BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0557–0572
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss1pp0557-0572

ANALYSIS OF APRIORI AND K-NEAREST NEIGHBOR (KNN) ALGORITHM IN RECOMMENDING APPROPRIATE LEARNING METHOD

Nuril Lutvi Azizah[1*], **Ade Eviyanti**[2], **Novia Ariyanti**[3], **Gita Wardani**[4], **Naila Farah Diba**[5]

[1,2,3,4,5]*Informatics Department, Faculty of Science and Technology, Universitas Muhammadiyah Sidoarjo*
Jln. Raya Gelam No 250, Candi, Sidoarjo, Jawa Timur, 71060, Indonesia

Corresponding author's e-mail: * **nurillutviazizah@umsida.ac.id**

Article Info

Article History:

Received: 21st April 2025
Revised: 2nd June 2025
Accepted: 21st July 2025
Available online: 24th November 2025

Keywords:

Apriori algorithm;
Classification;
K-Nearest neighbor;
Learning methods.

ABSTRACT

The study investigates the utilization of data mining techniques, especially the Apriori algorithm and K-Nearest Neighbor (KNN) classification, in recommending appropriate learning methods based on student data. The purpose of this research is to analyze patterns and groupings in students' behavior, preferences, and academic performance to support more informed and personalized educational strategies. The Apriori algorithm is used to identify frequent associations among learning related attributes, while KNN classification helps group students with similar learning characteristics. The analysis revealed that the digital learning method is the most preferred by students, with a percentage of 84.29%, followed by the traditional lecture method at 15.70%. These results reflect a notable trend toward technology-driven, flexible learning environments, although conventional approaches continue to hold relevance for a portion of learners. The research concludes that the integration of the Apriori algorithm and KNN clustering proves to be an effective analytical framework for facilitating adaptive learning. This approach allows educators and institutions to make data-driven decisions in tailoring instructional methods that align with the diverse needs and preferences of students.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

N. L. Azizah, A. Eviyanti, N. Ariyanti, G. Wardani, and N. F. Diba, "ANALYSIS OF APRIORI AND K-NEAREST NEIGHBOR (KNN) ALGORITHM IN RECOMMENDING APPROPRIATE LEARNING METHOD", *BAREKENG: J. Math. & App.*, vol. 20, no. 1, pp. 0557-0572, Mar, 2026.

Copyright © 2026 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

In the era of the Industrial Revolution 4.0, digital transformation has become an inseparable part of the world of education. Universities, as the highest-level institutions, have the responsibility to prepare the younger generation to compete in the global workforce that increasingly relies on technology [1]. One of the challenges faced by higher Education, especially mathematics courses such as calculus, which are compulsory courses in many study programs, is creating an inclusive learning experience that is in line with students' needs. According to research conducted, most students have difficulty in understanding basic mathematical concepts because the basics of mathematics, such as multiplication and division, during elementary school mathematics learning have not been mastered [2]. This challenge is even greater for students with special needs or those from less adequate educational backgrounds [3]. Counting in numeracy is not more than just about mastering mathematical material in college, but also involves the ability to think critically and logically on real problems that require problem solving [4]. In line with the Sustainable Development Goals (SDGs), especially goal 4 on quality education, innovative efforts are needed to ensure that every individual, regardless of social, economic, physical, or cognitive background, can obtain equitable and quality education [5]. The Independent Learning Curriculum implemented in Indonesia also emphasizes the importance of flexibility and independence in the learning process, which gives students the freedom to explore material according to their own pace and learning style [6].

In previous research, digital flipbooks have been applied to more complex learning, such as mathematics. Meanwhile, integration with project-based learning (PBL) for inclusive mathematics learning is a new approach that has not been widely studied, especially in the context of higher education. Through digital animation technology, mathematics learning will become more interesting, interactive, and applicable. Previous research that applies flipbook learning media includes the development of flipbook learning media based on scientific literacy in elementary school science learning [7]. The result of this study indicated that as many a 91% stated that flipbook learning media is very suitable for use as a science learning media. Development of thematic learning e-modules (EMOTIK) based on flipbooks in Elementary School [8]. This study measures the level of validity of flipbook-based thematic learning media. The results of this study indicate that research using flipbook media provides valid, effective, and interesting results so that it can be used as a learning medium. Research conducted [9]. Likewise, previously conducted research that discussed the relationship between active, innovative, and enjoyable learning can influence students' learning style [10]. Several studies related to learning media using flipbooks are used in science classes in elementary schools, similar to the Digital Flipbook-Based Learning Media Design study implemented in the growth of living things of elementary schools [11], which obtained satisfactory results related to the development of flipbook media. In addition, the development of character-based flipbook e-module media has been carried out in science learning in elementary school [12]. The results of previous studies indicate that the validation of the use of flipbook media as a learning medium has obtained good validation and can be continued to the development stage; however, the discussion related to the application of flipbooks in mathematics or numeracy learning combined with the PBL method in education has not been widely studied before.

In response to Sustainable Development Goal (SDG) 4 on inclusive and equitable quality education, there is a need for innovations that support flexible, student-centered learning. The "Independent Learning" policy in Indonesia emphasizes this direction by encouraging self-paced, adaptive instruction tailored to each learner's style and capability. While innovation such as digital flipbooks and project-based learning (PBL) has gained traction in science education, their integration into higher education mathematics, particularly through data-driven approaches, remains underexplored. This study addresses this gap by leveraging data mining techniques to assist in recommending the most appropriate learning methods for students based on behavioral and performance data.

Despite the growing use of digital tools in education, there is a lack of systematic methods to personalize learning method selection based on student data. This aims to develop a model that uses data mining techniques, specifically the Apriori algorithm [13] and the K-Nearest Neighbor (KNN) algorithm, to analyze student learning characteristics and recommend suitable instructional methods. The research questions are: What pattern in student learning behavior and preferences can be identified using the Apriori Algorithm? How accurately can the K-Nearest Neighbor (KNN) algorithm classify students based on learning attributes to predict suitable learning methods? And how does the integration of Apriori and KNN support educators in making data-driven instructional decisions?

