Advances in Social Science, Education and Humanities Research, volume 585
Proceedings of the 1st UMGESHIC International Seminar on Health, Social Science and Humanities (UMGESHIC-ISHSSH 2020)

Performance Comparison of Ensemble-based k-Nearest Neighbor and CART Classifiers for the Classification of Adaptive e-learning User Knowledge Levels

Utomo Pujianto, Harits Ar Rosyid, Aditya Cahyadi Putra

Universitas Negeri Malang, East Java, Indonesia
Corresponding author email: utomo.pujianto.ft@um.ac.id

Abstract. A person's learning style that refers to the preferred way of learning is the basis for the development of the Adaptive Educational Intelligent Hypermedia System (AEIHS) or adaptive e-learning. By knowing specific learning styles, the system can provide recommendations and offer instructions to someone how to optimize the learning process. The right learning style is needed, because it can support the achievement of a person's level of knowledge in learning. The problem in determining this level of knowledge is related to the performance of data mining methods as measured by algorithm performance. High performance is indicated by the optimal results of the algorithm used. On the other hand, the K-Nearest Neighbor (KNN) algorithm and Classification and Regression Trees (CART) have been shown to have good performance in various fields. Therefore, this study aims to compare the performance of data mining methods, namely K-Nearest Neighbor (KNN) and Classification and Regression Trees (CART). There are six scenarios carried out for comparison, including comparing the performance of the original algorithm and ensemble methods, namely bagging and boosting. The results of this study are the best performance results for the classification of the level of user knowledge of adaptive e-learning from several scenarios performed. CART boosting algorithm shows the best performance with accuracy of 94.0%, precision 94.2%, and recall of 94.0%. The best scenario of the algorithm used is expected to be a guide for developing AEIHS-based Education or adaptive e-learning.

Keywords: *Ensemble-Based K-Nearest, CART, e-learning*

1. INTRODUCTION

Education is a process of facilitating learning in acquiring knowledge, beliefs, values, skills and habits. Interaction in education is also referred to as the learning process [1]. The learning process is the process of acquiring or modifying new knowledge, behavior, skills, values, or existing preferences [2]. As a result of the learning process, humans will get knowledge from these activities. Knowledge is a condition where a person is in complex cognitive contact and deals with realities such as perception, communication, and reasoning [3]. Because knowledge involves complex cognitive. On the other hand, the ability of each person to absorb knowledge varies and are influenced by habits, environmental conditions, or personal characteristics that refer to learning styles. A person's learning style reflects the preferred way of learning, and it is the basic for the development of the modern education system [4]. An appropriate learning style is needed, because it can support the achievement of the knowledge level in learning[5].

Because support in the attainment of knowledge is needed, a modern education system called AEIHS (Adaptive Educational Intelligent Hypermedia System) or familiarly called adaptive e-learning was developed[6]. Adaptive e-learning is the development of conventional learning, which is the development of an education system by designing a system that can analyze user and domain models from users who use learning through digital media. So that it can accommodate the problem of differences in a person's speed in absorbing and understanding knowledge.

Adaptive e-learning has several processes, firstly the process of collecting user information (user profile) is useful for knowing all the activities that are being carried out. second, the process of developing a user model which is categorized into user knowledge. User Knowledge is a category that measures the level of user knowledge in understanding an object which will be divided into several categories, namely very low, low, middle, and high [7].

Based on the comparison of algorithms discussed before, this study discusses the classification of the knowledge level of adaptive e-learning users using the K-Nearest Neighbor (KNN) algorithm and Classification and Regression Trees (CART). The K-Nearest Neighbor (KNN) algorithm is a supervised learning algorithm which is based on finding the proximity of the research attributes and classifying them based on the most votes from the neighbors found. The advantages of the K-Nearest

Neighbor (KNN) algorithm are that it models a training process that is very fast, simple, has good performance against noisy training data, and is effective with large training data. On the other hand, K-Nearest Neighbor (KNN) also has weaknesses including biased K values, limited memory, and poor performance on irrelevant attributes[8]. The Classification and Regression Tree (CART) algorithm is an algorithm that uses the decision tree technique to describe the nonparametric statistics of research attributes to obtain a highly accurate data group as the center of the classification process. The advantages of the Classification and Regression Tree (CART) algorithm are that it produces results that are easy to interpret, accurate and fast in developing classification models, can be used on large data, suitable for many data types [9]. On the other hand, the weakness of the Classification and Regression Tree (CART) is that it is very sensitive to new data because it depends on the number of samples, and the selection depends on the value taken from one variable only. [10]. The testing process is carried out by creating several scenarios using the K-Nearest Neighbor (KNN) algorithm, Classification and Regression Trees (CART), and a combination of the bagging and boosting methods.

