Journal of Digital Technology and Computer Science
E-ISSN: 3030-8127; P-ISSN: 3031-0318
https://journal.lontaradigitech.com/DIGITECH

KNN Vs Naive Bayes: an Innovative Comparison in Predictive AI Learning With Association Data Support

Devi Miftahul Jannah*[1], Aprilianti Nirmala S [2]
[1,2,3,4,5] *Universitas Negeri Makassar, Indonesia*

Corresponding e-mail: devimiftahul734@gmail.com

ARTICLE INFO

Keywords:

Association Rule Mining; KNN; Naive Bayes; Predictive Learning; Artificial Intelligence

Article History

Received: August 15, 2025
Revised : September 25, 2025
Accepted : October 05, 2025

ABSTRACT

This study analyzes how Naive Bayes and K-Nearest Neighbor (KNN) predict learning outcomes based on artificial intelligence (AI). The main focus of this study is the difficulty of algorithms in handling complex learning data and the contribution of Association Rule Mining (ARM) attribute features in improving prediction accuracy. The methods applied include two classification algorithms (KNN and Naive Bayes) in an exploratory-comparative quantitative research design, as well as the application of ARM to uncover hidden patterns among variables using the apriori algorithm. Data for 368 students with prior experience in artificial intelligence technology was collected through an online survey. Although KNN outperforms in recall, the study results show that Naive Bayes has higher precision. By detecting hidden correlation patterns that cannot be identified by conventional classification methods, ARM improves classification results. The discussion emphasizes that the selection of the best algorithm depends on the application's objectives, namely whether the priority is on classification accuracy or the range of relevant results. Based on these findings, a hybrid technique combining KNN, Naive Bayes, and ARM is highly recommended for creating a more efficient and accurate prediction system to support AI-based education.

To cite this article : Author. (20xx). Title. Information Technology Education Journal, X(X), XX-XX. Doi. xxxx

INTRODUCTION

By facilitating learning personalization, real-time feedback, and data-driven analysis, artificial intelligence (AI) is gradually becoming a crucial tool in modern education that improves learning efficiency [1]. AI can dynamically identify the dynamics of learning behavior and adaptively adjust content based on individual needs [2]. Systems such as intelligent tutoring and adaptive learning are gradually improving responsive and scalable learning experiences [3]. In Indonesia, the adoption of artificial intelligence in educational institutions in Indonesia has increased by 35% in the past three years, with a significant impact on student learning performance [4]. Learning performance prediction, academic failure risk detection, and data-driven decision-making assistance [5] also use this technology. However, the success of predictive systems depends on the accuracy of classification algorithms in handling complex and changing variable relationships [6].

Among the most frequently used algorithms in classifying learning outcomes are K-Nearest Neighbor (KNN) and Naive Bayes (NB). While Naive Bayes is effective for large and highly scalable datasets [7], KNN works based on the proximity between samples in the feature space and works well on datasets that have a homogeneous distribution [8]. Although both have their advantages, no single method has been proven to be better for forecasting artificial intelligence-based learning outcomes [9]. Therefore, a comparative and integrative approach is required. Rule-based approaches such as association rule mining (ARM) also appear to be very promising in uncovering hidden relationships between features that cannot be discovered by traditional classification algorithms [10]. ARM helps one to understand the data structure better and improve the prediction accuracy of classification [11] by discovering patterns of coexistence.

Although more widely used in non-educational fields such as medical diagnosis and biological image classification [12], research combining classification algorithms with ARM result extraction features shows significant progress. Adding association features seems to improve the predictive context, but has rarely been quantitatively studied in education [13]. Especially on very large data [14]. The Apriori algorithm in ARM has been frequently used to extract co-existing patterns among traits and improve prediction accuracy. With KNN showing excellent performance [15]. In addition, to improve the dependability of the prediction model, a hybrid approach that combines both [16].

Increasingly researched is the application of ARM in education systems for better prediction accuracy. The SPRAR model, or Student Performance using Relational Association Rules, incorporates association rules among learning features to improve understanding of academic performance variables [17]. Apriori is also used to find correlations between subjects that affect students' academic success [18]. The integration of ARM with Naive Bayes was shown to improve the context of prediction in e-learning systems [19]. Unfortunately, integrative experiments using ARM and categorization are still far from perfection in digital learning systems [12]. Heart disease diagnosis has used a combination of ARM and Naive Bayes, but not much has been adapted for educational needs [20]. This shows the great capacity of ARM to create richer predictive features [21], [22].

The purpose of this study is to compare the performance of KNN and Naive Bayes in predicting AI learning outcomes as well as to investigate the contribution of ARM association features towards improving prediction accuracy. It is expected that the results of this study will provide practical advice in choosing the best classification way for AI-based learning systems.