In this research, the Apriori Algorithm is selected for its strength in mining association rules from categorical data, helping identify frequent patterns among student preferences for course content and chosen learning methods. The KNN algorithm is a robust, non-parametric classifier that effectively predicts labels (learning method types) for new student data based on proximity to labeled training data [14]. Together, these algorithms offer a hybrid data mining approach, Apriori for pattern discovery and KNN for predictive classification that supports informed, personalized learning method recommendations.

The use of the Apriori algorithm has traditionally been applied in transactional domains such as market basket analysis [13], [15], enabling the discovery of frequent item sets and association rules. However, its adaptation to educational data mining is still relatively rare. Hanifan and Putra [13] applied Apriori in product clustering, which mirrors the objective of identifying co-occurring learning behaviors in students. In an educational context, Apriori could be used to find meaningful patterns such as "students who choose method X also prefer media Y". Some studies have explored this potential indirectly. Rahmi and Mikola [16], for instance, used Apriori to detect consumer purchasing habits; this method could be analogously used to determine students' learning method selection patterns. Nonetheless, these studies have not examined Apriori's role in guiding instructional strategy formulation, which this study seeks to address.

The KNN algorithm is a widely adopted classification technique due to its simplicity and effectiveness in dealing with multi-dimensional data. KNN has been applied successfully in various educational contexts, such as predicting student graduation likelihood [17], tuition group classification [18], and disease risk modeling [19]. Sabri and Hamrizan [20] demonstrated KNN's reliability in predicting academic performance using multi-attribute datasets. Although KNN is well-documented in education-related prediction, few studies have leveraged its classification capability to recommend instructional approaches or learning methods based on students' prior learning behavior. The gap remains in applying KNN to classify and group students by learning needs or preferences, particularly in numeracy or mathematics subjects. Most of the existing literature treats Apriori and KNN in isolation and applies them to either non-educational domains or narrow prediction tasks. In this research, Apriori for association rule mining has been integrated with KNN for classification to analyze and recommend learning methods in the domain of inclusive mathematics education [21]. Moreover, studies on the use of these algorithms often lack a focus on personalized, data-driven instruction, which is increasingly necessary in a diverse, digital education landscape. This study contributes to filling this gap by synthesizing Apriori and KNN in a hybrid framework, applying the model to real-world student data, and focusing specifically on inclusive and adaptive learning method recommendations.

2. RESEARCH METHODS

This study uses a quantitative research method based on educational data mining to analyze student learning preferences and predict appropriate instructional methods. The data collection, preprocessing, algorithm implementation, and validation steps are detailed. The data on the influence of the learning method used to predict the appropriate method in learning is a dataset consisting of 150 responses was obtained from a structured questionnaire distributed to Informatics students during the odd Semester of 2024-2025. The data obtained is then processed to obtain lesson data that is in accordance with the needs of predicting the learning methods used. In making predictions of learning methods, there are 9 variables that will be used as attributes to process the dataset, consisting of 9 independent variables and 1 dependent variable. The independent variables consist of student name, student ID number, grades of the students, reason why the student chose the subject, course content, learning method, influence or impact of learning method, factors influencing learning method, learning models, while the dependent variable is the influence of learning method/media.

Data Preprocessing Steps:

1. Irrelevant attributes such as Name and ID were excluded from modeling.
2. Categorical variables were encoded into a numeric format using label encoding.
3. Data was normalized using Min-Max Normalization to ensure equal scale for all attributes.
4. The dataset was split into training (60%) and test (40%) using the train_test_split function in Python (Scikit-learn).

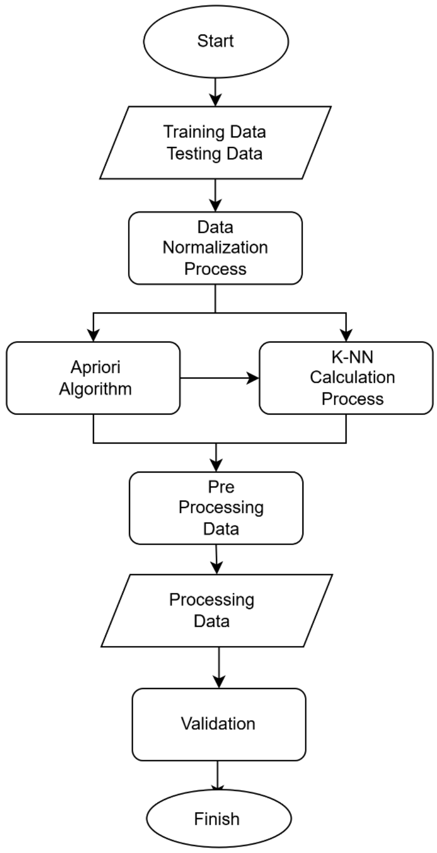

Figure 1. Research Steps

The research steps are shown in Fig. 1. It begins with the input of training data and testing data obtained from the analysis of learning method data collection needs. The input data will be normalized to minimize data redundancy and simplify the calculation process. The next process, using the Apriori algorithm, is used to find the ratio of the frequent item set in the given data set [22].

2.1 Apriori Algorithm

The Apriori Algorithm assumes that every nonempty subset of the concerned frequent set should be frequent. The Apriori algorithm was applied to the preprocessed categorical data to uncover frequent item sets and generate association rules. The following thresholds were used:

1. Minimum support : 0.05
2. Minimum confidence : 0.5
3. Lift threshold : ≥ 1.0

Sample rules extracted (top 3) in Table 1 below:

Table 1. Sample Rules Extracted

Antecedents	Consequents	Support	Confidence	Lift
{Digital Book, Grade: 6}	{Positive Impact}	0.22	0.85	1.20
{Lecture, Practical Task}	{CBL Model}	0.18	0.75	1.30
{Media: Interactive}	{E-learning Preferred}	0.25	0.88	1.45

These rules suggest strong relationships between learning attributes and preferences, helping educators understand students' learning behavior.