Previous research, with almost the same algorithm, first was conducted by Ramadhan & Wijanarto, entitled Implementation of Comparative K-Nearest Neighbor Algorithms and Classification and Regression Tree in Classification of Employee Performance Evaluation in Companies [11]. The results of this study are that the Classification and Regression Tree is said to be better because this algorithm has advantages that the K-Nearest Neighbor algorithm does not have, namely the Classification and Regression Tree is easier to interpret, more accurate and faster in its calculations.

In this study the K-Nearest Neighbor (KNN) algorithm and Classification and Regression Trees (CART) were tested to classify the level of knowledge of adaptive e-learning users. The results of this study are the best performance results for the classification of the level of knowledge of adaptive e-learning users from several trial scenarios that have been carried out, so that they can be developed into further research in making recommendations or guidelines for developing education based on the Adaptive Educational Intelligence Hypermedia System (AEIHS) or adaptive e-learning.

2. EVALUATION SYSTEM

This section discusses the Adaptive Educational Intelligent Hypermedia System (AEIHS) or adaptive e-learning which is used as a tool to collect adaptive e-learning user data. Then the discussion continues with the process of collecting data from adaptive e-learning users with this system.

2.1. Adaptive E-learning System

This system is an adaptive system that was built as a tool to measure the level of user knowledge in DC electronics courses at the Department of Electrical Engineering Education, Gazi University, Ankara, Turkiye. The User Model System in the Adaptive Educational Intelligent Hypermedia System (AEIHS) or adaptive E-Learning tracks and stores interactions between users and the system [12]. After evaluating the information carried out by machine learning, including information about the user's state of knowledge, choices, learning styles and other attributes of the user, it is obtained that adaptive e-learning regulates the interactions displayed on adaptive e-learning users and determines the material adapted to learning activities so that The needs and personalization of adaptive e-learning users are met, such as the level of difficulty of the topics tested will be adjusted to the level of user knowledge gradually from the personal attributes of each user that have been formed by their UM.

Users must log in first so that activity from users is recorded in the learning process. There are several processes in adaptive e-learning, including the process of collecting user information (user profile) which is useful for knowing all activities undertaken. Furthermore, the process of building a user model to be categorized into user knowledge. User knowledge is a category that measures the level of user knowledge in understanding an object / material which is further divided into several levels, namely very low, low, middle, and high.

2.2. Data Collection

Data collection was done by using User Modeling (UM) in adaptive e-learning, which is useful for storing information about the status of knowledge, learning styles, previous studies and students' personal identities. User Modeling (UM) is created using the information provided by the user during enrollment in the adaptive e-learning application. The User Modeling (UM) structure is adjusted to represent the level of students' knowledge about learning objects at a certain level.

In adaptive e-learning applications, it is possible for two types of adaptations at the content level and the navigation level. When users interact with the system, they provide feedback about the purpose of the given content, materials used for object education, exams and instructional object relationships. By evaluating this feedback and the

data obtained from the MUs via the intelligent system, information is required about the instructional object provided. In the session layer designed for students, there is an instruction page that students prepare to start an activity and the UM of the user is stored on a server where student actions are tracked and stored. Table I shows the attributes taken from MUs.

TABLE 1. List of Attributes In The Dataset

Attribute Name	Data Type	Attribute Explanation Value Range
STG	Numeric	Study Time (Objective Object)
SCG	Numeric	Number Of Repetitions (Relevant Object)
STR	Numeric	Study Time (Relevant Object)
LPR	Numeric	Exam Performance (Relevant Object)
PEG	Numeric	Test Performance (Objective Object)
UNS	Nominal	Level of Understanding

3. METHODOLOGY

This section discusses the proposed test scenarios and the algorithms used.

