BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0559–0570
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0559-0570

CLASSIFYING STUDENT'S DECISION MAKING ABILITY USING K-NEAREST NEIGHBOR FOR DETERMINING STUDENTS SUPPLEMENTARY LEARNING

Giyanti [1], Rina Oktaviyanthi [2*], Usep Sholahudin [3]

[1,2,3] *Department of Mathematics Education, Universitas Serang Raya*
Jl. Raya Serang-Cilegon Km. 5, Serang, Banten, Indonesia 42162

Corresponding author's e-mail: * ***rinaokta@unsera.ac.id***

Article History:

Received: 25th September 2022
Revised: 9th February 2023
Accepted: 18th February 2023

Keywords:

Classification algorithm;
Educational data mining;
k-nearest neighbor;
Mathematical decision making.

ABSTRACT

Mathematical decision-making ability is a complex cognitive process of finding problem solutions that is continuously explored and optimized for undergraduate students. The current research only focuses on the categorization on class score average into high, medium and low abilities. As a result, the lecturers do not have any standard categories to classify students' abilities as a reference in planning supplementary learning that could optimize undergraduate students' abilities. Therefore, the purpose of this study is to determine a classification model for undergraduate students' mathematical decision-making abilities that require supplementary learning based on the identification of the shortest distance of a new data from an existing data directory. The research method involved data mining techniques with the KNN classification model through the Knowledge Discovery in Database (KDD) process, starting from data selection, pre-processing, transformation, data mining and interpretation/evaluation [1]. *A total of 100 data were used as research samples which were divided into training data and testing data. Based on the test results, it is obtained that the accuracy of the classification model is 95% for the parameter value k = 15, meaning that each predicted testing data for the classification class is close to the actual condition with the number of neighbors 15 data from the training data.*

How to cite this article:

Giyanti, R. Oktaviyanthi and U. Sholahudin., "CLASSYFYING STUDENTS DECISION MAKING ABILITY USING K-NEAREST NEIGHBOR FOR DETERMINING STUDENTS SUPPLEMENTARY LEARNING," *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0559-0570, March, 2023.

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

Problem solving ability is one of the crucial skills that students should acquire at all levels of education. At the tertiary education, in each type of field study the undergraduate students must be able to implement the knowledge to solve the problems in the real context. In the problem solving process, students identify all possible ways that close to precise and accurate which will be chosen to be used as a solution [2], [3]. The identification of these possibilities according to [4] is an important part of the decision-making ability function that absolutely exists in every problem-solving process. Decision-making ability is a complex cognitive process of finding problems solutions that require skills in selecting, navigating and evaluating alternative of solving the problem [5], [6]. By studying the decision-making process, undergraduate students have an opportunity to train mental and instincts judgments in making choices and evaluating solution paths according to the situation as accurately as possible [7], [8]. Weaknesses in making decisions impact on incorrect results, inappropriate use of resources and time, and failed achieving targets as the goal [9], [10]. On the other hand, having the power to make decisions provides the opportunities to minimize failure, being closer to achieve the goals, optimize time and increase organizational productivity of resources [11], [12]. Therefore, developing and optimizing students' decision-making abilities becomes an important position in the learning process, especially for learning mathematics.

An important function of decision-making ability involves a variety of factors from undergraduate students that allow them to influence their decision-making abilities. The ideal decision-making process is related to the mental attitude of individuals in managing the information they receive to serve as rationalizations for determining the right alternative solutions from various aspects [13], [14]. As [15] mentioned this attitude is a derivative of personal control as an act of initiative to be consciously responsible for what he is facing. An attitude framework is known as self-directed learning which is defined by [16] as the process of individuals taking the initiative, with or without the help of others, in identifying learning needs, formulating learning objectives, selecting learning resources, using appropriate learning strategies and assessing the learning outcomes [17]. Self-directed learning has the potential to provide reinforcement for individuals in regulating and controlling the process of selecting problem solving in a rational, being directed and having responsible manner [18]. Therefore, the combination of undergraduate students' mathematical decision-making abilities and the self-directed learning is one of the efforts to synergize both skills.

To develop and optimize mathematical decision-making abilities, the lecturers should be able to track the level of mathematical decision-making abilities of their students. This information is essential in planning the appropriate actions for students based on their ability. The teachers need to have action planning in the learning activities to regulate the fulfillment of student achievement standards and controlling strategies in realizing the learning objectives [19]. However, the existing research only focuses on categorization on class score average into high, medium and low abilities. According to [20] this category is too general and need to be specified from various aspects. Moreover, grouping data by only standardizing it on the class average does not guarantee those students' latent abilities can be explored [21]. As the undergraduate students change every year, the reference for the category of decision-making ability also altered. As a result, the lecturers do not have any standard categories to classify students' abilities as a reference in planning learning and determining which students needs supplementary learning. Therefore, classifying students' decision-making abilities is an essential task.

