JPPIPA 10(11) (2024)

Jurnal Penelitian Pendidikan IPA
Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Supervised Machine Learning for Prediction of Minimum Completeness Criteria (KKM) Scores for Elementary School Students

Mustakim[1*], Arham Rahim[2]

[1] Primary Education Study Program, Teacher Training and Education Faculty, Universitas Madako Tolitoli, Tolitoli, Indonesia.
[2] Technology of Education Study Program, Teacher Training and Education Faculty, Universitas Madako Tolitoli, Tolitoli, Indonesia.

Received: June 25, 2024
Revised: September 08, 2024
Accepted: November 25, 2024
Published: November 30, 2024

Corresponding Author:
Mustakim
takim.physic@gmail.com

DOI: 10.29303/jppipa.v10i11.9258

Abstract: This study aims to predict potential declines in students' Minimum Completeness Criteria (KKM) in higher grades (4th, 5th, and 6th) by analyzing their cognitive, affective, and psychomotor scores from lower grades (1st, 2nd, and 3rd). Using a quantitative research method, various machine learning algorithms were applied, including Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Neural Networks. The dataset comprised students' scores across cognitive, affective, and psychomotor domains from the lower grades. After training and comparing the models, the Neural Network algorithm demonstrated the best performance, achieving 89% accuracy and 100% recall. These results indicate that the model can help teachers identify students at risk of struggling with KKM standards in higher grades, enabling early interventions. The study concludes that Neural Networks offer a promising tool for early detection of academic challenges in elementary education.

Keywords: Classification; KKM; Minimum completeness criteria; Supervised machine learning

Introduction

The minimum completeness criterion (KKM) is the minimum grade standard that must be achieved by students in a certain subject. KKM is used as a benchmark to determine whether students have reached the expected standards or not. In some cases, KKM is used as a condition to move up to the next level or get a graduation in a school (Mahdiansyah et al., 2017). Therefore, it is crucial to ensure that students achieve KKM in each subject. Minimum completeness is the minimum level of achievement that must be achieved by students to meet the competency standards that have been set by the government. The achievement of minimum completeness by students is very important to ensure equal access to education for all students, as well as to ensure that the quality of education provided by schools meets the set standards. The achievement of minimum completeness can also be an indicator of the success of an educational program and can help teachers and principals in improving teaching methods and curriculum to suit the needs of students (Selviani et al., 2022). This will have an impact on improving the quality of education provided and increasing students' ability to compete in the world of work.

Standards for assessing learning outcomes and learning processes have been established, especially in Indonesia. The determination of the assessment standards includes three main aspects, namely affective or attitude, cognitive or knowledge, and psychomotor or skills (Mahdiansyah et al., 2017). Affective refers to students' attitudes, values, and emotional responses, focusing on their social and behavioral development. Cognitive pertains to the intellectual abilities of students, such as their capacity to understand, analyze, and solve problems. Psychomotor involves the physical

How to Cite:
Mustakim, & Rahim, A. (2024). Supervised Machine Learning for Prediction of Minimum Completeness Criteria (KKM) Scores for Elementary School Students. *Jurnal Penelitian Pendidikan IPA*, *10*(11), 9216–9225. https://doi.org/10.29303/jppipa.v10i11.9258

skills and coordination necessary for tasks that require movement and manual dexterity. These three aspects are considered equally important in the assessment of learning outcomes and processes.

Learning in elementary school is divided into two parts, namely low grade learning, namely grades 1, 2 and 3 and high grade learning, namely grades 4, 5 and 6 (Zulvira et al., 2021) and research conducted by Dwi & Audina (2021) states that the low numeracy ability of elementary school students at the upper level is also caused by the mastery of basic numeracy skills such as addition, subtraction, multiplication, and division that is still lacking in the lower class. Based on this relationship, this study focuses on the prediction of the KKM scores of elementary school students at the upper level based on the achievements of students in the lower grades.

The urgency of this research stems from the critical role that KKM plays in determining students' academic progression and overall educational success. As KKM is often a deciding factor for students to advance to the next grade level or graduate, it is essential to ensure that students meet these benchmarks early on. However, without early detection of potential academic challenges, students who struggle in lower grades may continue to face difficulties in higher grades, ultimately affecting their ability to meet KKM standards. This could lead to negative consequences, such as delayed progression or failure to complete school on time. By developing a predictive model that can identify students at risk of falling below KKM standards, educators can intervene early and provide necessary support to ensure students' success. Such a model not only helps prevent academic failure but also contributes to improving the overall quality of education by enabling targeted interventions that address students' specific needs. Therefore, this research is crucial for fostering equitable education and ensuring that all students have the opportunity to succeed academically.

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that allows computers to learn and improve their performance in a specific task through experience and data provided (Charbuty & Abdulazeez, 2021). One of the intelligence from ML models or algorithms is obtained by applying the Supervised Machine Learning method (Tungadi et al., 2018). Supervised Machine Learning is one of the methods to train a model using past data that has labels so that the model makes it possible to predict future outcomes (Sen et al., 2020).

There have been several studies on the application of machine learning in predicting students' KKM scores. Hartanti et al. (2018) used the Naïve Bayes algorithm to classify KKM achievement and found that Naïve Bayes achieved an accuracy of 78%. Similarly, Purwaningsih & Nurelasari (2021) conducted research focusing on the use of the K-Nearest Neighbor (KNN) algorithm to classify student graduation levels, reporting an impressive accuracy of 96.49%. These studies, however, used different datasets, making it difficult to conclusively determine which algorithm is most effective for predicting KKM values.