3.1. The Proposed Test Scenario

In this subsection, the steps that have been carried out are discussed to see the comparison of the K-Nearest Neighbor (KNN) algorithm and the Classification and Regression Tree (CART), Classification and Regression Tree (CART) on the classification of the knowledge level of adaptive e-lerarning users.

The comparisons include the K-Nearest Neighbor (KNN) algorithm, K-Nearest Neighbor (KNN) bagging, K-Nearest Neighbor (KNN) boosting, Classification and Regression Tree (CART), Classification and Regression Tree (CART) bagging, and Classification. and Regression Tree (CART) boosting. The test details of these 6 scenarios are shown in Table II.

TABLE 2. List of Attributes In The Dataset

Scenario	Algorithm	Type of Testing
1	KNN	Grades K 1 – 21
2	KNN Bagging	Best K Value & Bagging
3	KNN Boosting	Best K Value & Boosting
4	CART	Best MNO Value
5	CART Bagging	Best MNO value & Bagging
6	CART Boosting	Best MNO value & Boosting

The K-Nearest Neighbor (KNN) algorithm was carried out by experimenting with finding the best K value to get the maximum algorithm performance, the K value shows as the closest neighbor class of a data. Testing the K value by trying the K value from 1 to 21.

The Classification and Regression Tree (CART) algorithm is carried out by experimenting with finding the Minimum number of observations to get the maximum algorithm performance, the Minimum number of observations shows the minimum number of observations as a parameter.

In the parent node. Testing the Minimum number of observations by trying the Minimum number of observations 1 to 10.

From testing the K-Nearest Neighbor (KNN) algorithm based on the best K value and the Classification and Regression Tree (CART) algorithm based on the best Minimum number of observations, the bagging and boosting methods of the two algorithms are applied. All tests use cross validation of 10 folds.

3.2. The Proposed Test Scenario

The calculation of the K-Nearest Neighbors (KNN) algorithm is a method for classifying new objects based on (K) its closest neighbors. K-Nearest Neighbors (KNN) is a supervised learning algorithm, where the results of new query instances are classified based on the majority of the categories in K-Nearest Neighbors (KNN). The following are the steps for classification using the KNN:

- Determine the value of K.
- Calculate the distance between training data and test data using equations.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (2.1)$$

Information:
Xi = value from training data
Yi = value of test
i = data variable
k = dimesion of data

- Sort neighbors by smallest value to largest value.
- Select as many neighbors as K from the sorted list.
- Determine the value of the most neighbors.

3.3. Classification and Regression Trees (CART)

The Classification And Regression Tree (CART) algorithm provides classification or regression results depending on the category or numeric dataset [13]. The steps for the Classification And Regression Tree (CART) algorithm are as follows:

- In the first step, arrange a candidate split. This compilation is carried out on all predictor variables completely (exhaustive). The list containing candidate branches is called the current branch candidate list.
- The second step is to assess the overall performance of the prospective branches on the most recent branch candidate list by calculating the value of the suitability value ϕ (s besarant).

$$\phi(s|t) = 2P_L P_R \sum_{j=1}^{\#Categoryi} |P(j|t_L) - P(j|t_R)| \quad (2.2)$$

Which,

$t_L = Candidates\ for\ the\ left\ branch\ of\ the\ decision\ dot\ t$

$t_R = Candidate\ for\ the\ right\ branch\ of\ the\ decision\ dot$

$$P_L = \frac{Number\ of\ records\ on\ the\ t_L\ left\ branch\ candidate}{Number\ of\ record\ on\ training\ data}$$

$$P_R = \frac{Number\ records\ on\ the\ t_R\ right\ branch\ candidate}{Number\ of\ record\ on\ training\ data}$$

$$P(j|t_L) = \frac{Number\ of\ records\ category\ j\ candidate\ left\ brancht_L}{Number\ of\ record\ in\ decision\ dot\ t_L}$$

$$P(j|t_R) = \frac{Number\ of\ records\ category\ j\ candidate\ right\ branch\ t_R}{Number\ of\ record\ in\ decison\ dot\ t_L}$$

- The third step is to determine which branch candidates will actually become branches by selecting the branch candidates that have the largest ϕ (s | t) suitability value. After that, a branch was formed. If there are no more decision dots, the CART algorithm will be stopped. However, if there are still decision dots, the algorithm implementation is continued by returning to the second step, by first removing the branch candidates that have successfully become branches so that they get a new list of prospective branches.

