Combination of Case Based Reasoning with Nearest Neighbor and Decision Tree for Early Warning System of Student Achievement

1st Mardiyyah Hasnawi
Faculty of Computer Science
Universitas Muslim Indonesia
Makassar, Indonesia
mardiyyah.hasnawi@umi.ac.id

2nd Nia Kurniati
Faculty of Computer Science
Universitas Muslim Indonesia
Makassar, Indonesia
nia.kurniati@umi.ac.id

3rd St. Hajrah Mansyur
Faculty of Computer Science
Universitas Muslim Indonesia
Makassar, Indonesia
hajrah.mansyur@umi.ac.id

4th Irawati
Faculty of Computer Science
Universitas Muslim Indonesia
Makassar, Indonesia
irawati.irawati@umi.ac.id

5th Tasrif Hasanuddin
Faculty of Computer Science
Universitas Muslim Indonesia
Makassar, Indonesia
tasrif.hasanuddin@umi.ac.id

***Abstract*—Student achievement is one of the main focuses to increase university credibility. An early warning system is needed to prevent more risks. The early warning systems (EWS) of student achievement has been possible with a combination of case based reasoning (CBR), k-nearest neighbor (K-NN), and decision tree. CBR is used to obtain a solution that stores knowledge so that it can predict student achievement. This research is combining the Case Based Reasoning, K-Nearest Neighbor (K-NN), and Decision Tree (DT) methods for the prediction of student achievement that applied in the early warning system. The attributes of an early warning system of student achievement are genders, distances of residence, ages, high schools, majors, and grade point average (GPA) for six semesters. The results show that accuracy rate is 60.5% of 55 data in the early warning system of student achievement and a model CBR for Early Warning System of Student Achievement.**

Keywords—case based reasoning, decision tree, k-nearest neighbor, early warning system

I. INTRODUCTION

An early warning system (EWS) is one strategy to improve and guarantee the continuity of organization or university. It is very crucial to prevent and reduce impact and risks as like a development of the monitoring and early warning system of students' learning [1]. It provides information quickly and precisely related to the problems that applied in the future.

The development of early warning systems has been widely developed using the latest technology such as warnings for the risks and impact of the internet of things based environment disasters [2]. It is also expanded in the field of education to predict students that success in university entrance exams [3], the student emergency at university [4], the final grades of student [5], and monitoring and prevention mechanisms [1].

Student achievement problems affect the credibility of the institution in passing students. Some private universities have the same issues as GPA's student getting very low. Several factors affecting student achievement are motivation, learning methods, facilities within the university and faculty, and the environment. The GPA is one measure of the success in student achievement. The student attributes required in determining student achievement are genders, distances of residence, ages, high school, majors, and GPA's Semester.

EWS can help stakeholders determine patterns of student achievement that will have an impact on study graduation. It is also currently developed with various methods and technologies. The Implementation of EWS at universities has been increasingly developed in cloud campus [1]. One of the benefits of EWS is providing early warning and plans by informing students and lecturing guidance, parental feedback and so on [1]. Four actors are very urgent in the Academic Early Warning Eco-system namely students, teachers, college and parents [6].

In EWS, the data are processed optimally to extract information and obtain new knowledge so that can be used to considerate decision making. Data collection is used to predict student achievement through the process of taking cases, grouping or classifying cases, assessing similarities, resolving cases recorded previously [7]. Therefore, to overcome these problems, an EWS based on CBR is developed that can predict student achievement index by utilizing cases stored in the student database.

According to L.D. Xu [8] that CBR was based on cognitive models and focused on cases stored in memory and ideas as past experiences and a way of solving problems. The problem solving focused on a building solution of new cases and a modifying solution of pre-cases. Development of the EWS based on CBR, there are several stages such as taking, selecting, classifying, assessing the similarity of cases on matching cases and adapting them to a system. The classifying case is conducted to find a pattern that describes and separates data classes from another. The classifying cases collaborate Decision Tree Algorithm while the grouping of similarity cases uses K-Nearest Neighbor as the predictive value of the new instance value. The EWS based on CBR is developed with classification methods such as Decision Tree, and K-Nearest Neighbor so that can predict student achievement.