METHOD

Sample and Research Design

This study uses an exploratory-comparative quantitative design with the aim of comparing the performance of two classification algorithms, namely KNN and Naive Bayes, in predicting the level of Learning AI. The research design applied is cross-sectional, which collects data at a single point in time to reflect the actual condition of the respondents. In addition, this research also uses the Association Rule Mining (ARM) approach through the Apriori algorithm [23], which aims to find patterns of association relationships between construct variables that are non-linear and unclear. This is expected to increase understanding in the predictive classification interpretation process [24]. The research sample consists of 368 Makassar State University students who have experience in using Artificial Intelligence (AI) through independent learning in academics or through homework projects. With a purposive sampling approach, the following criteria were used in the data collection technique:

a. Active undergraduate students majoring in engineering/informatics or computer engineering.
b. Willing to be a responden
c. Fill out the questionnaire validly and completely.

Data Collection

The data was collected using an online survey via Google Forms. The instrument used comprises nine main constructs:

E-ISSN: 3030-8127
P-ISSN: 3031-0318

a. AI learning
b. Job Replacement Anxiety
c. AI configuration
d. Socio-technical blindness
e. AI fear
f. Organizational culture and support
g. Socio-economic status
h. Previous experience with technology
i. Demographic data (gender, experience with AI and AI learning)

Each construct consists of a series of statements rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The data was converted into an average score for each construct variable for each respondent. [24].

Research Steps

A systematic procedure is developed to evaluate the performance of KNN and Naive Bayes algorithms in predicting AI-based learning outcomes following the stages of this research. In addition, this research integrates Association Rule Mining (ARM) using Apriori algorithm to improve prediction accuracy by identifying hidden relationships between variables. The first step in this research is data collection through questionnaires, which is then followed by a data pre-processing stage to prepare the data ready for analysis. After that, two main analytical approaches were applied: classification using KNN and Naive Bayes algorithms, and association exploration through ARM. The results of these two approaches are then compared to come up with stronger conclusions regarding the factors that influence the learning outcomes. The following flowchart illustrates the process undertaken in this research:

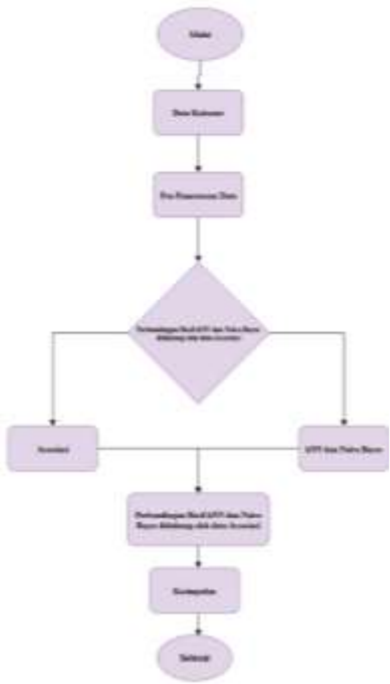

Figure 1. *Flowchart* Research step

The questionnaire data collected from respondents was then processed through a pre-processing stage to clean the data, handle missing data and outliers, and calculate the average score for each construct. By applying the equal-width binning discretization technique developed by Herath[25] on artificial intelligence learning target variables were divided into three categories: Low (1.00-2.99), Medium (3.00-3.99), and High (4.00-5.00). The analysis process was conducted through two main paths: first, using Association Rule Mining (ARM) with the Apriori algorithm to identify correlations between variables. Second, using KNN and Naive Bayes algorithms to classify AI learning levels. The performance of the model in classification is measured using metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix[26]. The results of ARM are then analyzed to determine its contribution in improving the accuracy of the classification model. Based on the evaluation and analysis conducted, conclusions regarding the best classification model can be drawn, along with recommendations for further implementation.

Analisis Data

Pre-Processing Data

Before the application of the algorithm, data pre-processing is the first and important step. The following are the stages performed:

a. Through the File section, clean data sets are uploaded. Since low data quality can cause errors in classification and association results [27], [28], the data cleaning process is quite important to ensure that the research conducted is accurate and reliable.
b. The organization and quality of the data is evaluated before further processing is done using data tables. This process helps to find possible errors in the dataset that may affect the performance of the model [29], [30].
c. The dataset is divided into training set to train the Naive Bayes and KNN models, and testing set to test and validate the models. This ensures the model can generalize new data, not just memorize. [31].

