ENSEMBLE ANALYSIS OF THE STUDENT'S LENGTH OF STUDY AT UNIVERSITY OF KLABAT

Niel Ananto
Faculty of Business and Economic, Klabat University
niel@unklab.ac.id

Abstract

The purpose of this study is to classify the student's length of study based on the status of graduating on time or not on time based on several independent variables observed, namely gender, Grade Point Average (GPA), place of residence, type of parent's occupation and school origin. The statistics used in this study is non-parametric statistics with a classification analysis method. The classification analysis is to find a training set model of the training set that distinguishes records into appropriate categories or classes. The method used is classification using ensemble techniques. The basic principle of the ensemble method is to develop a set of models from training data and combine a set of models to determine the final classification. The final classification is based on the largest collection of votes from a combination of a set of models. To get the best combination of models, the ensemble method enables the use of several different classification models. The ensemble method used in this study is Bagging and Boosting.

Keywords: Ensemble Analysis, Classification, Bagging, Boosting, Students Length of Study

INTRODUCTION

The development of databases is currently growing very rapidly, especially data in the field of education. From a lot of data, if not used, it will only become a collection of useless data. Every existing information becomes an important thing to determine every decision in a certain situation. This causes the provision of information to become a means to be analyzed and summarized into a knowledge of useful data when a decision is made. Berry (2004). Knowledge of data on information alone is not sufficient to make a decision. An analysis of any existing data is also needed to obtain consideration from the available information. By using data mining, each data collection or warehouse can provide important knowledge which is very valuable information for an organization, such as an educational organization. In educational organizations, an information system can be used to obtain information that supports each activity in making a decision. Vedayoko (2008). Higher education institutions need to detect student behavior so that the factors that cause the failure of students who graduate on time or not on time can be identified according to the predetermined study period. To find out this problem, an analysis of data mining is needed.

Data mining is a problem solving by analyzing the data presented in the database. Also, data mining is also used to determine data patterns, where each pattern has its own characteristics that can provide important information from the data. Witten *et.al.* (2011). Data mining can be interpreted as various branches of knowledge that come together, consisting of database systems, statistics, machine learning, visualization, and knowledge information. Data mining

  | ISSN: 2721-690X, E-ISSN: 2722-726X

has been successfully applied in various fields of science, such as business, bioinformatics, genetics, medicine, education, and so on (Jiawei Han, 2016).

Some of the techniques often used in data mining are clustering, association, estimation, and classification. In the field of machine learning, classification techniques are often used for various things, including prediction of the student study period, classification of disease types, predicting fraud in credit card transactions, and many other things that can be helped by using classification techniques Pristyanto *et.al.* (2017). Classification is the process of finding a model or pattern that can describe and differentiate classes in a dataset. The goal is that the model can be used to predict objects with unknown class labels. The model is based on training data analysis. The model from the classification results can be used to predict future data trends (Toms, 2014).

RESEARCH METHOD

The methodology used in this study for the classification of student study periods is shown in Figure 1. The research dataset was obtained from the information available in the database of the Klabat University academic section with the tributes of student ID number, student achievement index in semesters one to four, the number of credits taken in semester one. Up to four and the length of study while the sample taken was 1,717 students in the 2016 to 2018 graduation years.

Flow of Research Data Analysis

Here are the defining attributes and class labels for each data. The input data is preprocessed so that the data can be used in the Bagging and Boosting method. After that, the training data and testing data were randomly selected. Training data is data used to create a classification model in both the Bagging and Boosting methods. The classifier model generated by the training data is then used to predict the class from the testing data so that the class label testing data is not used. The prediction results in the form of class labels will be compared with the actual class labels to calculate accuracy. The prediction results of the two methods will be analyzed and seen which method is better able to predict.

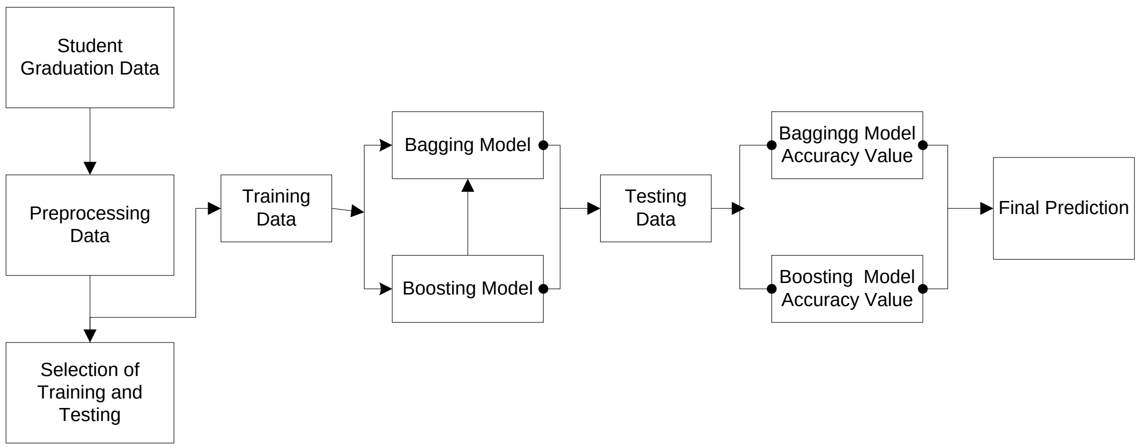

Figure 1
Research methodology for prediction of pass timeliness
Source: Witten et.al., 2011

  | ISSN: 2721-690X, E-ISSN: 2722-726X

Pre-processing Data

Data preprocessing is a step carried out to improve the input data so that it fits the data format used in the classification method. In this study, several variables that have two values will be simplified to 1 or 0. For example, the male gender variable is changed to 1, and the female becomes 0. Likewise, for class labels containing on time or not on time. On-time will be changed to 1 while not on time will be changed to 0.