978-1-5386-8050-6/18/$31.00 ©2018 IEEE

II. Research Method

A. Case Based Reasoning (CBR)

CBR has been applied in various cases such as a determining system for the status of volcanoes [9], a context of awareness systems [10], the autonomic managers [11], and searching for the same answers [12]. It was also developed with an Object-Oriented model using Situation Tree [13]. According to CBR, has four sub-processes [7]:

1) Retrieve: taking the most common problem or case

A similar case-taking process is carried out with a criterion that determines how to find and control a case in a base case so that it is more effective in assessing cases [9]. The searching process is carried out with cases by comparing new cases and content-based features [14]. The similarity assessment process using the K-Nearest Neighbor Algorithm determines the highest similarity value new case [9].

2) Reuse: reuse the case to try and solve it

This process is done to find solutions and problems by adapting to the query system after taking the case.

3) Revise: revise the offered solutions

If the solution offered is rejected then the cases revised until the case accepted as a solution.

4) Retain: store experiences for future purposes in other cases

This process adaptation made in the system. This process can be proposed to modify the level of similarity of cases with index structure modification.

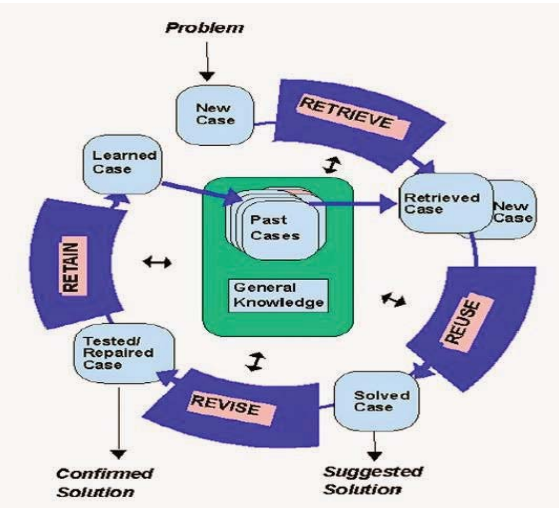

Fig. 1. Case Base Reasoning Cycle [8]

B. Decision Tree

Tree Structure, the internal node shows the test on the attribute, each branch shows a test result, and each leaf node holds the class label [15]. Several categories of Decision Tree methods are C.45, CHART, CHAID. The C4.5, Tree-Induction Algorithm makes the best and fastest classification accuracy among other algorithms [16],[26]. Classification and Tree Regression (CART) is a non-parametric method and does not require variables to be selected first [17]. CHAID can be used in prediction and classification, and for detecting interactions among variables. Both CHAID and CART algorithms can be applied to analyze the problem of regression or classification type [18]. The advantage of the CART is the ability to choose the most discriminatory features and classification conducted by a little calculation. The disadvantage is the complicated calculations occurs when it has more the problem spaces and the possible level of misclassification that occurs when training data is a few numbers of classes [19].

Methods related to decision tree classification algorithms have been developed in several studies such as Decision Tree Using Fast Splitting Attribute Selection (DTFS) [20], Classification by Clustering (CBC) [21]and EWS Model to predict student achievement [5].

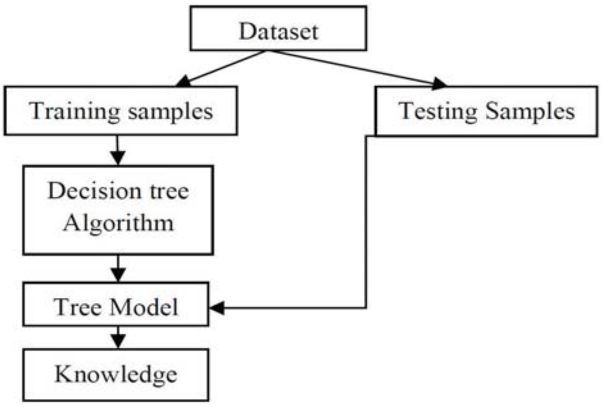

Fig. 2. Workflow of Decision Tree [19]

C. K-Nearest Neighbor (K-NN)

K-NN is one of the methods in supervised learning that widely used to identify and assess similarities and uses Euclidean metrics to determine similarities [22]. The value of k is used to category selection in the data training case. K-NN acts by finding groups of objects in the closest (similar) training data to objects in new data or testing data. K-NN has the advantage of training a lot of noise data that can be effective if the training data is massive. The K-NN method can also be combined with the CBR method to obtain similarities in cases such as volcano status systems [9], and chronic lung diagnosis [23]. The K-NN algorithm are [12]:

1) Determining parameter k.
2) Calculating the distance among data to be evaluated.
3) Sorting the distance ascending.
4) Determining the shortest distance to the sequence.
5) Pairing the class accordingly.
6) Looking for the number of classes from NN and setting the class as class data to evaluate.
7) Classification and Similarities of Stages.