Classification Using KNN and Naive Bayes

After the data pre-processing stage is complete, the next step in this research is to perform classification using two proven effective algorithms, namely K-Nearest Neighbor (KNN) and Naive Bayes. These two algorithms are applied to predict AI Learning categories based on pre-processed data [32].

a. Training

KNN and Naive Bayes are trained using the Training Set. At this stage, the algorithms learn patterns and relationships in the data that are relevant to the AI Learning category to be predicted.

b. Model Evaluation

After training, the model was evaluated using the Test and Score component. The performance metrics used in this evaluation include:

Accuracy

Accuracy is a measure used to assess how many predictions are correct compared to the total number of predictions made. This metric provides a general insight into how effective the model is in classifying the data as a whole [32]. Accuracy indicates the correctness of all predictions made by the model. While easy to understand, accuracy can be underrepresentative in cases of imbalanced data. The formula for calculating accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Description:
TP (True Positive): Correct prediction for positive class
TN (True Negative): Correct prediction for negative class
FP (False Positive): False prediction for positive class
FN (False Negative): False prediction for negative class

Precision

Precision is a metric measuring how many of the positive predictions are actually positive, describing the quality of the model's predictions in identifying positive categories [33]. The following is the formula for calculating precision:

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

Recall

Recall serves to measure how many positive cases are successfully identified by the model, it is important to assess the model's ability to detect all relevant examples in the positive category [34]. The formula for calculating recall:

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

F1-Score

F1-Score is the harmonic mean between precision and recall, providing a more balanced picture of model performance, especially when there is an imbalance between the two [35]. about the performance of the model, especially when there is an imbalance between the two [35]. Here is the formula for calculating F1-Score:

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

F1-Score provides a single value that unifies precision and recall to give an overall picture of the model's performance.

Prediksi

The trained model is used to predict the AI Learning category on the Testing Set. These prediction results are displayed using the Predictions component for KNN and Naive Bayes.

Association Rule Mining

Association Rule Mining (ARM) is a method in data mining used to identify unseen patterns or links between elements in a dataset. One of the most commonly used algorithms in ARM is Apriori, whose role is to find frequently occurring itemsets and form association rules. This algorithm runs on the principle that frequently occurring itemsets can be utilized to discover robust association rules, thus providing valuable insights in data analysis [36], [37].

a. Frequent Itemsets

In the initial stage, Apriori is used to detect frequent itemsets, which is a set of items that often appear together in transactions. This process starts by searching for single-item itemsets that meet the minimum support limit. Support measures how often itemsets appear in the dataset and is calculated by the following formula:

$$Support\ (X) = \frac{Number\ of\ transactions\ that\ contain\ X}{Total\ number\ of\ transactions}$$

Itemsets that meet the minimum support are considered frequent itemsets. The algorithm then expands its search to larger itemsets (e.g., two-item, three-item, and so on) until no larger itemsets meet the support threshold [38], [39]. This process is essential for building association rules that will be analyzed further, providing a solid foundation for data-driven decision making [40].

b. Association Rules

Setelah frequent itemsets ditemukan, langkah selanjutnya adalah menghasilkan association rules menggunakan dua metrik utama yaitu Confidence dan Lift.

- Confidence serves to measure the probability that item Y will appear if item X has appeared. It shows how strong the relationship between the two items is. Confidence Formula:

$$Confidence\ (X \rightarrow Y) = \frac{Support\ (X \cup Y)}{Support\ (X)}$$

  A high confidence indicates that if item X appears, then item Y is likely to appear as well, making it a useful indicator in behavior analysis [41].

- Lift is used to measure the strength of association between two items by comparing the probability of itemsets X and Y occurring together, compared to the probability of them occurring independently. If Lift is greater than 1, then the association is stronger than what would be expected if the two items were separate. Here is the lift formula:

$$Lift\ (X \rightarrow Y) = \frac{Confidence\ (X \rightarrow Y)}{Support\ (Y)}$$

  A lift greater than 1 indicates a strong positive association between X and Y, while a lift less than 1 indicates a negative association [42].

c. Feature Statistics

Feature Statistical Analysis is used to explore the distribution of features in the dataset and their contribution to the formation of association rules. Features that frequently appear in frequent itemsets will be analyzed to understand their relevance to the association rules found. This statistical analysis allows researchers to understand the variables that affect learning outcomes and identify factors that are closely related to AI learning [43], [44].

RESULTS AND DISCUSSION

Analisis Descriptif

This study involved 368 students from Makassar State University who had experience in using artificial intelligence (AI), either through academic courses or independent exploration. The data collected from the respondents included demographic information and various factors that could affect their AI learning outcomes [3]. The characteristics of the respondents obtained from this study are presented in Table 1 below:

Table 1. Respondent characteristics

Characteristics	Categoties	Proportion
Gender	Male	42,66%
	Female	57,34%
Learning AI Targetss	1 = Very Low	6,52%
	2 = Low	32,88%
	3 = Medium	47,55%
	4 = High	8,70%
	5 = Very High	4,35%
AI Technology Usage Experience	None	5,16%
	Beginner	49,46%
	Intermediate	43,21%
	Advance	2,17%

Table 1 shows the distribution of respondents based on their gender, AI experience, and AI learning targets. The majority of the respondents were female (57.34%), and most had AI technology experience at the beginner level (49.46%). This suggests that many respondents are just starting to use AI technologies, which may affect their understanding and use of AI learning systems. Previous research also states that technology experience has a great influence on one's ability to use AI-based tools in education [45].