The classification method is generally known as one of the techniques in the data mining to classify data based on data attachment to the data sample. The grouping process in data mining uses data references that have known groups or classes. Therefore, the data that does not have any group can be determined by a group through a comparison process with previously known data [22]. One of the algorithms in the classification technique that is commonly used is K-Nearest Neighbor (KNN). The KNN algorithm works based on the assumption that a data will have the same class or category as the data around [23]. The concept of the KNN algorithm is known as the neighboring concept, namely data that tend to be similar will be close to each other *(similarity metric)* [24]. The purpose of the KNN algorithm is to identify a number of k data objects called as training data that are closest to a given new data point which called as testing data), then assign a class label to the new data point with the highest number of votes [25] . The training data is used to train the algorithm in finding the appropriate model, while the testing data is used to test and determine the performance of the obtained model. Near or far neighbors are calculated through distance formulas including Euclidean Distance and Manhattan Distance. Data grouping using data mining algorithms according to [26] not only provides categorization information but can find hidden patterns from a set of data. In addition, classification with data mining techniques can facilitate analysis but with sufficient detailed information and time efficiency in determining decisions [27]. Thus, the KNN algorithm which is a classification technique in data mining can be used as a method of classifying student decision-making abilities as a reference for lecturers in determining supplementary learning.

The study is a collaborative research which involves fields of education, psychology and Artificial Intelligence [28]. The research was initiated by the field of psychology related to cognitive process theory and practically developed in the field of education and supported by the field of AI to design and identify a system to determine patterns and predictions. Research that raises the implementation of KNN in the field of education has been known as Education Data Mining (EDM) for approximately two decades and is still an interesting theme to study. Several studies regarding the implementation of the KNN algorithm focuses on the main topic of predicting student learning outcomes based on their participation in online learning platforms [29]. Another research discuss the use of predictive techniques in the provision of teachers in higher education institutions to facilitate distance learning [30]. Furthermore, KNN as a classification system to recommend high school students in engineering study programs [31]. In addition, the KNN technique in measuring the performance of students who take online learning through pedagogical video interactions [32]. Different research describe the role of KNN in predicting students' level of understanding and practice when dealing with virtual laboratories [33]. In mathematics education, especially the mathematical decision-making ability of students, the use of classification techniques using data mining is not widely used. Therefore, the results of this study are expected to contribute not only to the field of data mining but also to the field of mathematics education.

Based on the advantages and opportunities for optimizing undergraduate students' mathematical decision-making abilities and considering the implementation of the KNN algorithm as a research method, the focus of this research is to determine the classification model of students' mathematical decision-making abilities that require supplementary learning in mastering the theory and practice of mathematical concepts based on the identification of the closest distance to a new data from an existing data directory.

2. RESEARCH METHODS

The research method for determining the KNN classification model with this data mining technique refers to the Knowledge Discovery in Database (KDD) process starting from data selection, pre-processing, transformation, data mining and interpretation/evaluation [1]. The KDD process can be seen in the illustration below.

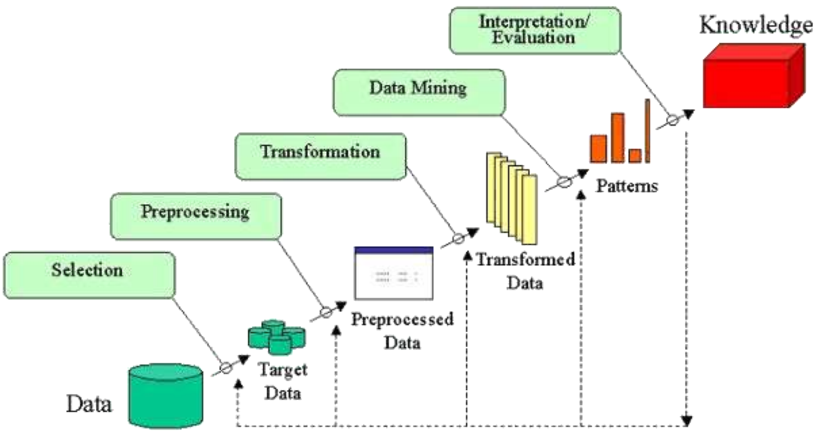

Figure 1. Illustration of the KDD Process [34]

2.1 Research Subject

The research subjects who participated in the study were 100 first-year undergraduate students who took the Calculus I course for the academic year 2020-2021 at Universitas Serang Raya. A number of research subjects were given an initial mathematical ability test containing diagnostic ability questions, a decision-making ability test containing mathematics problems from topic Limit Function and in accordance with the decision-making ability indicators. At last, the undergraduate students were asked to fill out a self-directed learning questionnaire which included statements related to management and self-control.