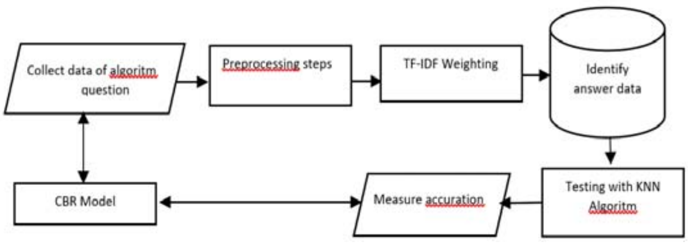

Fig. 3. The Classification Process and Similarity Stages [12]

III. Results And Discussion

The CBR and classification methods namely decision tree (CART) and K-NN are developed in the early warning system. The case study of the development of an early warning system on student achievement begins with gathering new cases into data collections. Data collections serve to form a data set. A case study in one of the private universities in Makassar with 55 data stored as a case base consisting of two datasets, namely profile data and academic values. The data collection has attributes used in classifications including student identification number, gender, distances of residence, address, subjects, grade point average (GPA) 6 semesters, credits, ages, majors, and high school.

The data set consists of two parts namely training data and testing data. The academic transcripts for four semesters as training data and two semesters as testing data. The original data set is urgent as input in the classification algorithm. At the preprocessing stage, extract the data sample to avoid noise and bias. The academic transcripts are as the target attribute. In the Decision tree model, built using 80% of the original dataset are determined as data training. Then it is tested using 20% of the dataset as data training.

Independent variable importance, namely gender, age, distance, majors and high school, is shown in Fig. 4.

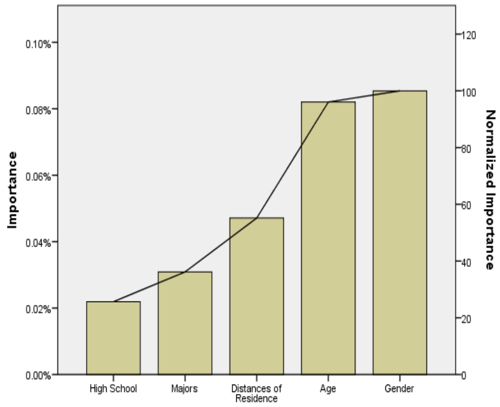

Fig. 4. Normalized Importance Attributes

TABLE I. Classification Cart Model

Classification					
Observed	**Predicted**				
	Excellent	**Good**	**Poor**	**Very Good**	**Percent Correct**
excellent	7	0	5	0	58.3%
good	0	0	4	0	0.0%
poor	1	0	23	2	88.5%
very good	2	0	5	6	46.2%
overall percentage	18.2%	0.0%	67.3%	14.5%	65.5%

Based on these datasets analyzed using the CART classification method by classifying an event into the category of poor, good, very good, excellent with an appropriate 65.5%. The calcification results from the decision tree become new knowledge for the next case. Based on the tree analysis model, gender is attribute having the highest percentages. The others have a crucial role in predicting. They are majors and the high school. These attributes are used as a case base to perform calculations using KNN analysis.

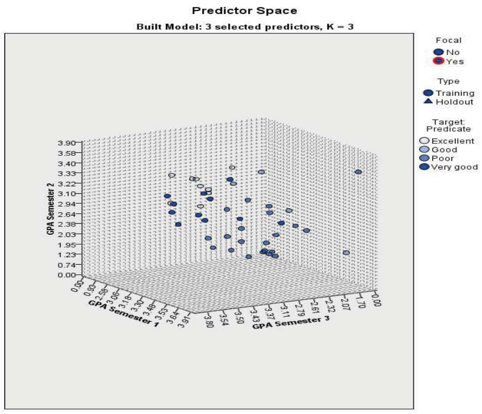

Fig. 5. Classification KNN of Student Data based on GPA

The results of the KNN analysis show the percentage of poor predicate having the highest percentage level. The analysis is used to develop the student's early warning system so that stakeholders take precautionary measures against the achievements of students who have a poor predicate. Development of early warning infrastructure at the application service level which is formed from the CBR development model by elaborating the process that occurs using a decision tree algorithm and the KNN algorithm. The results of a study of the development of a combination of CBR, KNN and CART methods, the EWS Model was obtained in Fig. 6.