Performance Evaluation of Naive Bayes and KNN on Training Data

Next, the performance of Naive Bayes and K-Nearest Neighbors (KNN) classification models on training data was evaluated. Training data is used to train both models before being tested on test data to assess their ability to predict learning outcomes. The evaluation results of both models on training data are presented in Table 2:

Table 2. *Test and Score Data Train Naïve Bayes*

Model	AUC	CA	F1	Precision	Recall	MCC
Naïve Bayes	0.661	0.373	0.387	0.624	0.373	0.200

Based on Table 2, the Naive Bayes model obtained an AUC value of 0.661, which reflects a moderate ability to distinguish between positive and negative classes. Although the resulting precision is quite high (0.624), the low recall value (0.373) indicates that the model has not been able to maximally recognize all relevant cases. This is an important concern in the context of data-driven prediction, where the ability to detect all positive examples is crucial [32].

Table 3. *Test and Score Data Train KNN*

Model	AUC	CA	F1	Precision	Recall	MCC
KNN	0.562	0.545	0.523	0.503	0.545	0.148

In comparison, Table 3 shows that despite the lower AUC value of KNN (0.562 compared to 0.661 in Naive Bayes), the KNN model shows higher accuracy (0.545 vs 0.373) and recall (0.545 vs 0.373). This finding suggests that while KNN is less effective in clearly distinguishing between positive and negative classes, it has the advantage of recognizing more relevant cases. This ability is important, especially in applications that demand a wider and more diverse range of predictions [33].

Comparison of KNN vs Naive Bayes on Test Data

After training the models, the evaluation continued on the test data to assess the generalization ability of the two models on data that had not been seen before. The evaluation results on the test data are presented in Table 4:

Table 4. Prediction Comparison of NB vs KNN

Model	AUC	CA	F1	Precision	Recall	MCC
Naïve Bayes	0.577	0.391	0.404	0.489	0.391	0.090
KNN	0.551	0.509	0.484	0.472	0.509	0.081

Table 4 shows that Naive Bayes is superior in terms of AUC (0.577 compared to 0.551) and precision (0.489 compared to 0.472) when compared to KNN. However, in terms of accuracy (0.545 vs 0.373) and recall (0.545 vs 0.391), KNN performed better. This indicates that while Naive Bayes is superior in prediction accuracy and class discrimination, KNN is more effective in recognizing more relevant cases, which is a plus in the context of wide-ranging predictions.

Confusion Matrix Analysis

The Confusion Matrix presents a more detailed picture of each model's performance in classifying data. This matrix shows the distribution of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), which allows to identify the types of errors that occur more frequently in each model. This is particularly useful for understanding the classification errors made by both algorithms and correcting them.The confusion matrix for Naive Bayes and KNN are presented in table 5 and table 6 which provide further insight into how each model predicts the following Classification results.

Table 5. Evaluation *Confusion Matrix Naïve Bayes*

	Predicted				
Actual	**Advance**	**Intermediate**	**Beginner**	**None**	**Σ**
Advance	2	0	0	0	**2**
Intermediate	22	4	15	7	**48**
Beginner	18	2	31	3	**54**
None	0	0	2	4	**6**
Σ	**42**	**6**	**48**	**14**	**110**

In Table 5, the confusion matrix generated by Naive Bayes shows a higher number of True Positives (TP), indicating that the model is more efficient in identifying positive categories. Despite having a high precision value (0.624), the model also shows a significant False Negative (FN) value, indicating that most of the cases that should have been predicted as positive were not successfully detected. This suggests that Naive Bayes tends to have difficulty in capturing all relevant cases despite its high accuracy. In contrast, in Table 6, the confusion matrix generated by KNN shows a higher recall than Naive Bayes, indicating KNN's ability to capture more relevant results. However, although KNN successfully identifies more positive cases (with a recall of 0.509), the model also generates a higher False Positive (FP), which indicates that a lot of negative data is misclassified as positive. As a result, although the recall of KNN is higher, the resulting precision (0.472) is lower, indicating a trade-off between capturing more relevant results and increasing positive misclassification.