3.4. Bagging

The Bootstrap Aggregating Method, also called bagging is a machine learning ensemble meta-algorithm method designed to improve the stability and accuracy of machine learning algorithms in statistical classification and regression [14]. This reduces variance and helps avoid overfitting. Although usually applied to the decision tree method, Bagging can be used with all types of methods. Bagging is a special case of the model averaging approach.

3.5. Boosting

Boosting is a ensamble learning machine to reduce data bias and too many variants [15]. Boosting combines several models into a strong learner. The boosting method can improve the performance of an algorithm and can handle class imbalances in a classification process. In general, the Boosting method can provide increased accuracy in the classification or prediction process used by increasing the combination of a model, with the prediction or classification results used is the model that has the best performance.

4. TEST AND ANALYSIS RESULT

In this section, a more in-depth exploration of the review of the comparison of the K-Nearest Neighbor (KNN) algorithm and the Classification and Regression Tree (CART) is carried out. In testing, all scenarios were carried out by the evaluation technique of the 10 folds cross validation model.

4.1. Classification Performance Measures

Tests were carried out on six research scenarios. In each scenario, a performance comparison is carried out. To measure performance, 3 classifier measures are used, namely accuracy, precision, and recall.

The evaluation results in the confusion matrix from the K-Nearest Neighbor (KNN) algorithm that have been carried out, namely trying 21 times using cross validation of 10 folds starting with a value of K = 1 to a value of K = 21. From these results the best performance is at a value of K = 4 The results of the K-Nearest Neighbor (KNN) algorithm experiment are shown in Table III.

TABLE 3. Comparison Experiment K Value KNN

K Value	Accuration	Class Error	Recall	Precission
1	85.1%	14.8%	85.1%	85.1%
2	83.6%	16.3%	83.6%	85.1%
3	87.6%	12.4%	87.6%	88.4%
4	**88.6%**	**11.4%**	**88.6%**	**89.6%**
5	86.8%	13.1%	86.8%	88.2%
6	88.6%	11.4%	88.6%	89.6%
7	85.6%	14.3%	85.6%	87.8%
8	88.1%	11.9%	88.1%	89.5%
9	85.6%	14.3%	85.6%	87.5%
10	88.1%	11.9%	88.1%	89.6%
11	86.4%	13.6%	86.4%	88.4%
12	87.3%	12.6%	87.3%	89.3%
13	85.4%	14.6%	85.4%	87.8%
14	86.6%	13.3%	86.6%	88.9%
15	84.6%	15.3%	84.6%	87.4%
16	86.4%	13.6%	83.69%	91.79%
17	83.1%	16.8%	81.41%	90.92%
18	85.9%	14.1%	81.91%	90.91%
19	83.6%	16.3%	79.55%	89.84%
20	85.1%	14.8%	81.26%	90.41%
21	82.4%	17.6%	78.01%	89.10%

Next, the evaluation results in the confusion matrix from the Classification and Regression Trees (CART) algorithm that have been done previously, namely trying 11 times using cross validation of 10 folds starting with the Minimum number of observations 0 - 10. The results of the best Minimum number of observations is the best performance of the Classification And Regression Trees (CART) algorithm. The experimental results of the Classification and Regression Trees (CART) algorithm based on the Minimum number of observations are shown in Table IV.

TABLE 4. Comparison Of The Experimental Results Of The MNO Value

MNO Value	Accuration	Class Error	Recall	Precission
0	92.8%	7.1%	92.8%	93.1%
1	**92.8%**	**7.1%**	**92.8%**	**93.1%**
2	92.3%	7.6%	92.3%	92.5%
3	92.1%	7.9%	92.1%	92.3%
4	91.6%	8.4%	91.6%	91.7%
5	90.8%	9.1%	90.8%	91.0%
6	91.3%	8.6%	91.3%	91.6%
7	92.3%	7.6%	92.3%	92.7%
8	92.6%	7.4%	92.6%	92.8%

9	92.6%	7.4%	92.6%	92.8%
10	92.1%	7.9%	92.1%	92.2%

Based on Table V, it shows the results of the evaluation metrics of accuracy, precision and recall of the two classification algorithms being compared. The results of each of the accuracy, precision and recall of the two algorithms show the best results in the same scenario, namely scenario 6, using the Classification And Regression Trees (CART) boosting algorithm.

