skenario S4 mengonfirmasi bahwa penyederhanaan data secara berlebihan justru menurunkan akurasi.

Hasil eksperimen menegaskan bahwa *preprocessing* bukan sekadar langkah teknis, melainkan strategi utama dalam membangun model prediksi. *Preprocessing* yang tepat terbukti dapat meningkatkan akurasi hingga 20–30% dibandingkan *baseline*. Bagi praktisi pendidikan, temuan ini memberikan wawasan bahwa integrasi faktor internal dan eksternal dengan pengolahan data yang sesuai mampu menghasilkan sistem prediksi capaian KKM yang lebih akurat. Dengan demikian, *preprocessing* skenario S3 dan algoritma *Naive Bayes* dapat direkomendasikan sebagai pendekatan terbaik untuk membangun model prediksi capaian KKM siswa.

Perbandingan hasil penelitian ini dengan studi Hartanti dkk. [10] menyoroti kemajuan signifikan dalam penggunaan algoritma *Naive Bayes* untuk klasifikasi KKM. Meskipun kedua penelitian menggunakan NB, studi [10] hanya mencapai akurasi 78% dengan *dataset* sebanyak 560 data namun terbatas pada 4 atribut. Sebaliknya, penelitian ini mencapai akurasi NB yang jauh lebih tinggi, yaitu 93%. Peningkatan performa yang substansial ini terjadi meskipun *dataset* penelitian ini jauh lebih kecil, yaitu 87 data latih. Keberhasilan ini menunjukkan bahwa kebaruan uta-

ma penelitian ini—yakni integrasi *dataset* berupa faktor internal dan eksternal siswa, serta penerapan *preprocessing* yang tepat—berhasil mengatasi keterbatasan penelitian sebelumnya. Temuan ini membuktikan bahwa kekayaan fitur dan strategi *preprocessing* yang baik memberikan kontribusi signifikan terhadap peningkatan akurasi *Naive Bayes*.

#### 4. Kesimpulan

Penentuan faktor eksternal siswa dalam penelitian ini dilakukan dengan memanfaatkan data hasil instrumen respon siswa serta informasi dari pihak orang tua. Data yang bersumber dari siswa mencakup aspek lingkungan sosial di sekolah, ketersediaan fasilitas sekolah, kondisi rumah tangga, faktor ekonomi, pemanfaatan teknologi, motivasi belajar, dan keterlibatan dalam kegiatan ekstrakurikuler. Sementara itu, informasi dari orang tua meliputi pendampingan proses belajar di rumah, pengawasan, dukungan emosional, kondisi ekonomi keluarga, serta ketersediaan sarana belajar. Hasil penelitian menunjukkan bahwa faktor eksternal tersebut berperan penting dalam meningkatkan akurasi model karena mampu merepresentasikan variasi lingkungan sosial maupun dukungan keluarga yang memengaruhi capaian akademik siswa.

Model prediksi KKM siswa berbasis faktor internal dan eksternal dibangun dengan menggabungkan data nilai, kehadiran, serta data dari siswa dan orang tua. Empat algoritma *supervised learning* (*Naive Bayes*, *Support Vector Machine*, *Neural Network*, dan *K-Nearest Neighbor*) diuji pada empat skenario *preprocessing*. Hasil terbaik diperoleh pada skenario S3, yaitu *normalization* dan *feature selection*, dengan algoritma *Support Vector Machine* (SVM) yang mencapai akurasi 100%. Untuk menghindari *overfitting*, direkomendasikan algoritma yang lebih stabil yaitu *Naive Bayes* dengan akurasi mencapai 93%. Hal ini menunjukkan bahwa kombi-

nasi faktor internal dan eksternal dengan *preprocessing* yang tepat mampu menghasilkan model prediksi yang sangat akurat.

**Kontribusi Penulis.** Arham Rahim: Metodologi, perangkat lunak, penulisan–persiapan draf asli. Mustakim: Penulisan–tinjauan dan penyuntingan, konseptualisasi, sumber daya. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

**Ucapan Terima Kasih.** Para penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penelitian ini dan dalam penyusunan manuskrip. Kami sangat menghargai editor dan reviewer atas masukan serta dukungannya dalam menyempurnakan karya ini.