Apart from it, if an item is assumed to be infrequent, then it is highly likely that its superset must be infrequent as well [22], on Fig. 2 below.

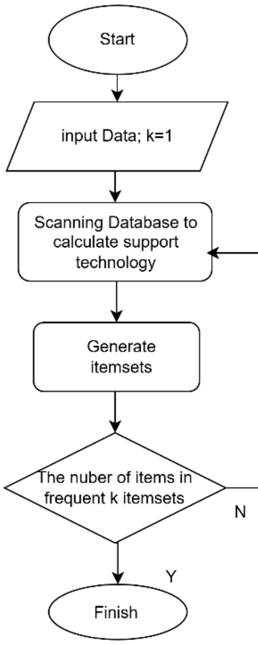

Figure 2. Flowchart of the Apriori Algorithm [22]

The Apriori Algorithm is used in association rule mining to discover frequent item sets in a transactional database. In Fig. 2, the process begins by inputting a dataset by looking for frequent item sets with 1 item. The next step scans the dataset to count how often each item appears, and it's also known as the support count. From the frequent 1-itemsets, the algorithm generates candidate item sets of size $K + 1$ using the Apriori property. The number of items in frequent K item sets are processed by counting the support for the new candidate item sets, pruning the item sets, and repeating the process. The process continues until no more frequent item sets can be generated. At the end, a list of all frequent item sets that meet the minimum support criteria.

2.2 K-Nearest Neighbor (KNN)

The KNN algorithm was used to classify student preferences based on learning related features.

1. K-Value Selecting and Validation
   a. The optimal value of k was determined through 5-fold cross-validation, testing k values from 1 to 15.
   b. The best classification performance was achieved at $k = 5$, which was then used for the final prediction.
2. Classification Performance Evaluation
   The model was evaluated using the test set (40%) with the following Table 2 metrics:

Table 2. Test Set Metrics

Metric	Value
Accuracy	84.29%
Precision	0.86
Recall	0.83
F1-Score	0.84

A confusion matrix was also generated to identify misclassification trends, and the implementation of algorithms and performance evaluation was conducted using Python, particularly the m1xtend and scikit-learn libraries [23].

The KNN calculation process is carried out, which begins by calculating the distance between the testing data and training data, then ranking the data based on the closest distance, and determining the value of the parameter k [24]. From the data that has been ranked based on the closest distance and the value of the parameter k, the prediction results of the appropriate learning method can be determined. The KNN algorithm is an algorithm that could produce good results and has been widely adopted for classification and regression problems [20].

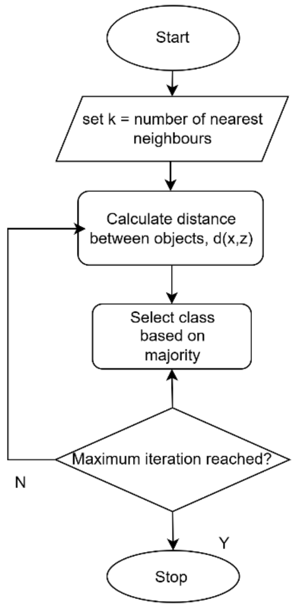

Figure 3. Flowchart of KNN Algorithm [20]

The flowchart in Fig. 3 explains that the KNN Algorithm begins the classification process. Then, the step begins by choosing the number of neighbors to consider. For a given dataset x, then compute the distance to the other point z. Selecting the class by looking at the k closest neighbors and take a majority vote on their class. For checking the value, a loop for validating or classifying multiple test values is needed. The process will end with the classification process if the value has been found.

2.3 Validation

At this stage, a frequency pattern analysis was conducted to identify item combinations that meet minimum support and confidence thresholds. To strengthen the findings, we computed confidence intervals for the top association rules and evaluated their statistical significance using lift values. For instance, rules with lift > 1.0 and narrow confidence interval (e.g., 95% CI: [0.80, 0.92]) indicate statistically significant and positively correlated preferences among students. This provides stronger empirical support for the conclusion that certain learning method combinations are more frequently and consistently preferred.

1. The combination, namely, X as a Data and Y as a different data from the X, and P is the probability for the data X or Y appear. The support value for each item is calculated using the following formula:

$$support\ (X \to Y) = P(X \cup Y) \tag{1}$$

$$support\ \left(\frac{X}{Y}\right) = \frac{\sum P(X \cup Y)}{N} \times 100\% \tag{2}$$

The support value of 2 items is obtained based on the formula below:

$$support\ (X \to Y) = P\left(\frac{X}{Y}\right) \tag{3}$$

$$support\ (X.Y) = \frac{\sum P(X \cap Y)}{N} \times 100\% \tag{4}$$

2. Once a high-frequency pattern is found, the next step is to find association rules that meet the minimum requirements for confidence. The formula for calculating confidence is as follows:

$$Confidence\ (X \to Y) = \frac{Support\ (X \cap Y)}{Support\ (X)} \times 100\% \tag{5}$$

3. After calculating the confidence value, the next step is to calculate the lift to understand the strength of the relationship between items X and Y.

$$Lift\ (X \to Y) = \frac{Confidence\ (X \to Y)}{Support\ (Y)} \tag{6}$$

Based on the procedures of KNN, the first step is to initialize the value of k, which is the number of nearest neighbors. The steps of the K-Means algorithm are as follows:

1. Determine k as the number of clusters you want to form.
2. Allocate objects into a cluster randomly.
3. Determine the cluster center (centroid) of the data in each cluster with the equation [25]:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \cdots + x_{nj}}{n}. \tag{7}$$

Where:

C_{kj} : the centers of k cluster on the variables of j ($j = 1,2, \dots, p$).
n : the sums of Data on the k cluster.