TABLE II. Classification Table Using Knn

Classification					
Observed	**Predicted**				
	Excellent	**Good**	**Poor**	**Very good**	**Percent Correct**
Excellent	7	0	2	1	70.0%
Good	0	0	4	0	0.0%
Poor	3	2	15	0	75.0%
Very good	3	0	2	4	44.4%

The development was started from the data collection process in the form of data sets. Through the pre-processing, data sets are grouped into classes as training data and testing data. The analysis process using Decision Tree Algorithm (CART) was modelled into Tree Model to produce useful knowledge for new cases. The new cases in the pre-processing were also taken new cases then analyzed by KNN algorithm to determine cases that had the highest similarity.

The results of the Retrieved Case were reused to obtain a solution. If no solution is found, it will be revised and corrected until the best solution is received. The solution was stored as a knowledge and could be reused if there is a recent case. Based on data as many as 55 samples of student data for 6 semesters in training data obtained 30.2% predicted to get excellent predicate.

Early Warning Infrastructure consists of 4 levels. These are user level, Interfaces Level, Application Services Level, and Devices and Servicers level. The user level showed that

4 actors who has role in the early warning system of student achievement. They are lecturers, students, parents and university management. While interfaces and Devices level is composed by devices, networks, and data center.

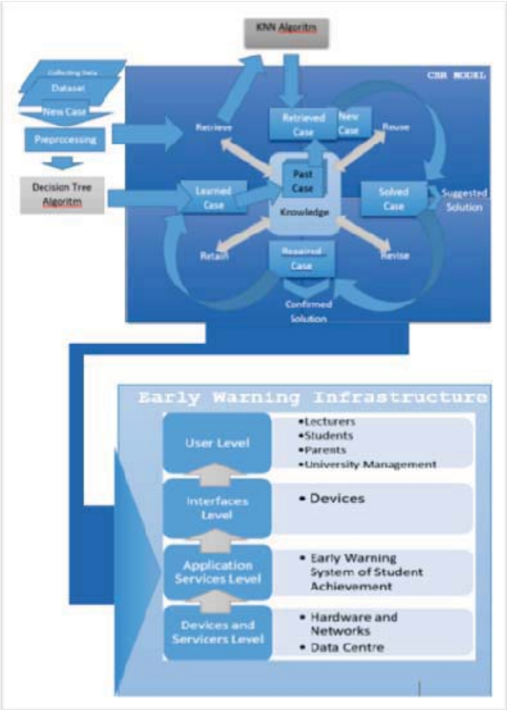

Fig. 6. Model CBR for Early Warning System of Student Achievement

IV. CONCLUSION

Based on the study method and testing the data using CBR, CART, and KNN on student achievement, the percentage level of accuracy between the KNN algorithm and CART algorithm has a significant difference. In the CART algorithm, the accuracy rate is 65.5% while the K-NN algorithm has the accuracy rate of 60.5%. Early warning system successfully predicts student achievement with features of 6 attributes and 55 data.

REFERENCES

[1] Z. FuSheng, Y. Ling, J. Z. Q. XinJun and Z. Mingkui, "The Monitoring and Early-warning System of Students' learning Based on the Campus Cloud," in The 12th International Conference on Computer Science & Education (ICCSE 2017), University of Houston, USA, 2017

[2] P. Stefan, E. M. Stuart, C. Fernando, T. Ran, N. Ocal and B. Ulrich, "A Semantic IoT Early Warning System for Natural Environment Crisis Management," IEEE , vol. 3, no. 2, pp. 246-256, 2015.

[3] H. Göker and H. I. Bülbül, "Improving an Early Warning System to Prediction of Student Examination Achievement," in 13th International Conference on Machine Learning and Applications, Detroit, MI, USA, 2014.

