Terbit *online* pada laman web jurnal : http://journal.unbara.ac.id/index.php/INTECH

INFORMATIKA DAN TEKNOLOGI (INTECH)

ISSN (Online) : 2722-7367

Learning Level Prediction System Based on Student Academic Ability By Looking at the Highest Similarity Value

Abdul Rahman [1] Pujianto [2]

[1,2] *Faculty of Engineering and Computer, Baturaja University, Indonesia*
[1] *abdulrahman@ft.unbara.ac.id* [2] *pujianto@ft.unbara.ac.id*

ARTICLE INFORMATION

Accepted Editor : 20 Februari 2022
Final Revision : 29 April 2022
Published Online : 29 Mei 2022

KEYWORDS

Students, Case Base Reasoning, Data Mining, Nearest Neigbor, Learning, Academic, Confusion Matrix.

ABSTRACT

There is a diversity of academic abilities of students in each Indonesian educational institution, each student has different learning abilities according to the level of learning. Educators in providing learning level predictions are done manually, so it takes a long time in predicting student learning levels. In this study, it was able to predict students' academic abilities in an easy and fast way. The method used in predicting learning levels is Case Based Reasoning. This method is able to predict the student's learning level to be (1) Very Bad, (2) Bad, (3) Average, (4) Good, and (5) Excelent. This level of learning will be used as a benchmark for educators to provide appropriate values to students. The results of this study for the academic performance of the very poor category are 0 students, the bad category is 3 students, the medium category is 79 students, the good category is 16 students and the excellent category is 12 students. The accuracy of academic performance recommendations using confusion matrix is 91.82%.

I. INTRODUCTION

Education is now very much determined by the learning performance of the younger generation, this is certainly dominated by students who are currently pursuing education. To improve student learning performance, there are several factors that influence it, including intellectual factors, the ability to learn and personality factors of each student [1].

Student performance in tertiary institutions is determined by several variables such as test scores, classroom interactions, practicum, attendance, participation in extra-curricular activities. When students are in a class that is not in accordance with academic abilities, students will find it difficult to participate in learning. Therefore it is necessary to analyze the results of student learning performance to help them keep learning and not drop out of college [1].

Several studies have been conducted to predict academic performance in tertiary education. Most researchers only use average data from previous semester programs, entry exam notes, work experience, age, gender, etc. [2]. Many studies that explain that there are many factors that are considered as the influence of academic performance [1].

In previous studies [1] obtained the eight best factors to predict student academic performance. These factors consist of Personal Data (PD), Study Strategies (SS), Belief in studying (BS) and Cognitive Skills (CS). And provides predictions of student performance which is divided into five categories namely Very Bad, Bad, Average, Good and Excelent.

Table. 1 Optimal selection of factors using Genetic Algorithms and Neural Networks

No	Factor
1	[PD] Age
2	[PD] Gender
3	[SS] Emphasize main ideas
4	[SS] Group Study
5	[SS] Take notes
6	[BS] Hard work
7	[CS] Mid-term grades
8	[CS] Finals grades

Table. 2 Rule Base Performa Akademic

No	Performa Academic				
	VeryBad	Bad	Average	Good	Excelent
1	>26	24 to 26	21 - 23	18 to 20	<= 18

No	Performa Academic				
	VeryBad	**Bad**	**Average**	**Good**	**Excelent**
2	-	-	-	Male	Female
3	Never	Rarely	Sometimes	Often	Always
4	Never	Rarely	Sometimes	Often	Always
5	Never	Rarely	Sometimes	Often	Always
6	Never	Rarely	Sometimes	Often	Always
7	0 - 25	25 - 50	50 - 75	75 - 90	90 - 100
8	0 - 25	25 - 50	50 - 75	75 - 90	90 - 100

From the data above the five predictions will be used as a base case or rule base of the past to provide recommendations on the performance of the latest student data. The results of these recommendations will be used as data to make groupings of students in a heterogeneous class. Case-based reasoning using the Nearest Neighbor algorithm will make it easier for teachers to solve this.