**Pembiayaan.** Penelitian ini tidak menerima pendanaan eksternal.

**Konflik Kepentingan.** Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

**Ketersediaan Data.** Tidak tersedia.

## Referensi

- [1] D. L. Putri, Desfiona, Welnita, and Julhadi, "Analisis KKM Dalam Evaluasi Pendidikan," *J. Kaji. Ilm. Multidisipliner*, vol. 9, no. 1, pp. 282–286, 2025.
- [2] S. R. Johnson and F. K. Stage, "Academic Engagement and Student Success: Do High-Impact Practices Mean Higher Graduation Rates?," *J. Higher Educ.*, vol. 89, no. 5, pp. 753–781, Sep. 2018, doi: 10.1080/00221546.2018.1441107.
- [3] A. Selviani, A. Martiah, and A. Pertiwi, "Strategi Guru dalam Pencapaian Kriteria Ketuntasan Minimal (KKM) pada Mata Pelajaran Ekonomi," *J. Teach. Educ.*, vol. 4, no. 2, pp. 405–411, 2022, doi: 10.31004/jote.v4i2.8215.
- [4] N. Nanang and R. Rusman, "Analisis kebutuhan pelatihan standar penilaian berbasis data pemetaan mutu pendidikan (PMP) pada jenjang sekolah dasar (SD) di kota Makassar," *J. Penelit. Ilmu Pendidik.*, vol. 12, no. 1, pp. 24–37, 2019, doi: 10.21831/jpipfp.v12i1.20605.
- [5] M. Roswita and Y. Prahagia, "Cara Belajar Siswa Dengan Nilai Di Bawah Kriteria Ketuntasan Minimal," *J. Pendidik. Vokasi dan Seni*, vol. 3, no. 1, pp. 25–35, 2024, doi: 10.52060/jpvs.v3i1.2766.
- [6] S. Nahdania and S. Q. Ain, "Menggali Penyebab Rendahnya Hasil Belajar Matematika di Kelas V SD Negeri 001 Tanjung," *Cetta J. Ilmu Pendidik.*, vol. 7, no. 4, pp. 195–205, 2024, doi: 10.37329/cetta.v7i4.3775.
- [7] N. Haliza and D. F. Dwi, "Analisis Faktor Kesulitan Belajar Siswa pada Pembelajaran IPAS Materi Organ Pernapasan Manusia Kelas V SD Negeri 106815 Marindal Tahun Pembelajaran 2024–2025," *Pendas J. Ilm. Pendidik. Dasar*, vol. 10, no. 3, pp. 348–361, 2025, doi: 10.23969/jp.v10i03.28346.
- [8] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [9] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," in *Emerging Technology in Modelling and Graphics*, 2020, pp. 99–111.
- [10] D. Hartanti, Kusri, and E. L. Taufiq, "Penerapan Naïve Bayes Dalam Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 1, pp. 15–22, 2018.
- [11] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax J. Inform.*, vol. 10, no. 1, pp. 46–55, 2021.
- [12] E. P. Saputra, M. Maulidah, N. Hidayati, and A. Saryoko, "Komparasi Evaluasi Kinerja Siswa Belajar dengan Menggunakan Algoritma Machine Learning," *J. Media Inform. Budidarma*, vol. 6, no. 4, pp. 2239–2246, 2022, doi: 10.30865/mib.v6i4.4786.
- [13] Mustakim and A. Rahim, "Supervised Machine Learning for Prediction of Minimum Completeness Criteria (KKM) Scores for Elementary School Students," *J. Penelit. Pendidik. IPA*, vol. 10, no. 11, pp. 9216–9225, 2024, doi: 10.29303/jppipa.v10i11.9258.
- [14] H. Bai, H. Yu, R. B. Bantsimba N., and L. Luo, "How college experiences impact student learning outcomes: Insights from Chinese undergraduate students," *Front. Psychol.*, vol. 13, pp. 1664–1078, 2022, doi: 10.3389/fpsyg.2022.1021591.
- [15] I. Y. A. Gultom, S. A. Sibagariang, and L. F. Simatupang, "Analisis Faktor Internal dan Eksternal Yang Mempengaruhi Hasil Belajar Kognitif Pada Mata Pelajaran IPS Kelas VIII SMP Negeri 4 Pematang Siantar Tahun Ajaran 2022/2023," *J. Darma Agung*, vol. 30, no. 3, pp. 492–497, 2022, doi: 10.46930/ojsuda.v30i3.2264.