TABLE 5. Performance Comparison Of KNN And CART

No.	Preprocessing	Accuration	Precission	Recall
1.	KNN	88.6%	89.6%	93.1%
2.	KNN Bagging	86.4%	87.4%	**93.1%**
3.	KNN Boosting	85.6%	85.9%	92.5%
4.	CART	92.8%	93.1%	92.3%
5.	CART Bagging	93.5%	93.5%	93.5%
6.	CART Boosting	94.0%	94.2%	94.0%

With an accuracy of 94.0%, the best scenario can be displayed in more detail with the confusion matrix shown in Table VI.

TABLE 6. Best Scenarios Confusion Matrix Results

No.	Actual Class				Recall Class
	Very Low	High	Low	Middle	
Very Low	45	0	5	0	90.0%
High	0	101	0	1	99.0%
Low	0	0	122	7	94.6%
Middle	0	2	9	111	91.0%
Class Precision	100.0%	98.1%	89.7%	93.3%	

The algorithm performance comparison based on the accuracy value in each scenario has a significant difference. It can be seen that the accuracy value of the Classification And Regression Trees (CART) algorithm for each the scenario is always better than all scenarios in the K-Nearest Neighbor (KNN) algorithm. The lowest accuracy is obtained from scenario III, namely the K-Nearest Neighbor (KNN) Bagging algorithm of 85.6% and the best accuracy is obtained from scenario VI, namely the Classification and Regression Trees (CART) Boosting algorithm of 94.0%. This means that boosting can provide an advantage, more effective and accurate classification than the standard method.

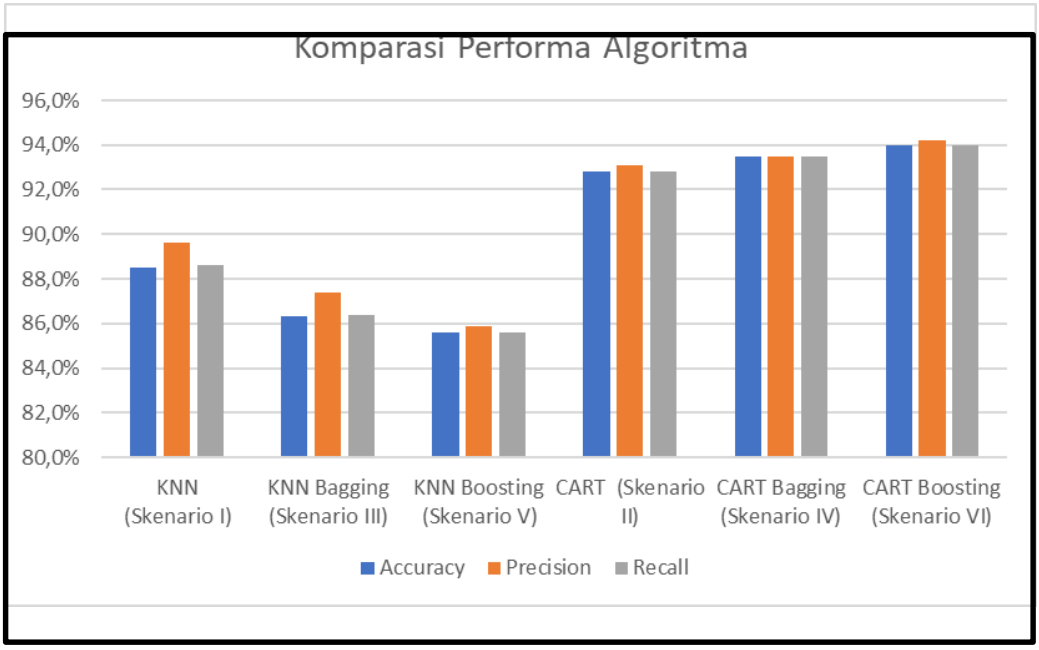

FIGURE 1. Performance Compariosn Alogaritthm.