Table 6. Evaluation *Confusion Matrix KNN*

	Predicted				
Actual	**Advance**	**Intermediate**	**Beginner**	**None**	**Σ**
Advance	0	2	0	0	**2**
Intermediate	0	24	24	0	**48**
Beginner	0	18	36	0	**54**
None	0	3	3	0	**6**
Σ	**0**	**47**	**63**	**0**	**110**

Association Rule Mining (ARM)

Penelitian ini turut menerapkan Association Rule Mining (ARM) dengan algoritma Apriori untuk mengungkap hubungan tersembunyi antar variabel dalam dataset. ARM dimanfaatkan untuk mengidentifikasi frequent itemsets, yaitu kombinasi item yang kerap muncul bersamaan dalam data, yang selanjutnya digunakan untuk membentuk association rules. Aturan-aturan ini memberikan pemahaman yang lebih mendalam terkait faktor-faktor yang memengaruhi hasil pembelajaran dalam konteks sistem berbasis AI. Table 7 menampilkan frequent itemsets yang ditemukan berdasarkan variabel jenis kelamin dan target pembelajaran AI, yang memberikan gambaran mengenai kombinasi variabel yang paling sering muncul bersama dalam dataset.

Table 7. Frequent Itemsets by Gender and AI Learning Target

Itemsets	Quantity	Support
Gender=Male	157	42,66%
Learning AI Targetss =1	11	2,99%
Learning AI Targetss=2	52	14,13%
Learning AI Targetss=3	72	19,57%
Learning AI Targetss=4	16	4,35%
Learning AI Targetss=5	6	1,63%
Gender=Female	211	57,34%
Learning AI Targetss=1	13	3,53%
Learning AI Targetss=2	69	18,75%
Learning AI Targetss=3	103	27,99%
Learning AI Targetss=4	16	4,35%
Learning AI Targetss=5	10	2,72%

Table 8 presents the association rules generated based on these itemsets, showing the relationship between gender and AI learning outcomes based on target achievement.

Table 8. Association Rules Based on Gender and Learning AI Targets

Antecedent	Consequent	Support	Confidence	Coverage	Strength	Lift	Leverage
Learning AI Targetss=1	Gender=Male	0,03	0,458	0,065	6,542	1,074	0,002
Learning AI Targetss=1	Gender=Female	0,035	0,542	0,065	8,792	0,945	-0,002
Learning AI Targetss=2	Gender=Male	0,141	0,43	0,329	1,298	1,007	0,001
Learning AI Targetss=2	Gender=Female	0,141	0,331	0,427	0,771	1,007	0,001
Learning AI Targetss=3	Gender=Male	0,188	0,327	0,573	1,744	0,995	-0,001
Learning AI Targetss=3	Gender=Female	0,196	0,329	0,476	0,897	0,964	-0,001
Learning AI Targetss=3	Gender=Male	0,196	0,459	0,411	1,115	0,964	0,007
Learning AI Targetss=4	Gender=Male	0,043	0,427	0,476	6,204	1,172	0,006
Learning AI Targetss=4	Gender=Female	0,043	0,087	0,043	6,594	1,159	0,002
Learning AI Targetss=5	Gender=Male	0,016	0,375	0,043	9,812	1,019	-0,001
Learning AI Targetss=5	Gender=Female	0,027	0,625	0,043	13,188	1,109	0,002

Learning AI Targetss=5	Gender=Female	0,188	0,573	0,573	0,573	0,995	0,002
Gender=Male	Learning AI Targets=2	0,141	0,573	0,476	0,771	1,007	0,002
Gender=Male	Learning AI Targets=3	0,196	0,459	0,476	0,897	1,028	0,006
Gender=Male	Learning AI Targets=4	0,043	0,427	0,043	6,594	1,109	0,002
Gender=Male	Learning AI Targets=5	0,027	0,043	0,043	0,879	1,109	0,002

Table 8 shows that there is a strong positive relationship between gender and AI learning target achievement. A lift value greater than 1 indicates that the relationship between the two variables is not random, but rather has a significant and meaningful association. The findings also reveal that Naive Bayes and K-Nearest Neighbor (KNN) algorithms have unique characteristics in predicting AI-based learning outcomes. Naive Bayes shows an advantage in precision, making it more suitable for applications that prioritize accuracy in recognizing positive classes. This is particularly important in contexts where classification errors, such as identifying high or low performing students, must be minimized [9]. However, the main drawback of the Naive Bayes model lies in its low recall value, which indicates that it tends to fail to detect some cases that actually belong to the positive category.

In contrast, KNN shows superiority in terms of recall, which is a key factor for systems that aim to capture as many relevant results as possible, despite the higher risk of generating False Positives (FP). This advantage in recall is particularly important in applications such as early detection of academic failure, where the main focus is to identify as many cases as possible that require attention, albeit with a higher probability of misclassification. Although the precision of KNN is at a lower level, this result reflects the ability of KNN to capture more variation in the data, which is particularly meaningful in applications that require broader predictive coverage [33].