II. LITERATUR RIVIEW

A. *Case Based Reasoning (CBR)*

Case-based reasoning (CBR), is a paradigm for solving problems by utilizing the knowledge of past cases to solve new cases. Past cases show situations that were previously experienced and that have been stored and studied, so that when there are new cases can be resolved with experience of past cases that have been stored [3].

Past cases are stored on a case basis, and are characterized from three aspects: 1) A description of the problem, which describes the situation when the case occurred; 2) Solution to the problem, which states the solution to that problem; 3) Results, which describe the state after the case occurred.

Based on previous cases, a new case is resolved in the following four steps: 1) Take the most similar past case. 2) Recommended solutions for new cases by reusing information and knowledge in the most similar last case. 3) Revise the proposed solution. 4) Save information and knowledge about solutions for new cases.

In this research, case-based reasoning is needed to recommend the academic performance of students included Very bad, Bad, Average, Good or Excelent categories. Later the results of learning performance recommendations will be used as data to form classes based on the merging of all categories that have been made as rule based.

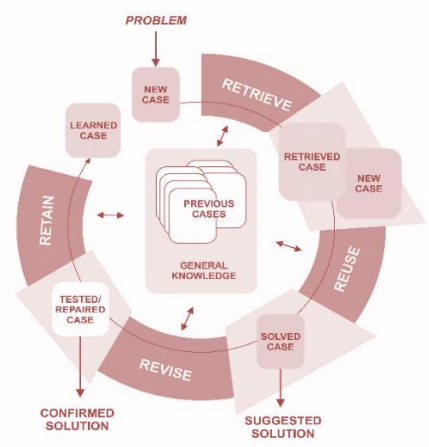


Figure 1 Case Based Reasoning [4]

B. *Nearest Neighbor Algorrithm*

Nearest Neighbor (NN) is one of the most popular algorithms and is included in 10 data mining methods. This is because the nearest neighbor algorithm is very simple in its implementation. How the nearest neighbors work calculates the proximity of each data and then selects the nearest neighbor.

The Nearest Neighbor algorithm is often combined with the CBR method in the diagnostic process, decisions and recommendations [5] [6]. According to [7] nearest neighbor algorithm is an approach to look for cases with the closeness between new cases with old cases, which is based on the weight matching of a number of existing features.

This method looks for distances to the destination of data that has been stored previously. After the distance is obtained then the closest distance is sought. The closest distance is used to find the identity of the destination.

The formula used in the calculation of proximity (similarity) as in the formula or formula 1 :

$$S(P,C) = \frac{(s1 * w1) + (s2 * w2) + \cdots + (sn * wn)}{w1 + w2 + \cdots + wn}$$

S = *Similarity* (Nilai kemiripan), W = *Weight* (Weight given), P = Problem, C = Case.

III. RESEARCH METHOD

In this study the method used is Case Based Reasoning (CBR) to recommend student academic performance, and to find the similarity value of the case by using the Nearest Neighbor algorithm.

Broadly speaking, the way CBR works with the Nearest Neighbor algorithm to recommend the results of student academic performance is as in Figure 3.