2.2 Research Instrument

The research instrument used to collect the data consists of three, namely the initial mathematical ability test, the decision-making ability test and the self-directed learning questionnaire. The initial mathematical ability test is consisting 15 diagnostic tests containing pre-calculus mathematics problems. Meanwhile, the mathematical decision-making ability test consist of 7 indicator questions related to the Limit Function material, and a self-directed learning questionnaire consisting of 22 statements divided into three self-directed learning indicators [35]. The three instruments were developed according to the research needs and based on the standard variable indicators. The results of developing instruments and validating process have been carried out previously and obtained a valid instrument; therefore, it can be used as a data collection tool [35].

2.3 Research Process

Furthermore, the data obtained is processed through a series of KDD activities as illustrated in **Figure 1**. At the data selection stage, which focuses on a subset of variables or data samples for the discovery process, the pre-processing stage is carried out to check for data inconsistencies and remove data duplication and select data as needed, the transformation stage focuses on the process of changing the data according to the data mining process and the algorithm used, the data mining stage covers the entire work of the algorithm in this study, and finally the interpretation/evaluation stage focuses on translating the patterns generated from data mining and checking the suitability of the patterns found with theory or facts in the field [36].

2.4 Data Mining Process

The data mining process in this study uses the KNN algorithm which has a workflow starting from (1) determining the k parameters (nearest neighbors), (2) calculating the distance of the testing data to all existing training data using the Euclidean distance formula, (3) sorting the distances used formed, and (4) determine the closest distance to the sequence k and then pair the corresponding classes [37]. The work of KNN algorithm data mining in this study is assisted by a data science platform, namely the RapidMiner application.

3. RESULTS AND DISCUSSION

3.1. Data Selection

The data used as a subset of variables in this study came from data on initial mathematical abilities, data on decision-making abilities, data on self-directed learning and data on student supplementary learning considerations. A total of 100 data from 100 research samples were obtained to be included in the data selection process. From the collected data, 80 data were used as training data and 20 data were used as testing data. The training data serves to train the algorithm to process the appropriate model and the testing data serves to check and evaluate the performance of the resulting model. Separation of training data and testing data to avoid overfitting is a training condition where the test results on the trained data are very good but tested by other data that are not used in training are very bad [38]. In addition, the data subset produced 9 attributes ranging from serial numbers to student supplementary learning considerations as compiled in the sample dataset in **Table 1**.

Table 1. Example of Student Ability Dataset

No.	Subject	Decision Making Ability (PK)		Early Mathematical Ability (KAM)		Self-directed Learning (SDL) Ability		Student Supplementary Learning Considerations Score
		Score	PK Category	Score	Category CAM	Score	SDL Category	
1	S1	83	Medium	92	High	51	Low	No Need
2	S2	100	High	92	High	59	Medium	No Need
3	S3	83	Medium	76	Medium	85	High	No Need
4	S4	67	Low	88	High	82	High	Need
5	S5	83	Medium	72	Medium	58	Medium	No Need
6	S6	50	Low	72	Medium	44	Low	Need
7	S7	83	Medium	64	Medium	85	High	No Need

No.	Subject	Decision Making Ability (PK)		Early Mathematical Ability (KAM)		Self-directed Learning (SDL) Ability		Student Supplementary Learning Considerations Score
		Score	PK Category	Score	Category CAM	Score	SDL Category	
8	S8	50	Low	64	Medium	90	High	Need
9	S9	50	Low	72	Medium	74	Medium	Need
10	S10	83	Medium	72	Medium	47	Low	No Need

3.2 Pre-processing

From the 9 attributes produced, the attributes that are only directly related to the analysis process are the attributes of the decision-making ability category, the category of early mathematical ability, the category of self-directed learning ability and consideration of student supplementary learning. Meanwhile, the attributes of serial number, subject identity and the score of ability were not used in the subsequent analysis.