4. Determine the distance of each object to each centroid by using the Euclidean distance measure in the equation [25]:

$$d_{kj} = \sqrt{\sum_{j=1}^{n} \left(x_{ij} - v_{kj}\right)^2}. \tag{8}$$

Where:

d_{kj} : Euclidean distance k for variables of j
k : cluster index
x_{ij} : The value of i object in the cluster for the j variable
$\boldsymbol{v_{kj}}$: Centroid cluster k for variables of j

Then, according to the attribute on the dataset, there are two types of data, namely numeric and categorical. In the KNN process, the calculation can be carried out if the data is of a numeric type, so that categorical data is processed into numeric data by giving weights to the subcategories. Then normalize the data by using the min-max normalization formula shown in Eq. (9):

$$X^* = \frac{x - \min(X)}{\max(X) - \min(X)}. \tag{9}$$

Where x is the value before normalization, $\max(X)$ for the maximum value of the parameter, $\min\ (X)$ for the minimum value of the parameter, and is the result value that has been normalized. After the data is normalized, the data can then be calculated using the KNN algorithm.

The KNN algorithm is used to classify data into several classes based on existing attributes. The KNN algorithm works by calculating the distance of data to classify based on data that has the greatest similarity [24]. The distance calculation in KNN uses the Euclidean Distance formula, namely [26]:

$$d_i \sqrt{\sum_{i=i}^{k} (a_i - b_i)^2}. \tag{10}$$

Where a_i is the value on the training data, b_i is the value on the testing data, while k is the limit of the amount of data. For the accuracy value of the KNN algorithm, use the confusion matrix formula in Eq. (11):

$$accuracy = \frac{\sum correct\ test}{\sum Testing\ Data} \times 100\%. \tag{11}$$

3. RESULTS AND DISCUSSION

The KNN calculation process is carried out, which begins by calculating the distance between the testing data and training data, then ranking the data based on the closest distance, and determining the value of the parameter k. From the data that has been ranked based on the closest distance and the value of the parameter k, the prediction results of the appropriate learning method can be determined. The deeper analysis of the research implications of the findings provides a data-driven approach for educators to align teaching strategies with student preferences, and the combined model is especially useful in inclusive mathematics education.

3.1 Apriori Association Rule Analysis

The Apriori algorithm was applied to identify frequent item sets and association rules from the normalized dataset. The analysis produced several significant rules with lift values greater than 1, indicating strong positive correlations between selected features. In Table 3 given the sample association rules are given from the Apriori algorithm.

Table 3. Sample Association Rules from Apriori Algorithm

Rule	Antecedents	Consequent	Support	Confidence	Lift	Interpretation
R1	Digital Book + Interactive Media	E-Learning	0.25	0.88	1.45	Students who prefer digital books and media also tend to choose E-Learning.
R2	Project-Based Learning	Digital Method	0.18	0.75	1.30	Learners involved in PjBL often select digital methods.
R3	Lecturing+ Basic Theory Preference	Lecture Method	0.22	0.81	1.33	Students inclined toward basic theory and lectures prefer traditional method.

A lift > 1 indicates a positive dependence between the antecedents and the consequent, meaning the presence of certain elements increases the likelihood of a student selecting a particular learning method. These insights help educators understand which combinations of factors influence learning preferences.

3.2 KNN Classification Result and Analysis

Using the optimized $K = 5$, the KNN classifier predicted students' preferred learning methods. The model achieved a classification accuracy of 84.29%, with evaluation metrics shown in Table 2. A confusion matrix in Table 4 was used to analyze prediction accuracy and misclassifications.

Table 4. Confusion Matrix

Actual/Predicted	Digital Method	Lecture Method
Digital Method	87	15
Lecture Method	6	12

The confusion matrix in Table 4 shows the following:

1. Most misclassifications occurred when the actual method was digital, but was predicted as a lecture.
2. Only a small number of lecture method users were misclassified.
3. This shows class imbalance and that the lecture method is harder to predict due to fewer training samples.

3.3 Data Processing

In the evaluation, there are several values of k that has been tested for the KNN algorithm to get the best result. Training data and testing data come from the data obtained from the results of questionnaires filled out by students from various batches, totaling 150 data sets in Table 5 below:

Table 5. Attribute Description

	Name	ID	Grade	Reason	Course	Method	Impact	Factors	Models
1	ID444	24044	2	1	1	6	1	4,5,9	1,3
2	ID178	21078	8	6	2	6	1	2,8,10	2,3
16	ID407	24107	2	1	3	1,2,6	1	2,9	1
17	ID431	24031	2	3	1	6	1	4,5,10	2
18	ID131	21131	8	3	2	5	1	4,8,10	3
19	ID246	22046	6	4	3	3,5,7	1	4	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
149	ID433	24133	2	4	2	4	2	6,9	1,3
150	ID273	22073	6	1	1	2,5	2	9	2

***Data source**: from the dataset taken from questionnaires*

The attributes used are 9 attributes consisting of sub-attributes that have been given on the Table 6 where K1 is attribute name, K2 is Students Identity, K3 is Student's Grade, K4 is the reason why the students choose the subject, K5 is the Student's Course, K6 is the Method used for Learning Material, K7 is the Impact of Learning Method, K8 is Factors influence the Learning Method, and K9 is Models of Learning Method. Sub-attribute is a part of an attribute that explains the attribute in more detail. In this research, the sub-attribute is given a score or weight based on level of importance, while K1, K2, and K3 are given a zero score. Weights assigned to categorical sub-attributes are numerical encoding for pre-processing, enabling algorithms such as KNN to compute distance. These are not statistical weights unless otherwise specified in Table 6.