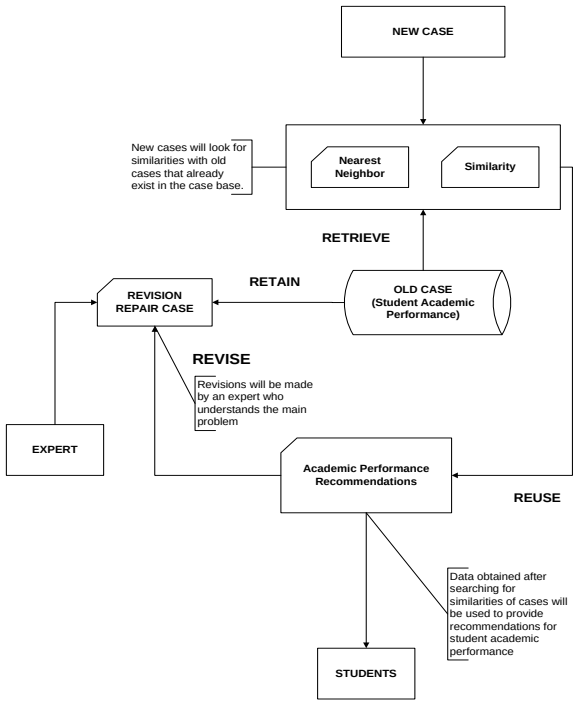


Figure 2 How CBR Works Using the Nearest Neighbor Algorithm

In Figure 3 it is explained that old cases are stored in a knowledge base. All old cases will be used as a knowledge base which will later become a reference for finding similarities in new cases owned by students. This new case was obtained from students' data while studying in the previous semester.

A. *Case Base Representation*

The collection of features used as a case base and recommendations for academic performance were obtained from previous research [1]. To predict the academic performance of these students there are five 8 factors and 5 predictions of student academic performance that are used as a case base. These factors consist of Personal Data (PD), Study Strategies (SS), Belief in studying (BS) and Cognitive Skills (CS). For PD the data are Age (A) and Gendre (G), for SS the data are Emphasize main ideas (EMI), Group Study (GS) and Take Notes (TN), for BS the data is Hard Work (HW), for CS the data are Mid-term Grades (MG) and Finals Grades (FG). To provide predictions of student performance divided into five categories: Very Bad (VB), Bad (B), Average (A), Good (G) and Excelent (E) [8], each of which has criteria as in Table 3.

Table. 3 Case Based Representation

Factor	Performa Academic				
	VB	B	A	G	E
[PD] A	>26	24 to 26	21 - 23	18 to 20	<= 18
[PD] G	-	-	-	M	F
[SS] EMI	N	R	S	Often	Als
[SS] GS	N	R	S	Often	Als
[SS] TN	N	R	S	Often	S
[BS] HW	N	R	S	Often	Als
[CS] MG	0 - 25	25 - 50	50 - 75	75 - 90	90 - 100
[CS] FG	0 - 25	25 - 50	50 - 75	75 - 90	90 - 100

N = Never, R = Rarely, S = Sometimes, O = Often, M = Male, F = Female, Als = Always.

B. *Student Academic Data*

Data to be grouped based on the results of student learning performance recommendations is the data of the previous semester, as in Table 4.

Table. 4 Student Academic Data

NO	A	G	EMI	GS	TN	HW	MG	FG
1	21	Male	S	R	S	Als	72	85
2	20	Male	Als	S	Als	Als	40	76
3	19	Male	R	R	S	Als	68	84
4	18	Male	S	S	S	Als	72	85
5	20	Male	R	S	Als	Als	64	83
6	25	Female	Als	Als	Als	Als	36	74
7	20	Male	S	N	Als	N	68	84
8	20	Male	Als	Als	Als	Als	40	76
9	19	Male	S	S	Als	S	72	85
10	19	Male	R	S	Als	S	72	85
11	19	Female	S	Als	N	Als	72	85
12	19	Male	Als	Als	Als	S	44	77
13	19	Male	S	S	S	R	72	85
14	20	Male	Als	Als	R	N	52	79
15	20	Male	S	S	S	Als	68	84
16	20	Male	R	S	S	Als	40	76
17	20	Male	S	Als	Als	Als	68	84
18	19	Male	Als	Als	N	Als	40	76
...	…	…	…	…	…	…	…	…
110	19	Male	N	S	Als	R	44	82

C. *Retrieval*

In this process the level of similarity of students academic performance will be sought based on the results of previous students' academic studies. This process is done by calculating the value of the similarity of new cases in this case represented by academic data of students in the following semester on the basis of existing knowledge cases. The highest similarity value will be used as a recommendation for students' academic performance. The flow of this process can be seen in Figure 3.