The application of Association Rule Mining (ARM) to explore the relationship between variables, especially related to the effect of gender on learning outcomes, is in line with findings that show that the addition of association features in classification models can improve the accuracy and interpretability of prediction results [13]. In this study, the ARM method is utilized to uncover hidden patterns between variables that may not be identified by conventional classification algorithms such as Naive Bayes and KNN. By adding this information, the model becomes more robust and is able to provide a deeper understanding of the factors that influence learning outcomes, as has been applied in the context of classification in the medical field[2].

These findings emphasize the need for a more comprehensive data-driven approach to strengthen the decision-making process in technology-enabled learning contexts. The application of ARM facilitates the identification of more complex and non-linear patterns among the various factors influencing learning outcomes, potentially improving the effectiveness of predictions in educational contexts [46]. Therefore, this research recommends using a hybrid approach that combines the strengths of Naive Bayes and KNN, as well as ARM, to build a more robust and accurate prediction system in AI-based learning. This fusion will not only improve accuracy, but also enrich the classification results with additional information that can help decision makers adjust learning interventions [47]. With this approach, AI-based learning systems can more effectively cope with the complexity of large and diverse educational data.

E-ISSN: 3030-8127
P-ISSN: 3031-0318

CONCLUSIONS

As a result of this study, K-Nearest Neighbor (KNN) and Naive Bayes differ in their capacity to predict the outcome of artificial intelligence (AI) learning. Better precision makes Naive Bayes suitable for applications that require accuracy in identifying positive classes. In contrast, KNN excels in terms of recall, which helps one to identify more relevant outcomes in the data. The integration of Association Rule Mining (ARM) with Apriori algorithm shows noteworthy results in uncovering underlying patterns among variables that cannot be discovered by conventional classification algorithms, therefore improving prediction accuracy by enriching the characteristics used in classification. Under the setting of artificial intelligence-based learning, hybrid model building that combines the strengths of Naive Bayes, KNN, and ARM can produce a more trustworthy and precise prediction system. This method can cope with the complexity of highly diverse and extreme education data. Further research is recommended to investigate the combination of other algorithms with ARM to improve the prediction results and find outside influences on the learning results. A more sophisticated prediction system can help guide more accurate data-driven choices, thus strengthening the use of artificial intelligence in education

REFERENCES

[1] A. Anwarudin, W. Andriyani, B. P. DP, and D. Kristomo, "The Prediction on the Students' Graduation Timeliness Using Naive Bayes Classification and K-Nearest Neighbor," *J. Intell. Softw. Syst.*, vol. 1, no. 1, p. 75, 2022, doi: 10.26798/jiss.v1i1.597.

[2] Jiao, "AI-Driven Personalization in Higher Education: Enhancing Learning Outcomes through Adaptive Technologies," *Adult High. Educ.*, vol. 6, no. 6, 2024, doi: 10.23977/aduhe.2024.060607.

[3] A. Akavova, Z. Temirkhanova, and Z. Lorsanova, "Adaptive learning and artificial intelligence in the educational space," *E3S Web Conf.*, vol. 451, pp. 1–4, 2023, doi: 10.1051/e3sconf/202345106011.

[4] A. Lei, "Use Cases of AI in Pharmaceutical Industry." [Online]. Available: https://www.linkedin.com/pulse/use-cases-ai-pharmaceutical-industry-seema-karwa/

[5] M. Munarsih and B. A. Ningsi, "Performance Comparison of Data Mining Classification Algorithms on Student Academic Achievement Prediction," *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 1, p. 29, 2023, doi: 10.24014/ijaidm.v6i1.21874.

[6] A. A. Permana and A. Arsanah, "Comparison of K-Nearest Neighbor Classification and Naive Bayes Classifier in Analysis of Heart Disease," *J. Intell. Comput. Heal. Informatics*, vol. 5, no. 1, pp. 35–42, 2024, doi: 10.26714/jichi.v5i1.11251.

[7] T. Sun, "Importance of Feature Extraction in Naïve Bayes Applied on Animal Classification," *Highlights Sci. Eng. Technol.*, vol. 124, pp. 244–248, 2025, doi: 10.54097/6dp2f003.

[8] I. A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naïve Bayesian," in *ICIT 2017 - 8th International Conference on Information Technology, Proceedings*, in, Ed., Amman: IEEE, 2017, pp. 909–913. doi: 10.1109/ICITECH.2017.8079967.

[9] S. Alneyadi and Y. Wardat, "ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools," *Contemp. Educ. Technol.*, vol. 15, no. 4, 2023, doi: 10.30935/cedtech/13417.

[10] R. N. Ichsan, L. Nasution, and S. Sinaga, "Studi kelayakan bisnis = Business feasibility study," *CV. Sentosa Deli Mandiri*.

[11] S. M. Ramesh, N. Sengottaiyan, D. Vanathi, R. Manoja, K. Tamizharasu, and P. Kalyanasundaram, "Prediction of Cardiac Disease Using Naive Bayes Algorithm," in *2nd International Conference on Sustainable Computing and Smart Systems, ICSCSS 2024 - Proceedings*, India: IEEE, 2024, pp. 994–997. doi: 10.1109/ICSCSS60660.2024.10624905.