[4] F. Li, "Study on Emergency Early Warning System for University Students," in International Conference on Information Management, Innovation Management and Industrial Engineering, Sanya, China, 2012.

[5] W. Liu, J. Wu, X. Gao and K. Feng, "An early warning model of student achievement based on decision trees algorithm," in 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Hong Kong, China, 2017.

[6] Z. FuSheng, J. ZongFu, Q. XinJun and Z. Mingkui, "The Monitoring and Early-warning System of Students' learning Based on the Campus Cloud," in 2017 12th International Conference on Computer Science and Education (ICCSE), Houston, TX, USA, 2017.

[7] E. P. A. Aamodt, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," IOS Press, vol. 7, no. 1, pp. 39-59, 1994.

[8] L. Xu, "Case Based Reasoning," IEEE, vol. 13, no. 5, pp. 10-13, 1995.

[9] F. Tempola, A. Arief and M. Muhammad, "Combination Of Case-Based Reasoning And Nearest Neighbour For Recommendation Of Volcano Status," in 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017.

[10] N. Selviandro, K. M. Sabariah and S. Saputra, "Context Awareness System on Ubiquitous Learning with Case Based Reasoning and Nearest Neighbor Algorithm," in 2016 Fourth International Conference on Information and Communication Technologies (ICoICT), Bandung, 2016.

[11] C. Khan and J. M. Khan, "Exploiting Fuzzy Clustering and Case-based Reasoning for Autonomic Managers," in 2016 IEEE International Conference on Autonomic Computing (ICAC), Wurzburg, 20 16.

[12] A. Y. Gerhana, R. A. Atmadja, B. W. Zulfikar and N. Ashanti, "The Implementation of K-Nearest Neighbor Algorithm in Case-Based Reasoning Model for Forming Automatic Answer Identity and Searching Answer Similarity of Algorithm Case," in 2017 5th International Conference on Cyber and IT Service Management (CITSM), Denpasar, 2017.

[13] K. R. Akhmadulin, N. I. N. Gluhih and Y. I. Karyakin, "An Object-Oriented Model of Case-Based Reasoning System Using Situations Tree," in 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, 2016.

[14] S. K. S. S. C. K. Pal, "Foundations of Soft Case-Based Reasoning," Wiley-Interscience publication, 2004

[15] K. M. HAN Jiawei, "Data Mining Concepts and Techniques," Higher Education Press, pp. 14-20, 2001.

[16] T. S. Lim, W. Y. Loh and Y. S. Shih, "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms," Machine Learning, vol. 40, no. 3, pp. 203-228, 2000.

[17] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum, 1981.

[18] S. kakavand, T. Mokfi and J. M. Tarokh, "Prediction the Loyal Student Using Decision Tree Algorithms," International Journal of Information and Communication Technology Research, vol. 4, no. 1, pp. 32-37, 2014.

[19] M. Balamurugan and S. Kannan, "Performance Analysis of Cart and C5.0 using Sampling Techniques," in 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India, 2016.

[20] F.-A. A, A. Carrasco-Ochoa J and y. M.-T. J. Sánchez-Díaz G, "Decision tree induction using a fast splitting attribute selection for large datasets," Expert Systems With Applications, pp. pp. 4290 - 14300, 2011.

[21] B. Aviad and G. Roy, "Classification by clustering decision tree-like classifier based on adjusted clusters," Expert Systems with Applications, vol. 38, no. 7, p. 8220–8228, 2011.

[22] K. W. a. L. Saul, "Distance metric learning for large margin nearest neighbor classification," The Journal of Machine Learning Research, vol. 10, pp. 207-244, 2009.

[23] S. L. M. T. a. L. J. Guessoum, "A Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease," Expert Systems With Applications An International Journal, vol. 41, pp. 267-273, 2014.

[24] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh and Q. Yang, "Top 10 algorithms in data mining," Springer-Verlag London, vol. 14, no. 1, pp. 1-37, 2008.

[25] S. C. K. S. Sankar K. Pal, Foundations of Soft Case-Based Reasoning, New Jersey, United States: Wiley InterScience, 2004.

[26] Budiman E., Haviluddin, Dengan N., Kridalaksana A.H., Wati M., Purnawansyah 2018. Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation. ICCST 2017. Lecture Notes in Electrical Engineering, vol 488. Springer, Singapore.