Saputra et al. (2022) expanded on this by comparing several classification algorithms, including Naïve Bayes, Support Vector Machines (SVM), Neural Networks (NN), and K-Nearest Neighbor (KNN). Using a dataset of 1,000 records from a dataset provider website, they concluded that the Neural Network algorithm performed best, with an accuracy of 99.50%. Despite these findings, the reliance on second-party data introduces the potential for bias in the results.

Given the varying results of these studies and the potential biases due to different data sources, there is a need for further research using primary data directly from schools. This study compares classification algorithms, including Naïve Bayes, K-Nearest Neighbor, Support Vector Machines, and Neural Networks, using actual KKM data to avoid biases and determine which algorithm offers the most accurate predictions. The models will be evaluated using a confusion matrix to assess their effectiveness in classifying students' KKM scores.

To assist teachers in identifying students at risk of a decline in KKM scores, a system is required to predict such issues early. Building a predictive model using machine learning algorithms could allow for human-level accuracy in analyzing student performance data. This intelligence is achieved through supervised machine learning models trained on historical data, including students' affective, cognitive, and psychomotor scores. Therefore, this study aims to determine which classification algorithm provides the highest accuracy for predicting students' KKM scores.

Method

This research adopts a quantitative approach with an experimental design to compare machine learning algorithms in predicting students' Minimum Completeness Criteria (KKM) scores in higher grades based on their performance in lower grades. First, data were collected from school records, focusing on the cognitive, affective, and psychomotor scores of students in grades 1 through 3. Second, the collected data underwent preprocessing, which involved cleaning, normalizing, and formatting to ensure their suitability for machine learning analysis. Third, four machine learning algorithms—Naïve Bayes, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Neural Networks—were trained using the preprocessed data. These models were designed to classify and predict KKM scores for students in grades 4, 5, and 6. Fourth, the models were evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine their predictive effectiveness. Finally, the models were compared, and the algorithm with the best overall performance was selected as the most suitable for predicting student achievement. This final model can assist educators in identifying students at risk of not meeting KKM standards, allowing for early intervention and tailored support.

The approach taken in this study is quantitative with the type of experimental research, where this study conducts test scenarios by comparing the algorithms of Naïve Bayes, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Neural Networks in supervised machine learning to measure, analyze, and classify data so that it can be used as a system that can predict student achievement. To clarify how the researcher solves the problem, namely how to build a model that can classify student KKM scores and how to determine the classification algorithm with the best accuracy in predicting student KKM scores, it can be seen in the research flow in Figure 1.

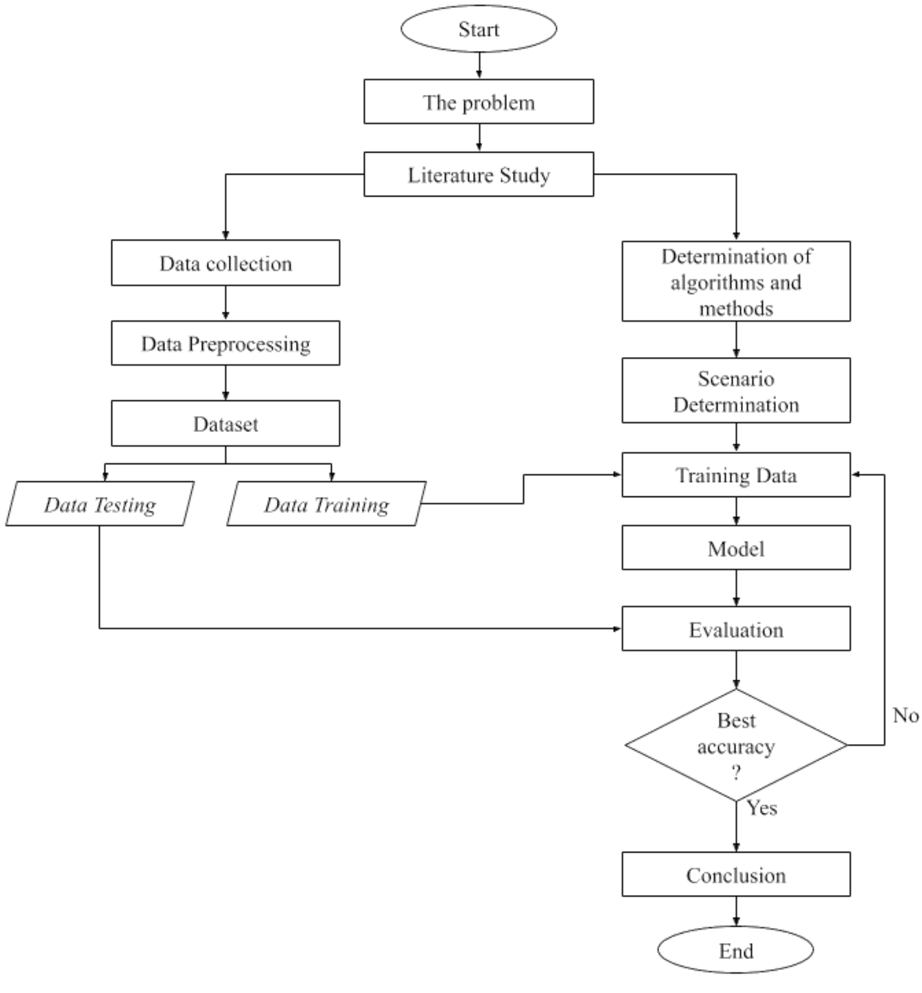

Figure 1. Research flow

Data Collection Methods

The data used in this study was collected through several methods to ensure completeness and accuracy, the first of which is documentation. The data documentation in this study focuses on collecting data on cognitive, affective, psychomotor and KKM scores of students. Next is an interview to explore more information related to student KKM data. Technical information related to the provision of student KKM scores by teachers is needed to support the description of this research.