Table 2. Example of Pre-processing Results of Student Ability Dataset

Decision Making Ability (PK)	Early Mathematical Ability (KAM)	Self-directed Learning (SDL) Ability	Student Supplementary Learning Considerations
Medium	High	Low	No Need
High	High	Medium	No Need
Medium	Medium	High	No Need
Low	High	High	Need
Medium	Medium	Medium	No Need
Low	Medium	Low	Need
Medium	Medium	High	No Need
Low	Medium	High	Need
Low	Medium	Medium	Need
Medium	Medium	Low	No Need

3.2. Transformation

Furthermore, the adjustment of the dataset with the algorithm used is K-Nearest Neighbor (KNN). The type of data that is processed in the KNN algorithm in this study is ordinal data. This is done as an effort to minimize data noise (the wrong data and outliers) [39]. For decision-making abilities, initial mathematical abilities and self-directed learning abilities, the data for the Low, Medium and High categories were converted into ordinal data into 1, 2 and 3. Meanwhile, for the consideration of supplementary learning, the category data needs and does not need to be changed into ordinal data into 1 and 2.

Table 3. Example of the Transformation Result Student Ability Dataset

Decision Making Ability (PK)	Early Mathematical Ability (KAM)	Self-directed Learning (SDL) Ability	Student Supplementary Learning Considerations
2	3	1	2
3	3	2	2
2	2	3	2
1	3	3	1
2	2	2	2
1	2	1	1
2	2	3	2
1	2	3	1
1	2	2	1
2	2	1	2

3.3. Data Mining

a. Determine the parameter k

The value of the parameter k chosen for experimentation in this study is k = 15 and k = 25. This k value indicates the number of nearest neighbors which is the basis for grouping the testing data.

b. Calculating the distance of testing data to all training data

The formula used in calculating the distance between the testing data and the training data is Euclidean Distance.

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \quad (1)$$

Information:

$dist(x_1, x_2)$ = distance between objects x_{1i} and x_{2i}

x_{1i} = Testing data

x_{2i} = Training data

n = Data dimension

The number of training data and testing data in this study were 80 and 20. In Table 4. the details of the distance from 5 testing data to 15 training data are shown as an example.

Table 4. Examples of Euclidean Distance 5 Testing Data on 15 Training Data

Subject	Decision Making Ability	Early Mathematical Ability	Self-directed Learning Ability	Supplementary Learning Consideration	Euclidean Distance				
					Test 1	Test 2	Test 3	Test 4	Test 5
S1	2	3	1	2	1	1.732051	2	1.732051	1
S2	3	3	2	2	2.44949	2.44949	1	2.236068	1.414214
S3	2	2	3	2	1.414214	1.414214	1	1.732051	0
S4	1	3	3	1	2.236068	1.732051	0	1.732051	1
S5	2	2	2	2	1.414214	2.44949	2.236068	2.236068	1.414214
S6	1	2	1	1	1	1	1.414214	1.414214	1
S7	2	2	3	2	1.732051	2.236068	1.414214	2.236068	1
S8	1	2	3	1	2.236068	1.732051	0	1.732051	1
S9	1	2	2	1	1.414214	1.414214	1	1.732051	0
S10	2	2	1	2	1.732051	2.236068	1.414214	2.236068	1
S11	3	2	2	2	2.236068	1.732051	0	1.732051	1
S12	3	2	2	2	0	1.414214	2.236068	1.414214	1.414214
S13	1	2	2	1	1	2.236068	2.44949	2	1.732051
S14	2	2	3	2	2.44949	2.44949	1	2.236068	1.414214
S15	2	2	1	2	2.236068	1.732051	0	1.732051	1

c. Sort the distance formed

Next, the distance of the testing data to the training data is sorted from the smallest distance value to the largest distance value. The following is an example of ordering the 1st testing data on 25 training data.

Table 5. Example of Ordering the 1st Testing Data from 25 Training Data

Subject	Decision Making Ability	Early Mathematical Ability	Self-directed Learning Ability	Supplementary Learning Consideration	Euclidean Distance Test 1
S12	3	2	2	2	0
S1	2	3	1	2	1
S6	1	2	1	1	1
S13	1	2	2	1	1
S18	1	2	3	1	1
S19	3	2	2	2	1
S3	2	2	3	2	1.414213562
S5	2	2	2	2	1.414213562
S9	1	2	2	1	1.414213562
S20	1	2	2	1	1.414213562
S7	2	2	3	2	1.732050808
S10	2	2	1	2	1.732050808
S16	1	2	2	1	1.732050808
S17	3	2	1	2	1.732050808
S4	1	3	3	1	2.236067977
S8	1	2	3	1	2.236067977
S11	3	2	2	2	2.236067977
S15	2	2	1	2	2.236067977
S2	3	3	2	2	2.449489743
S14	2	2	3	2	2.449489743
S21	3	2	2	2	3.741657387
S22	2	2	2	2	3.741657387
S23	1	2	2	1	3.741657387
S24	2	3	1	2	3.741657387
S25	1	1	3	1	3.741657387

d. Determine the closest distance to the order of k and pair the corresponding classes

The next step is to collect the smallest distance that has been sorted based on the value of the parameter k determined by ranking the distance according to the number of nearest neighbors. As an illustration, the 1st testing data grouping is presented using the parameter value $k = 15$.