Table 6. Attribute and Sub Attribute

Code	Attribute Name	Sub-Attribute	Weight
K1	Name	-	0
K2	ID	-	0
K3	Grade	a. 1-2	0
		b. 3-4	0
		c. 5-6	0
		d. 7-8	0
		e. >8	0
K4	Reason	a. numerical	1
		b. Basic Theory	2
		c. Knowledge	3
		d. Method	4
		e. Media	5
		f. Practical	6
K5	Course	a. Counting	1
		b. Theory	2
		c. Practical	3
K6	Method	a. Presentation	1
		b. Lecturing	2
		c. Writing	3
		d. Digital book	4

Code	Attribute Name	Sub-Attribute	Weight
		e. E Learning	5
		f. Assignment	6
K7	Impact	a. Yes	1
		b. No	2
K8	Factors	a. Upgrade Teaching Materials	1
		b. Teaching aids	2
		c. Interactive Teaching Materials	3
		d. Examples and Practice	4
		e. Learning in Depth	5
		f. Many Tasks	6
		g. Team Work Task	7
		h. Project-Based Learning	8
		i. Interactive Media Learning	9
		j. Interaction of Lecturer	10
		k. others	11
K9	Models	a. Interactive Media Learning	1
		b. Models Learning SCL	2
		c. Models Learning Based Project (PjBL)	3
		d. Models Learning CBL	4
		e. Others	5

From Table 6, the results are obtained in the form of raw data, which will later be processed. From the raw data obtained, the data is divided into training data and test data, and the next step is to normalize it so that it can be grouped based on item sets to calculate support and lift. The training data and test data collection is done by Python with train-test-split, namely by dividing the data into 60% training data and 40% test data in Table 7 and Table 8 below:

Table 7. The Training Data

	K1	K2	K3	K4.1	K4.2	K4.3	K5	K6.1	K6.2	K6.3	K7	K8.1	K8.2	K9.1	K9.2	K9.3
29	ID4106	24106	2	4	0	0	1	1	2	0	1	1	0	1	2	3
22	ID1145	21145	7	2	4	5	1	3	4	6	1	2	0	1	3	4
51	ID2141	22141	6	6	0	0	2	1	0	0	1	1	0	4	0	0
75	ID4057	24057	2	2	0	0	1	4	6	0	2	1	0	0	0	0
11	ID2106	22106	6	4	5	5	2	6	0	0	1	1	0	4	7	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
71	ID4076	24076	2	2	*5*	*0*	*3*	*5*	*0*	*0*	*1*	*1*	*0*	*4*	*9*	*0*
106	ID2005	24005	2	2	3	6	1	3	6	0	1	3	0	8	9	10
14	ID1031	21031	7	1	0	0	1	3	0	0	1	1	0	3	0	0
92	ID1035	21035	6	2	4	5	2	5	6	0	1	1	3	2	4	7
102	ID4052	24052	2	1	0	0	2	7	0	0	1	1	0	7	0	0

Table 8. Test Data

	K1	K2	K3	K4.1	K4.2	K4.3	K5	K6.1	K6.2	K6.3	K7	K8.1	K8.2	K9.1	K9.2	K9.3
68	ID1093	21093	8	2	0	0	3	5	6		0	1	3	0	6	7
147	ID3055	23055	4	2	5	6	2	6	0		0	1	2	3	1	2
96	ID2103	22103	6	1	2	0	1	1	3		6	1	4	0	4	0

	K1	K2	K3	K4.1	K4.2	K4.3	K5	K6.1	K6.2	K6.3	K7	K8.1	K8.2	K9.1	K9.2	K9.3
82	ID4015	24015	2	2	0	0	1	6	0		0	1	2	0	10	0
135	ID3050	23050	4	1	0	0	1	7	0		0	1	3	0	7	0
139	ID3095	23095	4	4	5	0	2	6	0		0	1	3	0	3	4
26	ID4050	24050	2	6	0	0	3	8	0		0	1	5	0	4	0
12	ID4001	24001	2	6	0	0	1	6	0		0	1	2	0	3	4
18	ID4045	24045	2	3	0	0	1	4	6		0	2	3	0	11	0
15	ID2073	22073	6	6	0	0	2	3	0		0	2	0	0	0	0
134	ID3117	23117	4	1	0	0	1	7	0		0	1	2	0	4	0

Data normalization is used to transform the numeric values in a dataset to the same scale, without destroying the differences in values or losing information. This is important for a variety of reasons, including avoiding scale imbalances between features, improving the performance of machine learning models, reducing data complexity, improving data consistency, and adapting the algorithms that require the same scale. One of the main goals of normalization is to reduce data duplication [27]. The next step is to normalize the data by using the min-max normalization formula shown in Eq. (9):

$$X^* = \frac{x - \min{(X)}}{\max(X) - \min{(X)}}$$

Normalization is used to scale the data of an attribute so that it falls in a smaller range, such as -1.0 to 1.0 or 0.0 to 1.0. It is generally useful for classification algorithms. For every feature, the minimum value of that feature gets transformed into a 0, the maximum value gets transformed into a 1, and every other value gets transformed into a decimal between 0 and 1. Where X is the attribute data, $\min(X), \max(X)$ are the minimum and maximum absolute values of X respectively. X^* is the new value of each entry in data. x is the old value of each entry in the data. Data normalization using Python is given in the following Table 9 below:

Table 9. Data Normalization

	K4.1	K4.2	K4.3	K5	K6.1	K6.2	K6.3	K7	KS.1	KS.2	K9.1	K9.2	K9.3
0	0.0	0.000000	0.000000	0.25	0.714286	0.0	0.0	0.0	0.2	0.000000	0.333333	0.5	0.9
1	0.6	0.000000	0.000000	0.75	0.714286	0.0	0.0	0.0	0.6	0.000000	0.166667	0.8	1.0
2	0.2	0.500000	0.666667	0.25	0.000000	0.5	1.0	0.0	0.2	0.666667	0.166667	0.9	1.0
3	0.2	0.000000	0.000000	0.25	0.285714	1.0	0.0	0.0	0.6	0.000000	0.333333	0.6	1.0
4	0.4	0.000000	0.000000	0.75	0.714286	0.0	0.0	0.0	0.2	0.666667	0.083333	0.2	0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
147	0.2	0.833333	1.000000	0.50	0.714286	0.0	0.0	0.0	0.4	1.000000	0.083333	0.2	0.3
148	0.2	0.000000	0.000000	0.25	0.714286	0.0	0.0	0.0	0.2	0.000000	0.333333	0.0	0.0
149	0.0	0.000000	0.000000	0.25	0.571429	0.0	0.0	1.0	0.2	0.000000	0.416667	0.0	0.0
150	0.0	0.333333	0.500000	0.75	0.428571	0.0	0.0	0.0	0.4	0.000000	0.083333	0.5	0.0
151	0.2	0.666667	0.000000	1.00	0.714286	0.0	0.0	0.0	0.6	0.000000	0.083333	0.4	0.8