Data Processing Methods

To train a machine learning model, input data is needed so that data is the life of a machine learning (Murray et al., 2021). Likewise, in this study, to find an algorithm with the best accuracy in classifying students' KKM scores in high grades, data in the form of affective, cognitive, and psychomotor scores of students in the two grades below is needed. The sample data used in this study can be seen in Figure 2.

	KELAS_1	MP_1_1_1	MP_1_1_2	MP_1_1_3	MP_1_1_4	MP_1_1_5	MP_1_1_6	MP_1_1_7	KET_1_1_1	KET_1_1_2	...
Code cell output actions		72	75	70	66	76	80	75	0	0	...
1	1	85	94	90	85	81	80	75	0	0	...
2	1	81	86	68	54	77	80	75	0	0	...
3	1	81	78	75	84	72	80	75	5	0	...
4	1	78	92	85	84	85	80	75	2	0	...
...	...	...	...	...	...	...	...	...	...	...	...
85	3	84	96	93	84	84	85	82	0	0	...
86	3	77	81	85	81	78	75	82	0	0	...
87	3	75	79	80	74	84	80	83	0	0	...
88	3	83	99	99	95	99	88	98	0	0	...
89	3	91	99	99	94	92	88	92	0	0	...

90 rows × 43 columns

MP_2_2_2	MP_2_2_3	MP_2_2_4	MP_2_2_5	MP_2_2_6	MP_2_2_7	KET_2_2_1	KET_2_2_2	KET_2_2_3	LABEL
79	75	77	76	80	85	0	0	1	1
95	90	90	84	82	90	0	0	1	1
72	75	72	60	83	80	3	0	2	0
87	80	75	63	86	80	2	0	4	1
86	90	88	84	85	90	1	0	2	1
...	...	...	...	...	...	...	...	...	...
79	80	75	80	80	80	5	0	2	0
80	89	86	83	84	81	0	4	0	0
78	67	74	80	80	80	5	0	1	0
80	90	90	87	90	90	0	0	0	1
80	85	83	84	90	85	8	0	0	1

Figure 2. Dataset

Data preprocessing always has an important effect on the performance of supervised machine learning algorithms (Alexandropoulos et al., 2019), so that to prepare data can be an input to machine learning algorithms, treatment such as instance selection to select a subset of relevant or representative data instances to be used in model training. The attributes used in this study are class, grades of Religious Education, PPKn, Mathematics, Science, Social Studies, SBDP, PJOK, Mulok as well as the number of permits, sick and without information; and missing value imputation to replace the missing value in the dataset with the estimated or calculated value.

Furthermore, the data labeling process for historical data of affective, cognitive and psychomotor values in each class is labeled "0" which does not reach the KKM score and "1" which reaches the KKM value when it has entered the upper class. Based on Figure 2, this researcher used 90 student data with 43 attributes consisting of 39 data with the label "0" and 51 data with the label "1".

Data grouping is also the most important part of training a supervised machine learning. This process is carried out by grouping the data that already has labels into two parts, namely training data (80%) which is used to train the model and testing data (20%) which is used to evaluate the model (Alexandropoulos et al., 2019).

Data Analysis Methods

The method used in this study is quantitative analysis using several algorithms, namely Naïve Bayes, Support Vector Machine, Neural Network and K-Nearest Neighbor which are made in the python programming language. Experiments on some of these algorithms were carried out to obtain the algorithm with the best accuracy for the problem of predicting students' KKM scores at the upper level based on student data at the lower level. Model training or model training is a process in machine learning where a model learns from existing data to make predictions or make decisions. During this process, machine learning algorithms use training data to find patterns or relationships that can be used to predict the output of new inputs. This process involves adjusting the model's parameters to minimize errors between the model's predictions and the actual results.

Naïve Bayes (NB) is an technique that uses simple probabilities based on Bayesian inference theory with strong assumptions of independence. NB is a classification statistic that can be used to estimate the likelihood that each particular class will have members. Nave Bayes is often used to categorize dataset data from intensive training. Bayesian theorem as follows:

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

Information:
P(A|C) is the probability of hypothesis A being given evidence C (posterior).
P(C|A) is the probability of proof C given that hypothesis A is true (likelihood).
P(A) is the initial probability of hypothesis A (prior).
P(C) is the probability of evidence C.

Support Vectore Machine (SVM) is a supervised learning algorithm that has good capabilities in analyzing data for classification and regression. SVM works to find the best hyperlane for splitting data into two classes, and maximizing the margin between the two classes (Suryani & Mustakim, 2022). Linear model used on SVM to generate optimal hyper lane using equations:

$$f(w,x) = w.x + b \quad (2)$$

Information:
w is the weight vector.
x is the feature vector.
b is the bias or intercept.