D. *Calculation of Case Similarity Value*

To calculate the similarity of cases owned by students with a predetermined knowledge base using the nearest neighbor algorithm, then we can provide the best recommendations in accordance with student academic performance.

At this stage will look for similarity values using formula 1 with W = Weight, S = Similarity, RWS =

Results of W * S, TW = Total Weight. The following is the calculation process for finding similarity values.

Table. 5 Rule Based Very Bad

Rule Base "Very Bad"		New Case
Factor	**Value**	**Value**
[PD] Age	>26	27
[PD] Gender	-	Male
[SS] Emphasize main ideas	Never	Rarely
[SS] Group Study	Never	Rarely
[SS] Take notes	Never	Always
[BS] Hard work	Never	Often
[CS] Mid-term grades	0 - 25	65
[CS] Finals grades	0 - 25	80

Table. 6 Calculation of Rule Based Very Bad With New Cases

W	S	RWS	Total Weight	Similarity With Algorithm NN (Total RWS/W)
5	1	5	40	**0.13**
5	0	0		
5	0	0		
5	0	0		
5	0	0		
5	0	0		
5	0	0		
5	0	0		

Table. 7 Rule Based Bad

Rule Base "Bad"		New Case
Factor	**Value**	**Value**
[PD] Age	24 to 26	27
[PD] Gender	-	Male
[SS] Emphasize main ideas	Rarely	Rarely
[SS] Group Study	Rarely	Rarely
[SS] Take notes	Rarely	Always
[BS] Hard work	Rarely	Often
[CS] Mid-term grades	25 - 50	65
[CS] Finals grades	25 - 50	80

Table. 8 Calculation of Bad Base Rule with New Cases

W	S	RWS	Total Weight	Similarity With Algorithm NN (Total RWS/W)
5	0	0	40	**0.25**
5	0	0		
5	1	5		
5	1	5		
5	0	0		
5	0	0		
5	0	0		
5	0	0		

Table. 9 Rule Based Average

Rule Base "Average"		New Case
Factor	**Value**	**Value**
[PD] Age	21 to 23	27
[PD] Gender	-	Male
[SS] Emphasize main ideas	Sometimes	Rarely
[SS] Group Study	Sometimes	Rarely
[SS] Take notes	Sometimes	Always
[BS] Hard work	Sometimes	Often
[CS] Mid-term grades	50 - 75	65
[CS] Finals grades	50 - 75	80

Table. 10 Calculation of Rule Based Average With New Cases

W	S	RWS	Total Weight	Similarity With Algorithm NN (Total RWS/W)
5	0	0	40	**0.13**
5	0	0		
5	0	0		
5	0	0		
5	0	0		
5	0	0		
5	1	5		
5	0	0		

Table. 11 Rule Based Good

Rule Base "Good"		New Case
Factor	**Value**	**Value**
[PD] Age	18 to 20	27
[PD] Gender	Male	Male
[SS] Emphasize main ideas	Often	Rarely
[SS] Group Study	Often	Rarely
[SS] Take notes	Often	Always
[BS] Hard work	Often	Often
[CS] Mid-term grades	75 - 90	65
[CS] Finals grades	75 - 90	80

Table. 12 Calculation of Rule Based Average With New Cases

W	S	RWS	Total Weight	Similarity With Algorithm NN (Total RWS/W)
5	0	0	40	**0.38**
5	1	5		
5	0	0		
5	0	0		
5	0	0		
5	1	5		
5	0	0		
5	1	5		

Table. 13 Rule Based Excelent

Rule Base "Excelent"		**New Case**
Factor	**Value**	**Value**
[PD] Age	<= 18	27
[PD] Gender	Female	Male
[SS] Emphasize main ideas	Always	Rarely
[SS] Group Study	Always	Rarely
[SS] Take notes	Always	Always
[BS] Hard work	Always	Often
[CS] Mid-term grades	90 - 100	65
[CS] Finals grades	90 - 100	80