Table 6. Grouping Testing Data 1 for Value $k = 15$

Sorting Distance Euclidean Test 1	Euclidean Distance Ranking Test 1	Student Supplementary Learning Considerations	Information	The Final Result
0	1	2		
1	2	2		
1	3	1		
1	4	1	Number of Labels 1 (Need) = 10 Number of Labels 2 (No Need) = 5	No Need
1	5	1		
1	6	2		
1.41421	7	2		
1.41421	8	2		
1.41421	9	1	Number of Labels 1 (Need) = 10 Number of Labels 2 (No Need) = 5	No Need
1.41421	10	1		
1.73205	11	2		

1.73205	12	2
1.73205	13	1

Sorting Distance Euclidean Test 1	Euclidean Distance Ranking Test 1	Student Supplementary Learning Considerations	Information	The Final Result
1.73205	14	2		
2.23606	15	1		

The same procedure was carried out for manual calculations of the 2nd testing data up to the 20th testing data on 80 training data, both for k = 15 and k = 25. After grouping the testing data based on neighboring values with two k parameters, the final results were obtained as presented in **Table 7** for k = 15.

Table 7. Test Results 20 Testing Data Determining Supplementary Learning with k = 15

No.	Subject	Student Supplementary Learning (System) Considerations		Student Supplementary Learning (Manual) Considerations k = 15		Test Result
1	S81	2	No Need	2	No Need	Correct
2	S82	1	Need	1	Need	Correct
3	S83	2	No Need	2	No Need	Correct
4	S84	2	No Need	2	No Need	Correct
5	S85	1	Need	1	Need	Correct
6	S86	2	No Need	2	No Need	Correct
7	S87	1	Need	1	Need	Correct
8	S88	2	No Need	2	No Need	Correct
9	S89	2	No Need	2	No Need	Correct
10	S90	1	Need	1	Need	Correct
11	S91	2	No Need	2	No Need	Correct
12	S92	2	No Need	2	No Need	Correct
13	S93	1	Need	1	Need	Correct
14	S94	2	No Need	1	Need	Wrong
15	S95	2	No Need	2	No Need	Correct
16	S96	1	Need	1	Need	Correct
17	S97	2	No Need	2	No Need	Correct
18	S98	1	Need	1	Need	Correct
19	S99	2	No Need	2	No Need	Correct
20	S100	1	Need	1	Need	Correct

e. Interpretation/Evaluation

Furthermore, the interpretation and evaluation of the classification model is carried out by calculating the level of model accuracy based on the parameter value k = 15 using the Confusion Matrix.

Table 8. Confusion Matrix k = 15

Actual Data	Classification Result Data	
	NO NEED	NEED
No Need	11	1
Need	0	8

Based on **Table 8** it is known that the True Positive (TP) or the number of positive data classified correctly by the system is recorded as many as 11 data. Then True Negative (TN) or the number of negative data classified correctly by the system is recorded as many as 8 data. While False Negative (FN) or the number of positive data but predicted to be negative by the system recorded 1 data and False Positive (FP) or the number of negative data but predicted positive by the system recorded 0 data. From the results of the Confusion Matrix, the accuracy of the classification model performance can be calculated using the formula below.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (2)$$

$$= \frac{11+8}{11+8+0+1} \times 100\% = \frac{19}{20} = 0.95 \times 100\% = 95\%$$

Based on the calculation of the accuracy of the classification model performance obtained a value of 95%. The accuracy describes the degree of closeness between the obtained value and the actual value. The higher the percentage of accuracy, the resulting classification model is closer to the actual condition. The same procedure, Confusion Matrix for parameter value k = 25 as presented in **Table 9**.

Table 9. Confusion Matrix k = 25

Actual Data	Classification Result Data	
	No Need	**Need**
No Need	7	1
Need	4	8

Referring to Table 9, it can be calculated the accuracy value of the classification model performance for k = 25 which is 75%. This value indicates that the classification model with the number of closest neighbors 25 can approach the actual condition by 75%.