[12] L. Chowdhury *et al.*, "A biological data-driven mining technique by using hybrid classifiers with rough set," *Int. J. Ambient Comput. Intell.*, vol. 12, no. 3, pp. 123–139, 2021, doi: 10.4018/IJACI.2021070106.

[13] A. Alaiad, H. Najadat, B. Mohsen, and K. Balhaf, "Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease," *J. Inf. Knowl. Manag.*, vol. 19, no. 1, p. 2040015, 2020, doi: 10.1142/S0219649220400158.

[14] P. Kaur *et al.*, "Does ChatGPT Add Value to University Students' Academic Performance?," *Int. J. Membr. Sci. Technol.*, vol. 10, no. 3, pp. 2800–2811, 2023, doi: 10.15379/ijmst.v10i3.2583.

[15] F. Al Azami, A. A. Riadi, and E. Evanita, "Klasifikasi Kualitas Wortel Menggunakan Metode K-Nearest Neighbor Berbasis Android," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 7, no. 1, p. 36, 2022, doi: 10.30645/jurasik.v7i1.413.

[16] A. Winantu and C. Khatimah, "Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa," *INTEK J. Inform. dan Teknol. Inf.*, vol. 6, no. 1, pp. 58–64, 2023, doi: 10.37729/intek.v6i1.3006.

[17] M. Zoromba, H. E. El-Gazar, A. Salah, H. O. Elboraie, A.-H. El-Gilany, and A. H. El-Monshed, "Effects of Emotional Intelligence Training on Symptom Severity in Patients With Depressive Disorders," *Clin. Nurs. Res.*, 2022, doi: 10.1177/10547738221074065.

[18] A. W. Syaputri, E. Irwandi, and M. Mustakim, "Naïve Bayes Algorithm for Classification of Student Major's Specialization," *J. Intell. Comput. Heal. Informatics*, vol. 1, no. 1, p. 17, 2020, doi: 10.26714/jichi.v1i1.5570.

[19] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Relationship between student engagement and performance in e-learning environment using association rules," in *EDUNINE 2018 - 2nd IEEE World Engineering Education Conference: The Role of Professional Associations in Contemporaneous Engineer Careers, Proceedings*, Buenos Aires: IEEE, 2018, pp. 1–6. doi: 10.1109/EDUNINE.2018.8451005.

[20] S. A. Hudli, A. V. Hudli, and A. A. Hudli, "Application of data mining to candidate screening," in *Proceedings of 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies, ICACCCT 2012*, Ramanathapuram, India: IEEE, 2012, pp. 287–290. doi: 10.1109/ICACCCT.2012.6320788.

[21] G. R. Thummala, R. Baskar, and N. Thiyaneswaran, "Prediction of Heart Disease Using Naive Bayes in Comparison with KNN Based on Accuracy," in *International Conference on Cyber Resilience, ICCR 2022*, Dubai, United Arab Emirates: IEEE, 2022, pp. 1–4. doi: 10.1109/ICCR56254.2022.9995841.

[22] A. A. M. Ahmed *et al.*, "Introductory Engineering Mathematics Students' Weighted Score Predictions Utilising a Novel Multivariate Adaptive Regression Spline Model," *Sustain.*, vol. 14, no. 17, p. 11070, 2022, doi: 10.3390/su141711070.

[23] M. Si *et al.*, "Triglycerides as Biomarker for Predicting Systemic Lupus Erythematosus Related Kidney Injury of Negative Proteinuria," *Biomolecules*, vol. 12, no. 7, pp. 1–12, 2022, doi: 10.3390/biom12070945.

[24] D. Dwiputra, A. Mulyo Widodo, H. Akbar, and G. Firmansyah, "Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis to

Discover Rules of Item Combinations," *J. World Sci.*, vol. 2, no. 8, pp. 1229–1248, 2023, doi: 10.58344/jws.v2i8.403.

[25] E. Herath and U. Wijenayake, "A Novel Approach to Enhance the Efficiency of Apriori Algorithm," *Proc. Conf. Transdiscipl. Res. Eng.*, vol. 1, no. 1, 2024, doi: 10.31357/contre.v1i1.7372.

[26] B. H. Situmorang, A. Isra, D. Paragya, and D. A. A. Adhieputra, "Apriori Algorithm Application for Consumer Purchase Patterns Analysis," *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 21, no. 1, pp. 15–20, 2024, doi: 10.33751/komputasi.v21i1.9260.

[27] Y. Ariyanto, B. Harijanto, and A. N. Asri, "Analyzing Student's Learning Interests in the Implementation of Blended Learning Using Data Mining," *Int. J. Online Biomed. Eng.*, 2020, doi: 10.3991/ijoe.v16i11.16453.