SVM learns the parameters by solving optimization problems using the following equation:

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

Information:
$w^T w$ is Manhattan Norm (also known as L1 norm).
C is Penalty parameter (can be an arbitrary value or a value selected using parameter-hyper tuning).
y_i' is Actual label.
$w^T x_i + b$ is Predictor function.

A Neural Network (NN) is an algorithmic model inspired by how neurons in the human brain work. Each neuron in the human brain is interconnected and information flows from each of the neurons. NN itself has three layers, namely input, hidden and output layers. A hidden layer can consist of one or two hidden layers and theoretically there is no basic research on how many hidden layers are adequate for the network. In a neural network there is a neuron, which allows it to map different inputs to an output, this is the most fundamental element in any NN (Ridwan et al., 2020).

K-Nearest Neighbors (KNN) is a machine learning algorithm used for classification and regression tasks. KNN is a non-parametric algorithm that does not make any assumptions about the distribution of data. It is an instance-based algorithm which means it stores all the training instances and classifies new instances based on similarities with the stored training instances. Here are the steps to calculate the K-Nearest Neighbor algorithm method to classify new data that is not yet known (Chasanah et al., 2022). Determine the number of neighbors, the number of k taken is generally odd because it avoids the same amount of distance in the classification process. From the data that has been given, then proceed to calculate the euclidean distance of each object. The result of the distance calculation then the objects are sorted from the smallest euclid distance. The results of the classification are collected according to the categories on the label. After the calculation is carried out, the prediction results from the category determined by the KNN method will appear. Here is the equation for the KNN Algorithm.

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

Information:
d_{ij} is Distance between object i and j.
x_{ik} is Object value i in the k variable.
x_{jk} is The value of object j on k variable.
p is Number of variables.

To obtain the model with the best accuracy, the experimental scenario will then be evaluated using the Confusion matrix method to assess the accuracy in classifying students' KKM scores so that it can be known which algorithm is most suitable for the problem.

Result and Discussion

Result

To find out whether a student will achieve a KKM score or not, this study compares the algorithms of Naïve Bayes, Support Vector Machine, Neural Network and K-Nearest Neighbor. The evaluation results for each algorithm can be seen in Figure 3.

In comparison, Naïve Bayes misclassified 3 students who met KKM (false negatives) and incorrectly predicted 2 students who did not meet KKM as passing (false positives). This demonstrates that Naïve Bayes has more difficulty in correctly classifying students who meet the criteria. SVM performed similarly, with 5 true positives, 3 false negatives, and 0 false positives, making it highly precise for class 0 but limited in recall. KNN had a more balanced performance, with 6 true positives, 2 false negatives, and 1 false positive, indicating a more reliable performance overall compared to Naïve Bayes and SVM. Based on Figure 3 using the conusion matrix, the values of accuracy, precision and recall are known. The test results of each model can be seen in Table 1.