Support is a measure of the frequency of an item or set of items in a data set, while lift is a measure of how strong the relationship is between two items in an association rule. In Apriori, support is used to determine interesting association rules [15]. Association rules that have high support indicate that the items frequently appear together. Lift is a measure of how strong the relationship between two items is in an association rule [16]. It measures the increase in the probability of purchasing the second item if the first item is also purchased. A high lift indicated that there is a strong relationship between the two items, and that purchasing one item is likely to encourage purchasing the other. The association rules value from the Data is given in Table 10 below:

Table 10. Association Rules Value

	Antecedents	Consequents	Support	Confidence	Lift
0	(0.08333333333333333)	0.0	0.223684	1.000000	1.000000
1	(0.142857 142857 14285)	0.0	0.052632	1.000000	1.000000
2	(0.1666666666666666)	0.0	0.111842	1.000000	1.000000
3	(0.2)	0.0	0.644737	1.000000	1.000000
4	0.0	(0.2)	0.644737	0.644737	1.000000
⋮	⋮	⋮	⋮	⋮	⋮
1854	(0.2, 1.0, 0.08333333333333333)	(0.5, 0.0, 0.4)	0.065789	0.555556	2.558923
1855	(1.0, 0.4, 0.08333333333333333)	(0.5, 0.2, 0.0)	0.065789	0.555556	1.796690
1856	(0.2, 0.4, 0.08333333333333333)	(0.5, 0.0, 1.0)	0.065789	0.476 190	1.447619
1857	(0.5, 0.08333333333333333)	(0.2, 0.0, 1.0, 0.4)	0.065789	0.500000	2.375000
1858	(1.0, 0.08333333333333333)	(0.5, 0.2, 0.0, 0.4)	0.065789	0.476 190	3.446712

After the association rules are obtained, the next step is to calculate the Euclidean value. Euclidean distance is used to measure two points in one dimension, which gives results like the Pythagorean calculation [28]. Euclidean distance is used in algorithms such as K-Nearest Neighbors (KNN) to classify data based on similarity, or in image segmentation to determine pixel similarity. Euclidean distance value is given in Table 11 below:

Table 11. Euclidean Distance

No	Id	Euclidean Distance	Learning Method
1	ID4106	0.000000	Digital Learning Method
2	ID1022	1.597383	Digital Learning Method
3	ID2141	0.725335	Digital Learning Method
4	ID4057	1.387466	Digital Learning Method
5	ID2106	1.667561	Digital Learning Method
⋮	⋮	⋮	⋮
117	ID4076	1.478035	Digital Learning Method
118	ID4005	1.848338	Lecture Method
119	ID1031	0.842924	Digital Learning Method
120	ID1031	1.827377	Lecture Method
121	ID4052	1.283863	Digital Learning Method

Based on the result in Table 11 of the Euclidean approach, it was found that values above 1.7 use the lecture learning method, while values below 1.7 use the digital learning method. It has been calculated using the Euclidean approach in Python, a total of 102 students chose to use the digital learning method learning methods, and 19 chose to use the lecture method. In other words, in today's digital era, students also want a digital learning method, while some students want learning with the lecture method, but the number is not significant. From these results, the accuracy was obtained using Eq. (11):

$$accuracy = \frac{\sum correct\ test}{\sum Testing\ Data} \times 100\%$$

The accuracy of selecting the appropriate learning method for students in the most frequently chosen subject, namely numerical, is achieved by using a digital learning method as shown in Eq. (11):

$$accuracy = \frac{102}{121} \times 100\% = 84.29\%.$$

With the result of 84,29%, it shows that this result has an accurate value. In addition, the digital learning method can be combined with various learning models, such as using interactive media learning, models learning SCL, Project-Based Learning, CBL, or others.

3.4 Error Analysis and Overfitting Risk

To evaluate the risk of overfitting, we performed 5-fold cross-validation with varying k values in Table 12 below:

Table 12. Accuracy and K-Value

K-Value	Accuracy (%)
1	78.57
3	81.43
5	84.29
7	82.86
9	80.00
11	78.57
13	76.43
15	75.00

Accuray peaked at $k = 5$. Accuracy declines as k increases, indicating underfitting at high k values and a balanced performance at $k = 5$. That happened because cross-validation was applied, and the accuracy curve is smooth; we conclude the model does not exhibit overfitting, despite the small dataset.

3.5 Implication for Learning Method Selection

The implication for learning Method selection:

1. Digital learning methods are strongly associated with interactive content and independent models (e.g., SCL, PjBL)
2. The Apriori rules and KNN predictions together provide a data-driven approach for educators to align teaching strategies with student preferences.
3. The combined model is especially useful in inclusive mathematics education, where one-size-fits-all instruction often fails.

4. CONCLUSION

The study demonstrates that the integration of the Apriori algorithm and the K-Nearest Neighbor (KNN) classification method can effectively support data-driven decision-making in recommending appropriate learning methods. The Apriori algorithm successfully revealed significant association rules among student attributes, preferences, and selected learning strategies, with lift values exceeding 1.0, indicating meaningful relationships. Simultaneously, the KNN model, optimized at $k = 5$, achieved a classification accuracy of 84.29%, supporting its suitability for educational data classification, even with a relatively small dataset. These findings directly address the research objectives by showing that Apriori can identify patterns in student learning behavior that inform teaching method choices. KNN can predict the most appropriate learning methods based on those patterns with high accuracy. The combined approach provides a practical framework for educators to personalize instruction, particularly in inclusive mathematics learning. Future studies should validate the findings on a larger, more diverse dataset to enhance generalizability and explore integration with real-time recommendation systems in adaptive learning environments.