3.4. Discussion

Based on the results of the experiments, the results of a classification model of students' mathematical decision-making abilities were obtained as the basis for determining the provision of supplementary learning in mastering the material both in theory and practice. The classification model built uses the KNN algorithm by experimenting with two k parameters, namely k = 15 and k = 25. Evaluation of the classification model using the Confusion Matrix shows that the accuracy level of the classification model performance is closest to the actual condition, which is 95%, when using the parameter k = 15. This means that each test data will have a predictable class classification with the number of neighbors 15 data from the training data. On this basis, it can be said that the test data located close to the 15 training data have similarities or proximity to each other. So that the test data for which the criteria are not known can finally be determined from the criteria for the 15 training data that appear the most. The KNN algorithm can process mathematical-based data to evaluate the value of existing criteria into a classification description **[40]**. The same thought was, the classification of KNN is based on unlabeled observations (test data) by assigning them to the most similar labeled sample class (training data) **[41]**.

The data processed in this study, later called training data, are data on decision-making abilities, initial mathematical abilities and self-directed learning to be a consideration for determining the lecturers in providing supplementary learning to students. The training data has been tested through the KNN procedure with an accuracy of 95% for k = 15. The implication of the test results is that lecturers can see which data have similarities for further processing, not only in the context of giving additional classes but also analyzing the cognitive abilities of students in their groups or based on their neighbors. This is conveyed implicitly by **[3]** which states that the results of the KNN algorithm classification are not only limited to attaching labels to the test data according to the class in the training data but also explore other criteria from previously unknown test data. This fact, according to **[42]** can provide an opportunity to find a certain pattern or regularity from the data and ultimately become the basis of a prediction.

The results obtained in this study provide the lecturers with an alternative way to consider giving learning treatment to groups of students who have been classified with the KNN algorithm. Since the performance of the classification model on the very good level, then it can be used as a framework for lecturers in classifying the students. Moreover, the classification model obtained could be a reference in predicting unknown classes as seen at **Table 10**. The **Table 10** illustrates the test data with scores and categories of decision-making abilities, early mathematical abilities and self-directed learning. Therefore, the lecturers could predict whether undergraduate students with the criteria in **Table 10** require supplementary learning or not by employing training data in a classification model using the KNN algorithm which already has very good model performance accuracy.

Table 10. New Test Data

Subject	Decision Making Ability (PK)		Early Mathematical Ability (KAM)		Self-directed Learning (SDL) Ability		Student Supplementary Learning Considerations
	Score	PK Category	Score	Category CAM	Score	SDL Category	
S1	67	Low	64	Medium	59	Medium	?

In its development, the formed classification model can also provide information about the hidden criteria of one undergraduate student as well as know the similarity of the undergraduate students' criteria with at least 15 other students who are neighbors in calculating KNN data. So that lecturers can be more flexible in taking action to improve the undergraduate students' abilities both from cognitive and non-cognitive aspects based on their groups. As [43] said that it could provide the opportunities for the lecturers to optimize group-based learning activities where students share knowledge with their colleagues. Additionally, [44] emphasized that such a learning approach can increase the students' motivational learning, the quality and the learning process to the social interaction of students during the learning process.

4. CONCLUSIONS

Regarding to the series of Knowledge Discovery in Database (KDD) activities, particularly in the data mining process, it is known the classification model to determine whether or not students obtain supplementary learning achieves 95% model performance accuracy for $k = 15$. Therefore, it can be concluded that the parameter $k = 15$ is the number of student previous data neighbors (training data) which is closest to and able to explain the actual condition of the data new student (test data). In addition, the classification model in this study can be used as a reference for lecturers in projecting unknown conclusions as illustrated in Table 10. However, the classification model using the KNN algorithm has not been able to explain a pattern. The resulting classification model with excellent performance accuracy can be a recommendation for the lecturers in classifying students according to the similarity of criteria. However, it is not the only way for the lecturers to make decisions in determining the planning of learning activities. Moreover, the results of this study become a lecturer's tool to optimize the undergraduate students' abilities in the learning process from each aspect that can be developed. Several recommendations for further research to increase the capacity and value of previous research results, including (1) conducting experiments using other classification algorithms KNN which have the opportunity to produce optimal accuracy values, and (2) conducting experiments by collaborating two algorithms so that the resulting classification model can work to predict and determine the pattern.

ACKNOWLEDGMENT

We would like to thank Kemdikbud for funding this research through the "Penelitian Dosen Pemula' scheme program.

REFERENCES

[1] M. Rogalewicz and R. Sika, "Methodologies of knowledge discovery from data and data mining methods in mechanical engineering," *Manag. Prod. Eng. Rev.*, vol. 7, no. 4, pp. 97–108, Dec. 2016.

[2] J. Orzechowski, E. Kruchowska, A. Gruszka, and B. Szymura, "Understanding factors behind the effectiveness of personal identification: Revolution – a new technique of creative problem solving," *Think. Ski. Creat.*, vol. 23, pp. 140–149, Mar. 2017.