The algorithm performance comparison based on the precision value in each scenario has a significant difference. It can be seen that the accuracy value of the Classification and Regression Trees (CART) algorithm in each scenario is always better than all scenarios in the K-Nearest Neighbor (KNN) algorithm. The lowest precision is obtained from scenario V, namely the K-Nearest Neighbor (KNN) Boosting algorithm at 85.9% and the best precision is obtained from scenario VI, namely the Classification and Regression Trees (CART) Boosting algorithm of 94.2%. Precision shows the results of the accuracy between the requested information and the results, so that the classification results of Classification and Regression Trees (CART) Boosting in the 6th scenario the accuracy of prediction with the actual class gets the best results compared to other scenarios.

The algorithm performance comparison based on recall value in each scenario has a significant difference. It can be seen that the recall value of the Classification and Regression Trees (CART) algorithm in each scenario is always better than all scenarios in the K-Nearest Neighbor (KNN) algorithm. The lowest recall is obtained from scenario V, namely the K-Nearest Neighbor (KNN) boosting algorithm of 85.6% and the best accuracy is obtained from scenario VI, namely the Classification and Regression Trees (CART) Boosting algorithm of 94.0%. Recall is the result of data that can be recovered by the system. In the classification of Classification and Regression Trees (CART) Boosting scenario 6 can recover the desired data better than other scenarios.

The results in this study, between the K-Nearest Neighbor (KNN) algorithm and the Classification and Regression Trees (CART), found that the performance of the Classification and Regression Trees (CART) algorithm is better than the K-Nearest Neighbor (KNN) algorithm using the dataset used seen. of the accuracy results. The results of the evaluation in this study indicate the accuracy of the algorithm.

K-Nearest Neighbor (KNN) is 85.895% and the accuracy of the Classification and Regression Trees (CART) algorithm is 88.46%, so it can be concluded that the performance of the Classification And Regression Trees (CART) algorithm has a better performance. Previous research has concluded that the Classification and Regression Trees (CART) algorithm performs well when used in classification techniques [12].

Classification and Regression Trees (CART) are said to be better because this algorithm has advantages that the K-Nearest Neighbor (KNN) algorithm does not, namely, Classification and Regression Trees (CART) are easier to interpret, more accurate and faster in calculations, besides it can handle large data sets [10]. The large number of data sets causes the Classification And Regression Trees (CART) algorithm to have better performance than the K-Nearest Neighbor (KNN) algorithm, the

calculation process for the Classification And Regression Trees (CART) algorithm is longer than the K-Nearest Neighbor (KNN) algorithm.

4.2. The Needs of Time To Build Models

In this study, apart from the performance measures of accuracy, precision, recall that have been discussed, the time required for the algorithm to make the model is also compared. The results of the time requirement test are displayed with a bar graph in Pic. 2.

From the graph shown, the comparison between the original method and the ensemble method has a significant difference, it can be seen that all the original methods have a shorter time requirement than the ensemble method.

From the time requirements obtained, the K-Nearest Neighbor (KNN) algorithm has the shortest model formation. Meanwhile, the algorithm that takes the most time to form the model is the bagging of Classification and Regression Trees (CART) algorithm.