Author Contributions

Nuril Lutvi Azizah: Conceptualization, Supervision, Writing – Review & Editing. Ade Eviyanti: Methodology, Validation. Novia Ariyanti: Data Curation, Formal Analysis. Gita Wardani: Software, Investigation. Visualization, Project Administration. Naila Farah Diba: Project Administration, Visualization. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was funded by the Program RisetMu Batch VIII Tahun 2024–2025, Majelis Pendidikan Tinggi Penelitian dan Pengembangan Pimpinan Pusat Muhammadiyah-Hibah RISETMU Batch VIII Program 2024-2025, with contract number 0258.218/I.3/D/2025 for the funding and financial support.

Acknowledgment

With my gratitude to Allah SWT for His grace and guidance. We also extend our appreciation to Universitas Muhammadiyah Sidoarjo (UMSIDA) for the institutional support, facilities, and collaborative environment provided during the course of this study. Lastly, we thank all of the team and participating students who contributed valuable data through the questionnaire process.

Declarations

The authors declare no competing interests.

REFERENCES

[1] S. B. W. Sukoco, Badri Munir, Akhmaloka, *STRATEGI PENINGKATAN KUALITAS MENUJU PERGURUAN TINGGI BERKELAS DUNIA*. 2023. [Online]. Available: http://dinkes.sulselprov.go.id/page/download

[2] J. M. Sihombing, S. Syahrial, and U. S. Manurung, "KESULITAN PESERTA DIDIK DALAM PEMBELAJARAN MATEMATIKA MATERI PERKALIAN DAN PEMBAGIAN DI SEKOLAH DASAR," *J. Didakt. Pendidik. Dasar*, vol. 7, no. 3, pp. 1003–1016, 2023. doi: https://doi.org/10.26811/didaktika.v7i3.1177

[3] J. M. Adnan.K, Afandi F, "HUBUNGAN KEMAMPUAN LITERASI NUMERASI DENGAN HASIL BELAJAR MATEMATIKA SISWA KELAS V SD GUGUS II," pp. 423–430, 2021.

[4] M. Mariamah, S. Suciyati, and H. Hendrawan, "KEMAMPUAN NUMERASI SISWA SEKOLAH DASAR DITINJAU DARI JENIS KELAMIN," Tunas : J. Penelit. Pendidik. Dasar, vol. 1, no. 2, pp. 17–19, Dec. 2021. [Online]. Available: https://jurnal.fkip.unmul.ac.id/index.php/tunas/article/view/818

[5] U. Umar and A. Widodo, "ANALISIS FAKTOR PENYEBAB RENDAHNYA KEMAMPUAN AKADEMIK SISWA SEKOLAH DASAR DI DAERAH PINGGIRAN," *J. Educ. FKIP UNMA*, vol. 8, no. 2, pp. 458–465, 2022. doi: https://doi.org/10.31949/educatio.v8i2.2131

[6] F. T. P. Pangesti, "MENUMBUHKEMBANGKAN LITERASI NUMERASI PADA PEMBELAJARAN MATEMATIKA DENGAN SOAL HOTS," *Indones. Digit. J. Math. Educ.*, vol. 5, no. 9, pp. 566–575, 2018, [Online]. Available: http://idealmathedu.p4tkmatematika.org

[7] H. Nurhayati and N. W., Langlang Handayani, "JURNAL BASICEDU. JURNAL BASICEDU," *J. Basicedu*, vol. 5, no. 5, pp. 3(2), 524–532, 2020, [Online]. Available: https://journal.uii.ac.id/ajie/article/view/971

[8] D. S. Damayanti and P. I. Perdana, "PENGEMBANGAN E-MODUL PEMBELAJARAN TEMATIK (EMOTIK) BERBASIS FLIPBOOK PADA TEMA 8 SUBTEMA 1 KELAS V DI SEKOLAH DASAR," *J. Basicedu*, vol. 7, no. 5, pp. 2886–2897, 2023. doi: https://doi.org/10.31004/basicedu.v7i5.5932

[9] N. L. Azizah, V. Liansari, and A. I. Kusuma, "NUMERACY DEVELOPMENT TRAINING FOR ELEMENTARY SCHOOL STUDENTS AT SD MUHAMMADIYAH 2 WARU SIDOARJO," *Community Empower.*, vol. 8, no. 7, pp. 1033–1039, 2023. doi: https://doi.org/10.31603/ce.8781

[10] V. Liansari and N. L. Azizah, "THE RELATIONSHIP BETWEEN THE USE OF ACTIVE, INNOVATIVE, CREATIVE, AND FUN LEARNING TECHNIQUES AND ONLINE ENGLISH LEARNING BY MULTIDISCIPLINARY STUDENTS," *KnE Soc. Sci.*, vol. 2022, pp. 18–23, 2022, doi: https://doi.org/10.18502/kss.v7i10.11205

[11] A. M. Zen, A. Muhith, and S. S. Wahono, "DIGITAL FLIPBOOK-BASED LEARNING MEDIA DESIGN IN INTEGRATED THEMATIC LEARNING ON THE THEME OF GROWTH AND DEVELOPMENT OF LIVING THINGS IN CLASS III INCLUSIVE …," *Attadib J. …*, vol. 8, no. 1, 2024, [Online]. Available: https://www.jurnalfai-uikabogor.org/index.php/attadib/article/view/2700%0Ahttps://www.jurnalfai-uikabogor.org/index.php/attadib/article/download/2700/939. doi: https://doi.org/10.32507/attadib.v8i1.2700

[12] S. Silfia, "PENGEMBANGAN MEDIA PEMBELAJARAN FLIPBOOK DIGITAL BERBASIS LITERASI SAINS UNTUK SISWA KELAS IV SEKOLAH DASAR," no. July, pp. 1–23, 2020.