Table. 14 Calculation of Rule Based Excelent with New Cases

W	**S**	**RWS**	**Total Weight**	**Similarity With Algorithm NN (Total RWS/W)**
5	0	0	40	**0.13**
5	0	0		
5	0	0		
5	0	0		
5	1	5		
5	0	0		
5	0	0		
5	0	0		

From the calculation table above for the similarity value of Very Bad is **0.13**, Bad is **0.25**, Average is **0.13**, Good is **0.38** and Excelent is **0.13**. The highest value from the data above is on **Good** performance with a value of **0.38**. So it can be concluded that this new case recommended the academic performance of students namely **Good**.

IV. RESULTS AND DISCUSSION

A. The Testing of Accuracy

The results of this study are for the academic performance of the Very Bad category is 0 students, the Bad category is 3 students, the Average category is 79 students, the Good category is 16 students and the Excelent category is 12 students. Testing accuracy using the confusion matrix method, the data consisted of 110 students.

Table. 15 Dataset of Student Prediction Results

NO	**..**	**MG**	**FG**	**ACADEMIC PERFORMANCE PREDICTION**
1	..	72	85	AVERAGE
2	..	40	76	AVERAGE
3	..	68	84	AVERAGE
4	..	72	85	AVERAGE
5	..	64	83	AVERAGE
6	..	36	74	EXCELENT
7	..	68	84	AVERAGE
8	..	40	76	EXCELENT
9	..	72	85	AVERAGE
10	..	72	85	AVERAGE
11	..	72	85	AVERAGE
12	..	44	77	AVERAGE
13	..	72	85	AVERAGE
14	..	52	79	GOOD
15	..	68	84	AVERAGE
..	..	..	..	..
110	..	44	82	AVERAGE

Here is a matrix obtained from the dataset using WEKA.

Table. 16 Matrix Dataset

Class	Average	Excelent	Good	Bad
Average	81	2	3	0
Excelent	1	11	0	0
Good	2	0	6	0
Bad	1	0	0	3

To calculate the accuracy value, we use confusion matrix classification using WEKA, we will get the value as in Table 16.

Table. 17 Confusion Matrix Classification

Actual Value		**Prediction**	
		Positive	Negative
Actual Value	Positive	101 (TP)	9 (FP)
	Negative	0 (FP)	0 (TN)

Table Confusion matrix classification shows a TP (True Positive) of 101 and FP (False Positive) of 9, then the Accuracy results are as follows :

$$\boldsymbol{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} X\ 100\%$$

$$\boldsymbol{Accuracy} = \frac{101 + 0}{101 + 0 + 9 + 0} X\ 100\% = \mathbf{91.82}\%$$

Confusion matrix testing results using WEKA software with 70% training data and 30% testing data from existing datasets. The results can be seen in Figure 4.

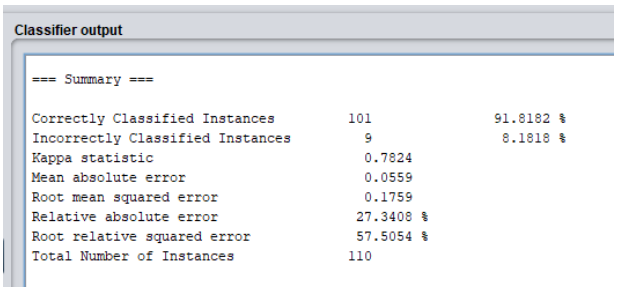


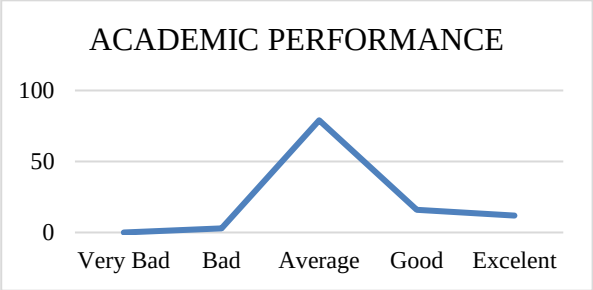
Figure 3. Test results using WEKA

The accuracy calculation results above show that the academic performance of students by using similarity

weights with algorithms is suitable for predicting students academic performance with accuracy **91,82%**.