[28] K. Pawluszek-Filipiak and A. Borkowski, "On the importance of train-test split ratio of datasets in automatic landslide detection by supervised classification," *Remote Sens.*, vol. 12, no. 18, 2020, doi: 10.3390/rs12183054.

[29] A. D. W. M. Sidik, I. Himawan Kusumah, A. Suryana, Edwinanto, M. Artiyasa, and A. Pradiftha Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," *Fidel. J. Tek. Elektro*, vol. 2, no. 2, pp. 34–38, 2020, doi: 10.52005/fidelity.v2i2.111.

[30] J. Smucny *et al.*, "Predicting conversion to psychosis using machine learning: response to Cannon," *Front. Psychiatry*, vol. 15, no. January, pp. 1–6, 2024, doi: 10.3389/fpsyt.2024.1520173.

[31] P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of Heart Failure," *Ilk. J. Ilm.*, 2022, doi: 10.33096/ilkom.v14i2.1148.120-125.

[32] M. Zuo, B. Yu, and L. Sui, "Application of EEG-based Machine Learning in Stock Trading-related Emotion Recognition," 2024, doi: 10.4108/eai.27-10-2023.2342016.

[33] N. Van Thieu, "PerMetrics: A Framework of Performance Metrics for Machine Learning Models," *J. Open Source Softw.*, vol. 9, no. 95, p. 6143, 2024, doi: 10.21105/joss.06143.

[34] G. Mostafa, H. Mahmoud, T. Abd El-Hafeez, and M. E. ElAraby, "Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00944-3.

[35] A. Zeng, H. Yu, Q. Da, Y. Zhan, and C. Miao, "Accelerating ranking in e-commerce search engines through contextual factor selection," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 13212–13219, 2020, doi: 10.1609/aaai.v34i08.7026.

[36] L. Hamdad and K. Benatchba, "Association Rules Mining Exact, Approximate and Parallel Methods: A Survey," *SN Comput. Sci.*, vol. 2, no. 6, 2021, doi: 10.1007/s42979-021-00819-x.

[37] P. Mehrannia, B. Moshiri, and O. Basir, "Knowledgebase approximation using association rule aggregation," *Int. J. Data Sci. Anal.*, vol. 13, no. 3, pp. 225–237, 2022, doi: 10.1007/s41060-021-00304-x.

[38] M. Peng, S. Lee, A. G. D'Souza, C. T. A. Doktorchik, and H. Quan, "Development and validation of data quality rules in administrative health data using association rule mining," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, 2020, doi: 10.1186/s12911-020-1089-0.

[39] B. Bouaita, A. Beghriche, A. Kout, and A. Moussaoui, "A New Approach for Optimizing the Extraction of Association Rules," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 2, pp. 10496–10500, 2023, doi: 10.48084/etasr.5722.

[40] A. Mokkadem, M. Pelletier, and L. Raimbault, "Association rules and decision rules," *Stat. Anal. Data Min.*, vol. 16, no. 5, pp. 411–435, 2023, doi: 10.1002/sam.11620.

[41] F. Bao, L. Mao, Y. Zhu, C. Xiao, and C. Xu, "An Improved Evaluation Methodology for Mining Association Rules," *Axioms*, vol. 11, no. 1, 2022, doi: 10.3390/axioms11010017.

[42] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "MS-Net: Multi-Site Network for Improving Prostate Segmentation with Heterogeneous MRI Data," *IEEE Trans. Med. Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020, doi: 10.1109/TMI.2020.2974574.

[43] L. Szathmary, "Closed association rules," *Ann. Math. Informaticae*, vol. 51, pp. 65–76, 2020, doi: 10.33039/ami.2020.07.009.

[44] D. W. Wardani, "Measuring positive and negative association of apriori algorithm with cosine correlation analysis," *Baghdad Sci. J.*, vol. 18, no. 3, pp. 554–564, 2021, doi: 10.21123/BSJ.2021.18.3.0554.

[45] N. Kim, H. Oh, and J. K. Choi, "A privacy scoring framework: Automation of privacy compliance and risk evaluation with standard indicators," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 1, pp. 514–525, 2023, doi: 10.1016/j.jksuci.2022.12.019.

[46] S. H. Jin, K. Im, M. Yoo, I. Roll, and K. Seo, "Supporting students' self-regulated learning in online learning using artificial intelligence applications," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 37, 2023, doi: 10.1186/s41239-023-00406-5.

[47] I. Riadi, R. Umar, and R. Anggara, "Comparative Analysis of Naive Bayes and K-NN Approaches to Predict Timely Graduation using Academic History," *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, pp. 1163–1174, 2024, doi: 10.12785/ijcds/160185.