[13] R. Hanifan, T. D. Putra, and D. Hartanti, "IMPLEMENTASI ALGORITMA APRIORI UNTUK PENGELOMPOKKAN PRODUK TERBAIK PADA PANGKALAN SUDIAWATI," Komputa : J. Ilm. Komput. dan Informatika, vol. 11, no. 2, pp. 59–67, 2022. doi: https://doi.org/10.34010/komputa.v11i2.7363

[14] B. Hardiyanto and F. Rozi, "PREDIKSI PENJUALAN SEPATU MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *JOEICT(Jurnal Educ. Inf. Commun. Technol.*, vol. 04, no. 02, pp. 13–18, 2020.

[15] M. H. Santoso, "APPLICATION OF ASSOCIATION RULE METHOD USING APRIORI ALGORITHM TO FIND SALES PATTERNS CASE STUDY OF INDOMARET TANJUNG ANOM," *Brill. Res. Artif. Intell.*, vol. 1, no. 2, pp. 54–66, 2021. doi: https://doi.org/10.47709/brilliance.v1i2.1228

[16] A. N. Rahmi and Y. A. Mikola, "IMPLEMENTASI ALGORITMA APRIORI UNTUK MENENTUKAN POLA PEMBELIAN PADA CUSTOMER (STUDI KASUS : TOKO BAKOEL SEMBAKO)," *Inf. Syst. J.*, vol. 4, no. 1, pp. 14–19, 2021, [Online]. Available: https://jurnal.amikom.ac.id/index.php/infos/article/view/561. doi: https://doi.org/10.24076/infosjournal.2021v4i1.561

[17] A. Sanjaya and T. Wahyana, "PENERAPAN METODE K-NEAREST NEIGHBOUR UNTUK SISTEM PREDIKSI KELULUSAN SISWA MTS NURUL MUSLIMIN BERBASIS WEBSITE," *J. Transform. Mandalika*, vol. 3, no. 1, pp. 31–47, 2022, [Online]. Available: https://www.ojs.cahayamandalika.com/index.php/jtm/article/view/866%0Ahttps://www.ojs.cahayamandalika.com/index.php/jtm/article/download/866/863

[18] Y. Adriyani and R. Aulia, "PREDIKSI KELOMPOK UKT MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (THE PREDICTION UKT OF STUDENTS USING THE K-NEAREST NEIGHBOR ALGORITHM)," vol. 8, pp. 121–130, 2020. doi: https://doi.org/10.30595/juita.v8i1.6267

[19] M. N. Maskuri, K. Sukerti, and R. M. H. Bhakti, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN) UNTUK MEMPREDIKSI PENYAKIT STROKE," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 4, no. 1, pp. 130–140, Mei, 2022. doi: https://doi.org/10.46772/intech.v4i01.751

[20] N. M. Sabri and S. F. A. Hamrizan, "PREDICTION OF MUET RESULTS BASED ON K-NEAREST NEIGHBOUR ALGORITHM," *Ann. Emerg. Technol. Comput.*, vol. 7, no. 5, pp. 50–59, 2023 doi: https://doi.org/10.33166/AETiC.2023.05.005

[21] K. W. Mahardika, Y. A. Sari, and A. Arwan, "OPTIMASI K-NEAREST NEIGHBOUR MENGGUNAKAN PARTICLE SWARM OPTIMIZATION PADA SISTEM PAKAR UNTUK MONITORING PENGENDALIAN HAMA PADA TANAMAN JERUK," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 2, no. 9, pp. 3333–3344, 2018.

[22] Y. Zheng, P. Chen, B. Chen, D. Wei, and M. Wang, "APPLICATION OF APRIORI IMPROVEMENT ALGORITHM IN ASTHMA CASE DATA MINING," *J. Healthc. Eng.*, vol. 2021, 2021, doi: https://doi.org/10.1155/2021/9018408

[23] P. WiraBuana, S. Jannet D.R.M., and I. Ketut Gede Darma Putra, "COMBINATION OF K-NEAREST NEIGHBOR AND K-MEANS BASED ON TERM RE-WEIGHTING FOR CLASSIFY INDONESIAN NEWS," *Int. J. Comput. Appl.*, vol. 50, no. 11, pp. 37–42, 2012, doi: https://doi.org/10.5120/7817-1105

[24] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "IMPLEMENTATION OF K-NEAREST NEIGHBOR (K-NN) ALGORITHM FOR PUBLIC SENTIMENT ANALYSIS OF ONLINE LEARNING," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 2, p. 121, 2021, doi: https://doi.org/10.22146/ijccs.65176

[25] M. R. Zuhdi, H. Syihabuddin, A. Jauhar, A. Achmad, and R. Fernandes, "COMPARISON OF DBSCAN AND K-MEANS CLUSTER ANALYSIS WITH PATH- ANOVA IN CLUSTERING WASTE MANAGEMENT BEHAVIOUR PATTERNS PERBANDINGAN ANALISIS CLUSTER DBSCAN DAN K-MEANS DENGAN PATH-ANOVA DALAM PENGELOMPOKAN POLA PERILAKU PENGELOLAAN," vol. 6, no. 1, pp. 105–112, 2025. doi: https://doi.org/10.52436/1.jutif.2025.6.1.4183

[26] R. Sakti and A. Daulay, "ANALISIS KRITIS DAN PENGEMBANGAN ALGORITMA K-NEAREST NEIGHBOR (KNN): SEBUAH TINJAUAN LITERATUR," vol. 4, no. 2, pp. 131–141, 2024. doi: https://doi.org/10.47709/jpsk.v4i02.5055

[27] Z. Efendy, "NORMALISASI DALAM DESAIN DATABASE," *J. CoreIT*, vol. 4, no. 1, pp. 34–43, 2018.

[28] Y. Miftahuddin, S. Umaroh, and F. R. Karim, "PERBANDINGAN METODE PERHITUNGAN JARAK EUCLIDEAN , HAVERSINE , (STUDI KASUS : INSTITUT TEKNOLOGI NASIONAL BANDUNG)," vol. 14, no. 2, pp. 69–77, 2020. doi: https://doi.org/10.36787/jti.v14i2.270