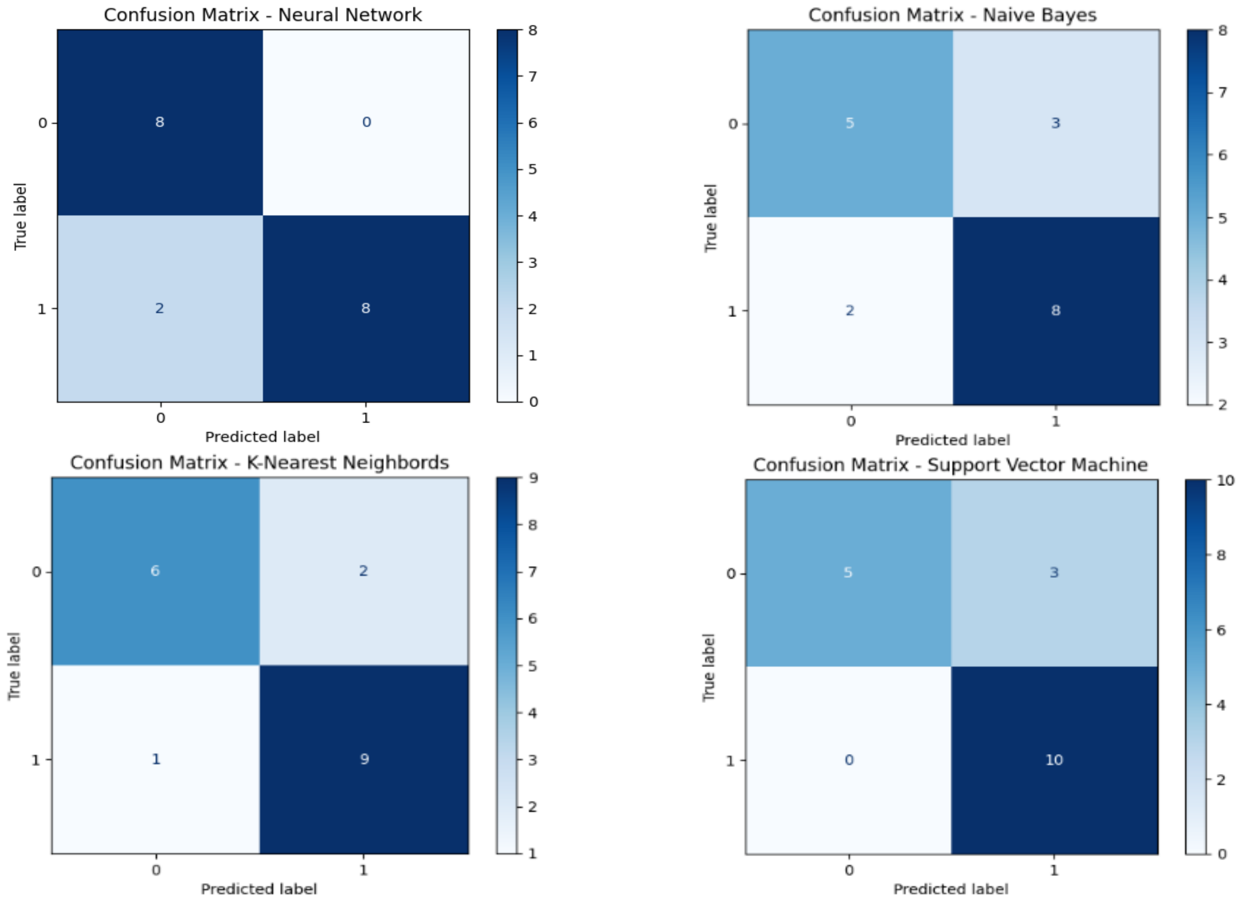

Figure 3. Confusion matrix test model results

Table 1. Recap of the results of the testing model

Model	Accuracy (%)	Class	Precision (%)	Recall (%)
Naïve bayes	72	0	71	62
		1	73	80
Support vector machine	83	0	100	62
		1	77	100
Neural network	89	0	80	100
		1	100	80
K-nearest neighbor	83	0	86	75
		1	82	90

As shown in Table 1, the Neural Network model outperformed the other models with an accuracy of 89% and a perfect recall of 100% for class 0, indicating that it successfully identified all students who met KKM. In contrast, Naïve Bayes had the lowest accuracy at 72%, with a recall of only 62% for class 0, suggesting it struggled to correctly predict students who met KKM. SVM and KNN both achieved an accuracy of 83%, but KNN had a higher recall of 75% for class 0, making it more effective in identifying students who met KKM compared to SVM, which had a recall of 62%. The following results of the comparison of accuracy and recall in class 0 can be seen in Figure 4.

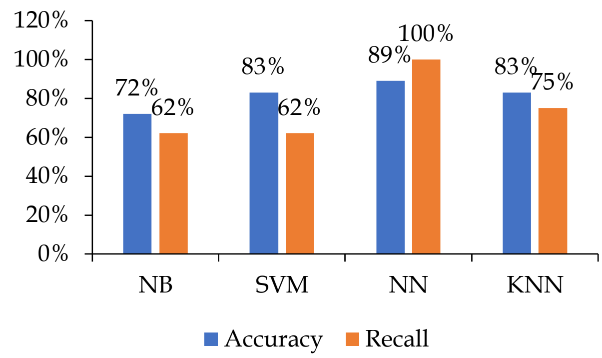

Figure 4. Comparison of accuracy and recall between models

Based on Figure 4, This chart visually highlights the trade-offs between precision and recall across the models, showing that while SVM had high precision, it was less effective in identifying all students who met KKM compared to KNN and Neural Network. In conclusion, the Neural Network model was the best-performing algorithm, with a balanced combination of high accuracy (89%) and perfect recall (100%) for students who meet KKM. This suggests that it is the most effective tool for early detection of students who are likely to meet KKM, allowing educators to focus their attention on students who may need further support. KNN, with its balanced recall and accuracy, also performed well, while SVM, though precise, showed limitations in recall. Naïve Bayes consistently underperformed, making it less suitable for predicting student success in this context.

Discussion

The application of artificial intelligence technology, especially machine learning, offers solutions that can help schools and teachers to identify students who need special attention from an early age (Ahadi et al., 2015; Riza et al., 2024; Sinaga et al., 2024; Sudais et al., 2022; Suharti, 2024; Sulistiyo et al., 2020). This study uses a supervised machine learning-based prediction model trained using data on affective, cognitive, and psychomotor scores of students at low levels (grades 1, 2, and 3), to predict KKM achievement at high levels (grades 4, 5, and 6).

The results of this study Neural Network model demonstrated the highest overall performance, achieving an accuracy of 89% and a perfect recall of 100% for class 0 (students who meet KKM). This suggests that Neural Networks are particularly well-suited for this type of predictive task, as they can accurately detect all students who are on track to meet KKM while maintaining strong overall prediction accuracy (Almeida & Azkune, 2018; Gawlikowski et al., 2023; Juniawan et al., 2024). The model's high recall is critical in educational contexts, as identifying students who meet KKM helps ensure that resources and interventions are directed towards those who are truly at risk of underperforming. The success of this model could be attributed to its ability to capture complex patterns in the dataset, which includes cognitive, affective, and psychomotor scores (Abiodun et al., 2019; Alwosheel et al., 2018).