[3] Y. Zhang, G. Cao, B. Wang, and X. Li, "A novel ensemble method for k-nearest neighbor," *Pattern Recognit.*, vol. 85, pp. 13–25, Jan. 2019.

[4] D. D. Sheu and J. Hong, "Prioritized relevant effect identification for problem solving based on similarity measures," *Expert Syst. Appl.*, vol. 100, pp. 211–223, Jun. 2018.

[5] L. Qi and C. Gonzalez, "Math matters: mathematical knowledge plays an essential role in Chinese undergraduates' stock-

and-flow task performance," *Syst. Dyn. Rev.*, vol. 35, no. 3, pp. 208–231, Jul. 2019.
[6] K. Skagerlund, M. Forsblad, G. Tinghög, and D. Västfjäll, "Decision-making competence and cognitive abilities: Which abilities matter?," *J. Behav. Decis. Mak.*, vol. 35, no. 1, Jan. 2022.
[7] I. Dasgupta and S. J. Gershman, "Memory as a computational resource," 2020.
[8] M. Witt, F. Lewis, and B. Knight, "'I wasn't worried because I wasn't being judged': The development of pre-service teacher professional capital, pedagogical instinct and discretionary judgement during an overseas teaching placement," *Teach. Teach. Educ.*, vol. 112, p. 103642, Apr. 2022.
[9] M. Chegeni, P. M. Shahrbabaki, M. E. Shahrbabaki, N. Nakhaee, and A. Haghdoost, "Why people are becoming addicted to social media: A qualitative study," 2021.
[10] M. E. David, F. R. David, and F. R. David, "Closing the Gap between Graduates' Skills and Employers' Requirements: A Focus on the Strategic Management Capstone Business Course," 2021.
[11] Y. H. Yuan, C. H. Liu, and S. S. Kuang, "An Innovative and Interactive Teaching Model for Cultivating Talent's Digital Literacy in Decision Making, Sustainability, and Computational Thinking," *Sustain. 2021, Vol. 13, Page 5117*, vol. 13, no. 9, p. 5117, May 2021.
[12] J. Mitton and A. Murray-Orr, "Exploring the connection between playfulness and learning: Making learning memorable in a culturally and economically diverse grade 5 classroom," *Think. Ski. Creat.*, vol. 43, p. 101005, Mar. 2022.
[13] A. K. Sahu, R. K. Padhy, and A. Dhir, "Envisioning the future of behavioral decision-making: A systematic literature review of behavioral reasoning theory," *Australas. Mark. J.*, vol. 28, no. 4, pp. 145–159, Nov. 2020.
[14] P. Xu and J. Cheng, "Individual differences in social distancing and mask-wearing in the pandemic of COVID-19: The role of need for cognition, self-control and risk attitude," *Pers. Individ. Dif.*, vol. 175, p. 110706, Jun. 2021.
[15] K. D. Vohs, R. F. Baumeister, B. J. Schmeichel, J. M. Twenge, N. M. Nelson, and D. M. Tice, "Making choices impairs subsequent self-control : A limited-resource account of decision making, self-regulation, and active initiative," *Self-Regulation and Self-Control*, pp. 45–77, Jan. 2018.
[16] J. L. van der Walt, "The Term 'Self-Directed Learning'—Back to Knowles, or Another Way to Forge Ahead?," *https://doi.org/10.1080/10656219.2019.1593265*, vol. 28, no. 1, pp. 1–20, Jan. 2019.
[17] R. Haworth, "Personal Learning Environments: A Solution for Self-Directed Learners," *TechTrends 2016 604*, vol. 60, no. 4, pp. 359–364, Apr. 2016.
[18] T. H. Morris and M. Rohs, "The potential for digital technology to support self-directed learning in formal education of children: a scoping review," *https://doi.org/10.1080/10494820.2020.1870501*, 2021.
[19] R. G. Brockett and R. Hiemstra, "Self-direction in adult learning: Perspectives on theory, research and practice," *Self-Direction Adult Learn. Perspect. Theory, Res. Pract.*, vol. 2, pp. 1–276, Jan. 2018.
[20] K. Karatas and I. Arpaci, "The Role of Self-directed Learning, Metacognition, and 21st Century Skills Predicting the Readiness for Online Learning," *Contemp. Educ. Technol.*, vol. 13, no. 3, p. ep300, Mar. 2021.
[21] C. Bratianu, G. Prelipcean, and R. Bejinaru, "Exploring the latent variables which support SMEs to become learning organizations," *Manag. Mark.*, vol. 15, no. 2, pp. 154–171, Jun. 2020.