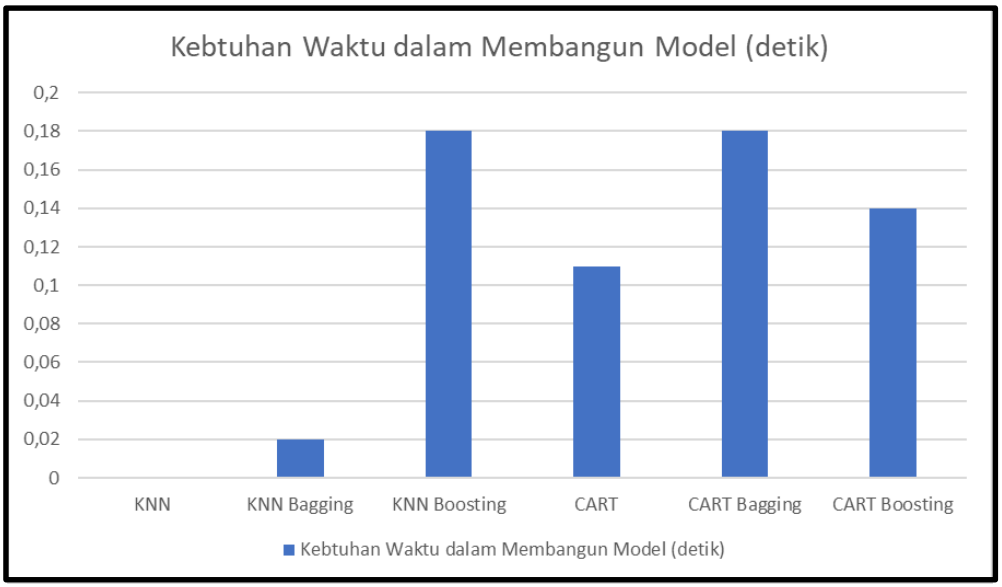

FIGURE 2. The Needs Of Time To Build Models.

5. CONCLUSION

Based on the results of the discussion of this study it can be concluded that the K-Nearest Neighbor (KNN) algorithm and the Classification and Regression Tree (CART) as in the previous research, also combined with Bagging and Boosting in scenarios II and III for K-Nearest Neighbor (KNN) and Scenario V and VI for Classification and Regression Tree (CART). From this combination, it is proven that it can improve performance measured from accuracy, precision, and recall. With the best results obtained in scenario VI which applies the Classification and Regression Tree (CART) Boosting algorithm.

REFERENCES

[1] B. Yazan, "The Qualitative Report - Three Approaches to Case Study Methods in Education: Yin, Merriam, and Stake," *Teach. Learn.*, vol. 20, no. 2, pp. 134–152, 2015.

[2] J. Walker, "Psychology: The Science of Mind and Behaviour (4th edition)," *Nurse Educ. Today*, vol. 22, no. 6, pp. 507–508, 2005, doi: 10.1054/nedt.2002.0766.

[3] S. Roush, *The difference between knowledge and understanding*, no. June. 2017.

[4] M. Grochowski and N. Jankowski, "Comparison of Instance Selection Algorithms II. Results and Comments," pp. 580–585, 2010, doi: 10.1007/978-3-540-24844-6_87.

[5] H. M. Truong, "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities," *Comput. Human Behav.*, vol. 55, pp. 1185–1193, 2016, doi: 10.1016/j.chb.2015.02.014.

[6] H. A. Rosyid, "ADAPTIVE SERIOUS EDUCATIONAL GAMES USING," 2018.

[7] "Tangerang - Indonesia," 1978.

[8] J. S. Informasi and F. Teknik, "Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika," *J. Gamma*, no. September, pp. 130–134, 2014.

[9] B. P. Keputusan, A. Shoddiq, B. Asmoro, W. Sakti, G. Irianto, and U. Pujianto, "Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa," vol. 4, no. 2, pp. 84–89, 2018.

[10] F. E. dan I. Z. Pratiwi, "Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara," *J. Sains Dan Seni Pomits*, vol. 3, no. 1, pp. 2337–3520, 2014.

[11] T. A. Wijaya, "Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016," *Fik*, vol. 1, no. 1, pp. 1–2, 2016, doi: 10.1021/jf901375e.

[12] H. T. Kahraman, S. Sagiroglu, and I. Colak, "The development of intuitive knowledge classifier and the modeling of domain dependent data," *Knowledge-Based Syst.*, vol. 37, pp. 283–295, 2013, doi: 10.1016/j.knosys.2012.08.009.

[13] D. Alverina, A. R. Chrismanto, and R. G. Santosa, "Perbandingan Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 2, p. 76, 2018, doi: 10.14710/jtsiskom.6.2.2018.76-83.

[14] L. Breiman, J. Friedman, C. J.Stone, and R. A. Olshen, "Classification Algorithms and Regression Trees," *Wadsworth Stat.*, pp. 246–280, 1984.

[15] T. Report, A. Classifiers, L. Breiman, and C. Berkeley, "Bias, variance, and arcing classifier," no. April, 1996.