B. Student Classification

After being given recommendations on student academic performance, existing data will be used to group data into 3 classes. For student academic performance data can be seen in Graph 1.



Graph. 1 Student Academic Performance

The new class groups according to students' academic performance recommendations are as follows. For class A consists of 1 Bad, 26 Average, 5 Good and 4 Excelent. Class B consists of 1 Bad, 27 Average, 5 Good and 4 Excelent. Class C consists of 1 Bad, 26 Average, 6 Good and 4 Excelent. For complete data, see Table 18.

Table. 18 Classroom Results

ACADEMIC PERFORMANCE	**A**	**B**	**C**	**TOTAL**
Very Bad	0	0	0	0 Student
Bad	1	1	1	3 Student
Average	26	27	26	79 Student
Good	5	5	6	16 Student
Excelent	4	4	4	12 Student

V. CONCLUSION

From the results of the above research it can be concluded that the nearest neighbor algorithm by adapting the Case Based Reasoning method is able to recommend Student academic performance with 91,82% accuracy. Then the prediction results can be used to classify students randomly based on their abilities.

VI. REFERENCE

[1] O. A. Echegaray calderon and D. Barrios aranibar, "Optimal selection of factors using Genetic Algorithms and Neural Networks for the prediction of students academic performance," *Lat. Am. Congr. Comput. Intell.*, pp. 40–46, 2015, doi: 10.1109/LA-CCI.2015.7435976.

[2] M. Paliwal and U. A. Kumar, "A study of academic performance of business school graduates using neural network and statistical techniques," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7865–7872, 2009, doi: 10.1016/j.eswa.2008.11.003.

[3] A. Flores, L. Alfaro, and J. Herrera, "Proposal model for e-learning based on Case Based Reasoning and Reinforcement Learning," in *EDUNINE 2019 - 3rd IEEE World Engineering Education Conference: Modern Educational Paradigms for Computer and Engineering Career, Proceedings*, 2019, pp. 1–6, doi: 10.1109/EDUNINE.2019.8875800.

[4] A. Rahman, R. A. Mutiarawan, A. Darmawan, Y. Rianto, and M. Syafrullah, "Prediction of students academic success using case based reasoning," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2019, pp. 171–176, doi: 10.23919/EECSI48112.2019.8977104.

[5] A. Rahman and A. Qosim, "SISTEM CERDAS PENGELOMPOKAN MAHASISWA BERDASARKAN PREDIKSI PERFORMA BELAJAR DENGAN METODE CASE BASED REASONING," *J. Edik Inform. J. Edik Inform.*, vol. 8, no. 1, pp. 13–25, 2021, doi: http://dx.doi.org/10.22202/ei.2021.v8i1.5030.

[6] F. Tempola, A. Arief, and M. Muhammad, "Combination Of Case-Based Reasoning And Nearest Neighbour For Recommendation of Volcano Status," *Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 348–352, 2017, doi: 10.1109/ICITISEE.2017.8285525.

[7] A. Rahman and U. Budiyanto, "Case based reasoning adaptive e-learning system based on visual-auditory-kinesthetic learning styles," in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2019, pp. 177–182, doi: 10.23919/EECSI48112.2019.8976921.

[8] Abdul Rahman, Destiarini, and J. Kuswanto, "Fuzzy Logic Recommended Student Learning Levels," *J. Inform. Polinema*, vol. 7, no. 2, 2021, doi: 10.33795/jip.v7i2.531.