- [16] C. M. Burke, L. P. Montross, and V. G. Dianova, "Beyond the Classroom: An Analysis of Internal and External Factors Related to Students' Love of Learning and Educational Outcomes," *Data*, vol. 9, no. 6, p. 81, 2024, doi: 10.3390/data9060081.
- [17] A. Rusli, Gusmaweti, W. Hendri, and R. T. Sari, "Relationship of External Factor Caused Students' Learning Difficulties and Biology Learning Outcome In Class XI IPA MAN 3 Padang City," *Int. J. Educ. Teach. Zo.*, vol. 2, no. 1, pp. 105–112, 2023, doi: 10.57092/ijetz.v2i1.57.
- [18] H. T. C. Tho, N. Van Dinh, N. T. T. Vinh, L. T. T. Thuy, and N. T. Phuong, "Factors Influencing Student Learning Outcomes through Learning Motivation: A Case Study at a University in Vietnam," *J. Ecohumanism*, vol. 3, no. 8, pp. 5011–5021, 2024, doi: 10.62754/joe.v3i8.5143.
- [19] E. Sitepu, "Analysis of the Influence of External Factors on Student Learning Achievement in Junior High Schools," *L'Genius J. Lang. Gener. Intellect. Soc.*, vol. 13, no. 2, pp. 64–74, 2024, doi: 10.35335/genueus.v13i2.
- [20] J. Li and E. Xue, "Dynamic Interaction between Student Learning Behaviour and Learning Environment: Meta-Analysis of Student Engagement and Its Influencing Factors," *Behav. Sci. (Basel)*, vol. 13, no. 1, p. 59, 2023, doi: 10.3390/bs13010059.
- [21] D. Hartanti, Kusri, and E. L. Taufiq, "Penerapan Naïve Bayes Dalam Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa," *Jusikom Prim (J. Sist. Inf. Ilmu Komput. Prima)*, vol. 2, no. 1, pp. 15–22, 2018.
- [22] J. García-Fernández, Á. Postigo, M. Cuesta, C. González-Nuevo, Á. Menéndez-Aller, and E. García-Cueto, "To be Direct or not: Reversing Likert Response Format Items," *Span. J. Psychol.*, vol. 25, pp. 1–9, 2022, doi: 10.1017/SJP.2022.20.
- [23] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction," *Front. Bioinform.*, vol. 2, no. Jun., pp. 1–17, 2022, doi: 10.3389/fbinf.2022.927312.
- [24] Y. Kim, M. Keun, N. Fu, J. Liu, J. Wang, and J. Srebric, "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," *Sustain. Cities Soc.*, vol. 118, no. Oct. 2023, p. 105570, 2025, doi: 10.1016/j.scs.2024.105570.
- [25] J. M. Muñoz, P. Rafael, and P. Mejías, "Dimensionality reduction through clustering of variables and canonical correlation," *J. Korean Stat. Soc.*, vol. 54, no. 1, pp. 63–90, 2025, doi: 10.1007/s42952-024-00290-3.
- [26] E. Novianto, A. Hermawan, and D. Avianto, "Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree untuk prediksi Status Kelulusan Mahasiswa S1," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, 2023, doi: 10.36341/rabit.v8i2.3434.
- [27] Suryani and Mustakim, "Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine," *Bull. Informatics Data Sci.*, vol. 1, no. 2, pp. 81–88, 2022.
- [28] Ridwan, H. Lubis, and P. Kustanto, "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa," *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 286–293, 2020, doi: 10.30865/mib.v4i2.2035.
- [29] D. N. Chasanah, A. M. Siregar, and Rahmat, "Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan Algoritma," *Sci. Student J. Information, Technol. Sci.*, vol. 3, no. 1, pp. 51–58, 2022.