While SVM and KNN both achieved similar accuracy levels (83%), the trade-offs between precision and recall are evident. SVM's perfect precision for class 0, paired with lower recall, suggests that while it can identify students who meet KKM with high confidence, it may fail to recognize some students who actually meet the criteria. This might be problematic in educational interventions, where missing students who need support could hinder their academic success. On the other hand, KNN strikes a more balanced performance with a recall of 75%, showing it is better at identifying students who meet KKM compared to SVM, though not as good as Neural Networks. Naïve Bayes, with an accuracy of 72% and a recall of 62%, exhibited the weakest performance overall. This can likely be attributed to the model's assumption of independence between features, which may not hold true in a dataset that includes complex interactions between cognitive, affective, and psychomotor aspects (Blanquero et al., 2021; Chen et al., 2020; Zhang et al., 2018). In educational contexts, such a model would not be ideal, as its lower recall indicates a higher likelihood of misclassifying students who meet KKM, potentially leading to inaccurate intervention strategies.

The results from this study have clear implications for educational practice, particularly in early detection of students' academic performance. Positive relationships between teachers and students in the learning process can narrow the achievement gap for students who have difficulty in the established achievement process (Kincade et al., 2020; Mayanda et al., 2024). The superior performance of the Neural Network model suggests that it could be used to support teachers and administrators in identifying students at risk of not meeting the KKM early on, allowing for timely interventions (Azizah et al., 2024; Kroesch et al., 2022; Nurita et al., 2024; Zilz & Pang, 2021). The Minimum Completeness Criteria (KKM) having a far-reaching impact on student academic and career future (Cole et al., 2016; Johnson & Stage, 2018; Karalekas et al., 2020). By incorporating such predictive models into school systems like this study, educators could prioritize resources more effectively, tailoring support to students who need it most.

However, while machine learning models like Neural Networks provide strong predictive capabilities, educators must remain aware of the limitations inherent in the data. Factors beyond cognitive, affective, and psychomotor scores, such as socioeconomic status, family environment, and other external influences, may also impact student performance and should be considered when designing intervention strategies (Aulia et al., 2024; Ayuningtyas & Kuswandi, 2024; Burchinal et al., 2020; Engelhardt et al., 2019; Marks, 2016; Rafidah et al., 2024; Xing, 2023). Additionally, continuous evaluation and improvement of these models are necessary to ensure they remain effective as educational practices and student profiles evolve. Future research could expand on these findings by incorporating additional features into the models, such as social or environmental factors, to further enhance their predictive accuracy. Furthermore, different machine learning techniques, such as ensemble methods, could be explored to see if they can outperform Neural Networks in this context. Investigating how models like Neural Networks can be implemented in real-world educational settings would also be valuable to determine their practical utility and impact on student outcomes.

Conclusion

Based on the findings of this study, students' cognitive, affective, and psychomotor scores in the lower grades can be effectively used to build a predictive model for determining the likelihood of students meeting the Minimum Completeness Criteria (KKM) in higher grades. Among the machine learning algorithms tested, including Naïve Bayes, Support Vector Machine, Neural Network, and K-Nearest Neighbor, the Neural Network algorithm emerged as the most effective, with an accuracy of 89% and a recall of 100%. These results demonstrate the potential of machine learning, particularly Neural Networks, in identifying students at risk of falling below KKM standards, thus enabling timely intervention to enhance student outcomes. This approach highlights the significant role that machine learning can play in improving the quality of education in Indonesia by supporting educators in addressing students' academic challenges early on.

Acknowledgments
The authors express their gratitude to the Headmaster of SDN Lantatapan, Tolitoli Regency, and all relevant parties for their valuable support and permission to facilitate data collection for this research.

Author Contributions
Conceptualization, writing—original draft preparation, writing—review and editing, and data visualization, M. and A.R.; methodology and formal analysis, A.R.; data source, M. All authors have read and agreed to the published version of the manuscript.

Funding
The authors gratefully acknowledge the financial support provided by the Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia, which made this research possible.

Conflicts of Interest
The author declares no conflict of interest.

References

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. A., Gana, U., & Kiru, M. U. (2019). Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. *IEEE Access*, *7*, 158820–158846. https://doi.org/10.1109/ACCESS.2019.2945545

Ahadi, A., Lister, R., Haapala, H., & Vihavainen, A. (2015). Exploring Machine Learning Methods to Automatically Identify Students in Need of Assistance. *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*, 121–130. https://doi.org/10.1145/2787622.2787717

Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data Preprocessing in Predictive Data Mining. *The Knowledge Engineering Review*, *34*(e1), 1–33. https://doi.org/10.1017/S026988891800036X

Almeida, A., & Azkune, G. (2018). Predicting Human Behaviour with Recurrent Neural Networks. *Applied Sciences*, *8*(2), 1–13. https://doi.org/10.3390/app8020305

Alwosheel, A., Cranenburgh, S. V., & Chorus, C. G. (2018). Is Your Dataset Big Enough? Sample Size Requirements When Using Artificial Neural Networks for Discrete Choice Analysis. *Journal of Choice Modelling*, *28*, 167–182. https://doi.org/10.1016/j.jocm.2018.07.002

Aulia, I., Sujana, A., & Sunaengsih, C. (2024). Application of the Climate Kids Interactive Web to Build Understanding of Concepts and Environmental Awareness of Class VI Students on Global Warming Material. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 246-253. https://doi.org/10.29303/jppipa.v10i1.6386

Ayuningtyas, D., & Kuswandi, P. C. (2024). PteridophytaE-Encyclopedia of Kelud Mountain to Study of the Diversity of Fern Species in the Environment on the RICOSRE Model to Improve Student's Information Literacy. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 254-260. https://doi.org/10.29303/jppipa.v10i1.5917