[22] C. Jalota and R. Agrawal, "Analysis of Educational Data Mining using Classification," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 243–247, Feb. 2019.
[23] F. Alshareef, H. Alhakami, T. Alsubait, and A. Baz, "Educational Data Mining Applications and Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 729–734, Jun. 2020.
[24] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, p. e01250, Feb. 2019.
[25] G. Ramaswami, T. Susnjak, A. Mathrani, J. Lim, and P. Garcia, "Using educational data mining techniques to increase the prediction accuracy of student academic performance," *Inf. Learn. Sci.*, vol. 120, no. 7–8, pp. 451–467, Sep. 2019.
[26] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, pp. 1–19, Dec. 2022.
[27] S. Shrestha and M. Pokharel, "Educational data mining in moodle data," *Int. J. Informatics Commun. Technol.*, vol. 10, no. 1, pp. 9–18, Apr. 2021.
[28] T. Jülicher, "Education 2.0: Learning Analytics, Educational Data Mining and Co.," pp. 47–53, 2018.
[29] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in Educational Data Mining for Prediction Tasks," *Appl. Sci. 2020, Vol. 10, Page 90*, vol. 10, no. 1, p. 90, Dec. 2019.
[30] C. Herodotou, B. Rienties, · Avinash Boroowa, Z. Zdrahal, and · Martin Hlosta, "A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective," *Educ. Technol. Res. Dev.*, vol. 67, pp. 1273–1306, 2019.
[31] M. Garanayak, S. Sahoo, S. N. Mohanty, and A. K. Jagadev, "An Automated Recommender System for Educational Institute in India," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. "7," no. 26, pp. 1–13, Feb. 2020.
[32] P. M. Moreno-Marcos, T. C. Pong, P. J. Munoz-Merino, and C. D. Kloos, "Analysis of the Factors Influencing Learners' Performance Prediction with Learning Analytics," *IEEE Access*, vol. 8, pp. 5264–5282, 2020.
[33] T. Hristova, K. Gabrovska-Evstatieva, and B. Evstatiev, "Prediction of engineering students' virtual lab understanding and implementation rates using SVM classification," *J. e-Learning Knowl. Soc.*, vol. 17, no. 1, pp. 62–71, Jul. 2021.
[34] U. M. Fayyad, "The Secrets of Data Science Deployments," *IEEE Intell. Syst.*, vol. 37, no. 4, pp. 30–34, Jul. 2022.
[35] R. Oktaviyanthi and R. N. Agus, "INSTRUMEN EVALUASI KEMAMPUAN PENALARAN ADAPTIF MATEMATIS MAHASISWA," *AKSIOMA J. Progr. Stud. Pendidik. Mat.*, vol. 9, no. 4, pp. 1123–1136, Dec. 2020.
[36] I. Triguero, D. García-Gil, J. Maillo, J. Luengo, S. García, and F. Herrera, "Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 2, p. e1289, Mar. 2019.
[37] Y. Sun, Z. Li, X. Li, and J. Zhang, "Classifier Selection and Ensemble Model for Multi-class Imbalance Learning in Education Grants Prediction," *https://doi.org/10.1080/08839514.2021.1877481*, vol. 35, no. 4, pp. 290–303, 2021.
[38] E. M. Dos Santos, R. Sabourin, and P. Maupin, "Overfitting cautious selection of classifier ensembles with genetic algorithms," *Inf. Fusion*, vol. 10, no. 2, pp. 150–162, Apr. 2009.
[39] M. Tang, R. Pérez-Fernández, and B. De Baets, "Fusing absolute and relative information for augmenting the method of

nearest neighbors for ordinal classification," *Inf. Fusion*, vol. 56, pp. 128–140, Apr. 2020.
[40] K. K. Sharma and A. Seal, "Spectral embedded generalized mean based k-nearest neighbors clustering with S-distance," *Expert Syst. Appl.*, vol. 169, p. 114326, May 2021.
[41] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, Jun. 2016.
[42] J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, and H. Yang, "A generalized mean distance-based k-nearest neighbor classifier," *Expert Syst. Appl.*, vol. 115, pp. 356–372, Jan. 2019.
[43] E. O. Kimbrough, A. D. McGee, and H. Shigeoka, "How Do Peers Impact Learning? An Experimental Investigation of Peer-to-Peer Teaching and Ability Tracking," *J. Hum. Resour.*, vol. 57, no. 1, pp. 304–339, Jan. 2022.
[44] M. Stigmar, "Peer-to-peer Teaching in Higher Education: A Critical Literature Review," *http://dx.doi.org/10.1080/13611267.2016.1178963*, vol. 24, no. 2, pp. 124–136, Mar. 2016.