Azizah, N., Istiyono, E., & Wilujeng, I. (2024). Development of Student Cognitive Learning Outcomes Tests Based on Differentiated Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 194-200. https://doi.org/10.29303/jppipa.v10i1.5080

Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable Selection for Naïve Bayes Classification. *Computers & Operations Research*, *135*, 105456. https://doi.org/10.1016/j.cor.2021.105456

Burchinal, M., Foster, T. J., Bezdek, K. G., Bratsch-Hines, M., Blair, C., & Vernon-Feagans, L. (2020). School-Entry Skills Predicting School-Age Academic and Social–Emotional Trajectories. *Early Childhood Research Quarterly*, *51*, 67–80. https://doi.org/10.1016/j.ecresq.2019.08.004

Charbuty, B., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Chasanah, D. N., Siregar, A. M., & Rahmat, R. (2022). Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan Algoritma K-Nearest Neighbor. *Scientific Student Journal for Information, Technology and Science*, *3*(1), 51–58. Retrieved from https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/421

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A Novel Selective Naïve Bayes Algorithm. *Knowledge-Based Systems*, *192*, 105361. https://doi.org/10.1016/j.knosys.2019.105361

Cole, S., Paulson, A., & Shastry, G. K. (2016). High School Curriculum and Financial Outcomes: The Impact of Mandated Personal Finance and Mathematics Courses. *Journal of Human Resources*, *51*(3), 656–698. https://doi.org/10.3368/jhr.51.3.0113-5410R1

Dwi, D. F., & Audina, R. (2021). Analisis Faktor Penyebab Kesulitan Belajar Matematika Kelas IV Sekolah Dasar Negeri. *Cybernetics: Journal Educational Research and Social Studies*, *2*(3), 94–106. https://doi.org/10.51178/cjerss.v2i3.256

Engelhardt, L. E., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2019). Accounting for the Shared Environment in Cognitive Abilities and Academic Achievement with Measured Socioecological Contexts. *Developmental Science*, *22*(1), e12699. https://doi.org/10.1111/desc.12699

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., & Zhu, X. X. (2023). A Survey of Uncertainty in Deep Neural Networks. *Artificial Intelligence Review*, *56*(1), 1513–1589. https://doi.org/10.1007/s10462-023-10562-9

Juniawan, E. R., Sumarni, W., & Prasetya, A. T. (2024). Development of Ethno-STEM-Loaded Digital Science Teaching Materials the Process of Making Traditional Sidoarjo Snacks Material of Force and Object Motion to Train Science Literacy in Elementary School Students. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 325-337. https://doi.org/10.29303/jppipa.v10i1.5948

Hartanti, D., Kusrini, K., & Taufiq, E. L. (2018). Penerapan Naïve Bayes Dalams Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa. *Jusikom Prima*, *2*(1), 15–22. Retrieved from https://jurnal.unprimdn.ac.id/index.php/JUSIKOM/article/view/147

Johnson, S. R., & Stage, F. K. (2018). Academic Engagement and Student Success: Do High-Impact Practices Mean Higher Graduation Rates?. *The Journal of Higher Education*, *89*(5), 753–781. https://doi.org/10.1080/00221546.2018.1441107

Karalekas, G., Vologiannidis, S., & Kalomiros, J. (2020). Europa: A Case Study for Teaching Sensors, Data Acquisition and Robotics Via a ROS-Based Educational Robot. *Sensors (Switzerland)*, *20*(9). https://doi.org/10.3390/s20092469

Kincade, L., Cook, C. R., & Goerdt, A. (2020). Meta-Analysis and Common Practice Elements of Universal Approaches to Improving Student-Teacher Relationships. *Review of Educational Research*, *90*(5),003465432094683. https://doi.org/10.3102/0034654320946836

Kroesch, A. M., Jozwik, S., Douglas, K. H., Chung, Y.-C., Uphold, N. M., & Baker, E. (2022). Using Technology to Support Academic Learning. *The Journal of Special Education*, *56*(3), 158–167. https://doi.org/10.1177/00224669211070563

Mahdiansyah, M., Sembiring, M. S., Supriyadi, T., Ulumudin, I., & Fujianita, S. (2017). Sistem Penilaian Hasil Belajar dan Kemampuan Guru dalam Melaksanakan Penulisan Berdasarkan Kurikulum 2013 Handal. In *Policy Brief Pusat Penelitian Kebijakan Pendidikan dan Kebudayaan*.

Marks, G. N. (2016). The Relative Effects of Socio-Economic, Demographic, Non-Cognitive and Cognitive Influences on Student Achievement in Australia. *Learning and Individual Differences*, *49*, 1–10. https://doi.org/10.1016/j.lindif.2016.05.012

Mayanda, I., Yennita, Y., & Islami, N. (2024). The Effect of Wordwall-Assisted Brain-Based Learning to Cognitive Learning Outcomes on Optical Equipment Material. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 261-269. https://doi.org/10.29303/jppipa.v10i1.5518

Murray, D. G., Šimša, J., Klimovic, A., & Indyk, I. (2021). Tf.Data: A Machine Learning Data Processing Framework. *Proceedings of the VLDB Endowment*, *14*(12), 2945–2958. https://doi.org/10.14778/3476311.3476374

Nurita, T., Yuliati, L., Mahdiannur, M. A., Ilhami, F. B., Fauziah, A. N. M., Hendratmoko, A. F., & Puspitarini, S. (2024). Increasing Pre-Service Science Teacher Creativity Through STEM Problem-Solving. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 72-79. https://doi.org/10.29303/jppipa.v10i1.6335

Purwaningsih, E., & Nurelasari, E. (2021). Penerapan K-Nearest Neighbor untuk Klasifikasi Tingkat Kelulusan pada Siswa. *Syntax: Jurnal Informatika*, *10*(1), 46–56. https://doi.org/10.35706/syji.v10i01.5173

Rafidah, H. N., Rachmadiarti, F., & Prastiwi, M. S. (2024). Stepping Together with Nature of Malang Raya: The Development Environmental Changes E-Book Based on Problem Based Learning (PBL). *Jurnal Penelitian Pendidikan IPA*, *10*(7), 3556-3568. https://doi.org/10.29303/jppipa.v10i7.7377

Ridwan, R., Lubis, H., & Kustanto, P. (2020). Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Media Informatika Budidarma*, *4*(2), 286. https://doi.org/10.30865/mib.v4i2.2035

Riza, S., Rizki, D., & Ihsan, M. A. N. (2024). The Effect of The Use of Contextual Teaching and Learning (CTL) Learning Model on The Cognitive Value of Students of Elementary School. *Jurnal Penelitian Pendidikan IPA*, *10*(5), 2702-2710. https://doi.org/10.29303/jppipa.v10i5.6988

Saputra, E. P., Maulidah, M., Hidayati, N., & Saryoko, A. (2022). Komparasi Evaluasi Kinerja Siswa Belajar dengan Mengggunakan Algoritma Machine Learning. *Jurnal Media Informatika Budidarma*, *6*(4), 2239–2246. https://doi.org/10.30865/mib.v6i4.4786

Selviani, A., Martiah, A., & Pertiwi, A. (2022). Strategi Guru dalam Pencapaian Kriteria Ketuntasan Minimal (KKM) pada Mata Pelajaran Ekonomi. *Journal on Teacher Education*, *4*(2), 405–411. https://doi.org/10.31004/jote.v4i2.8215

Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised Classification Algorithms in Machine Learning: A Survey and Review. In J. K. Mandal & D. Bhattacharya (Eds.), *Emerging Technology in Modelling and Graphics* (pp. 99–111). Springer Singapore. https://doi.org/10.1007/978-981-13-7403-6_11

Sinaga, Y. R. A., Boleng, D. T., Maasawet, E. T., Akhmad, A., & Rambitan, V. M. M. (2024). Development of Neuroscience-Based Biology Learning Media to Increase Learning Motivation and Cognitive Learning Outcomes of Tenggarong High School Students. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 2916–2926. https://doi.org/10.29303/jppipa.v10i6.7314

Sudais, M., Safwan, M., & Ahmed, S. (2022). Students' Academic Performance Prediction Model Using Machine Learning. *Research Square*, 1–20. https://doi.org/10.21203/rs.3.rs-1296035/v1

Sulistiyo, B., Surarso, B., & Syafei, W. A. (2020). Improving the Accuracy of Student Problem Identification Using Rule-Based Machine Learning. *E3S Web of Conferences*, *202*, 0–7. https://doi.org/10.1051/e3sconf/202020215012

Suharti, D. I. (2024). Validity of Creative Interactive-Web and Seamless Learning Media and Learning Models to Improve Students' Creative Thinking Skills and Cognitive Learning Outcomes in High School Biology Subjects. *Jurnal Penelitian Pendidikan IPA*, *10*(7), 3770-3779. https://doi.org/10.29303/jppipa.v10i7.8279

Suryani, S., & Mustakim, M. (2022). Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine. *Bulletin of Informatics and Data Science*, *1*(2), 81–88. https://doi.org/10.61944/bids.v1i2.36

Tungadi, E., Thalib, I., Nur, M., & Utomo, Y. (2018). Machine Learning Penentuan Penerima Beasiswa Peningkatan Prestasi Akademik (PPA) Menggunakan Metode Jaringan Saraf Tiruan (JST). *Seminar Nasional Teknik Elektro dan Informatika 2018*. September, 391–396.

Xing, Z. (2023). Explore How Family Factors Affect Students' Academic Performance-Based on Literature Analysis Zhuoran. *Journal of Education, Humanities and Social Sciences*, *10*, 91–98. https://doi.org/10.54097/ehss.v10i.6897

Zhang, N., Wu, L., Yang, J., & Guan, Y. (2018). Naive Bayes Bearing Fault Diagnosis Based on Enhanced Independence of Data. *Sensors*, *18*(2), 1–17. https://doi.org/10.3390/s18020463

Zilz, W., & Pang, Y. (2021). Application of Assistive Technology in Inclusive Classrooms. *Disability and Rehabilitation: Assistive Technology*, *16*(7), 684–686. https://doi.org/10.1080/17483107.2019.1695963

Zulvira, R., Neviyarni, N., & Irdamurni, I. (2021). Karakteristik Siswa Kelas Rendah Sekolah Dasar. *Jurnal Pendidikan Tambusai*, *5*(1), 1846–1851. https://doi.org/10.59188/jurnalsostech.v3i6.810