Bagging Techniques

Bagging or bootstrap aggregating is a method of machine learning that is built in an ensemble for stability and accuracy in both classification and regression. The concept of an ensemble with bagging is carried out by combining many estimated values into one estimated value by using a sample in the form of bootstrap to generate random data samples that will be used as a training process for each tree. In general, bagging will use a voting mechanism to be combined as the main classifier of the resulting final model Zhou (2012). The basic idea of the Bagging method is to use random resampling with returns to the original dataset in order to obtain a new dataset. The new dataset has the same size as the training data by taking a random sample of size n with the return of the training data (bootstrap sample Sk from DK). The new dataset is then used to generate a multi-version classification tree. The classification tree from each version is then combined to obtain a final prediction (Breiman *et.al.*1984).

Based on the name, it can be estimated that there are two main stages in this analysis, namely bootstrapping, which is nothing but a sampling from the sample data held (resampling), and aggregating, which is to combine many estimated values into one estimated value. Thus, the process of making a bagging guess using a tree is as follows:

1. a. Drag a random sample of size n from the training data set (bootstrap step).
   b. Arrange the best trees based on the data obtained in a.
   c. Repeat steps a-b k times, so you get k random trees
2. Estimate the aggregate based on the k trees, for example, using the majority vote for the classification case or the average for the regression case.

Maimon *et.al.* (2018). Mentioned that in addition to using the concept of a majority vote to combine estimates from many trees, it is also possible to use the sum of the estimated odds for each class. It is important to remember that the guess produced by a tree can be an opportunity value. In the illustration above, it can be stated how many chances of "Yes" and how many for "No". From each tree, the probability of each response category can be calculated and then added up. The category with the greatest number of chances is the category that is the final guess.

Tibshirani (2008) states that the summation can be done by giving a certain weight to the probability value of each tree, for example, using the entropy value. The use of bagging is very helpful, especially in overcoming the instability of the classification tree and single regression, as previously mentioned. Breiman (1996) explained that the bagging process could reduce the standard error generated by a single tree. This can be clearly seen because by doing the average, for example, the variety of assumptions will be reduced, while the level of bias in allegations is not affected.

In addition, Larose (2006). Noted that in the many data sets he tried, bagging was able to reduce the misclassification rate in classification cases. This of course does not apply as a whole. Noted several cases where the allegation of bagging had a greater range of suspicions or a greater bias. This occurs, among others, in cases with very unbalanced response variable categories and also in conditions of explanatory variables whose distribution has a high degree of elongation. Regarding how many bootstrap iterations are required, a study from Breiman (1996). Showed that using bootstrap repetitions 50 times for classification cases and 25 times for regression cases can provide satisfactory results. Larose (2006). Disclosed the results of their study that to obtain good bagging results, it is not necessary to take samples with recovery (sampling with replacement) repeatedly, but it can be done without recovery (sampling without replacement) if the sample size is large enough.

Adaptive Boosting Technique (Adaboost)

Boosting is one of the popular methods used in machine learning. Boosting itself is designed for problems related to classification and is applied to weak classifiers. Adaptive Boosting (Adaboost) is a boosting algorithm for use with classifiers. Adaboost can improve the accuracy of various classification methods such as Decision stumps, Decision trees, Multi-Layer perceptrons, and Support Vector Machines (SVM). Adaboost is a method that combines standard classifiers where it is iteratively constructed from a weighted resampling sample, with Weights are adjusted adaptively at each step to give increased weight to misclassified cases. The following are the advantages of using AdaBoost, according to (Larose, 2006).

Adaboost works by giving more weight to objects that are not prsoperly classified by a weak classifier, which is denoted by $h(x_i)$ and then these weak classifiers will be combined to form a strong classifier (final classifier), which is denoted by $H(x_i)$. This method begins with initializing the initial weights of the training data $\{(x_i, y_i, \dots, (x_N, y_{N)}\}$ where $Y \epsilon \{-1,1\}$ each object will be given the same weight $w_t(i)$ If the training data consists of N objects, then the initial weight for each object is $\frac{1}{N}$, then weights are used for resampling the data at each subsequent step, depending on the degree of misclassification of classifiers made in the previous step.

The following is the classification error formula in equation (1), which is denoted by ε_t.

$$\varepsilon_t = \sum_{i=1}^{n} w_t(i)\, \mathbf{I}\big(y_i \neq h_t(x_i)\big) \tag{1}$$

Weights are denoted by w_t and $\mathbf{I}\big(y_i \neq h_t(x_i)\big)$ is an indicator function with a value of 1 if $y_i \neq h_t(x_i)$ and has a value of 0 if $y_i = h_t(x_i)$ then calculates the weighted voting with the formula presented in equation (2).

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \tag{2}$$

Furthermore, the weighting votes will be calculated α_t on the weak classifier $h_t(x)$, which then α_t used to update the weights. In the next step, the formula is presented in equation (3).

$$w_{t+1}(i) = \frac{w_t(i) \exp(-\alpha_t\, y_i h_t(x_i))}{Z_t} \tag{3}$$

Where,

$$\exp(-\alpha_t\, y_i h_t(x_i)) = \begin{cases} < 1 & y_i = h_t(x_i) \\ > 1 & y_i \neq h_t(x_i) \end{cases} \tag{4}$$

$$Z_t = \sum_{i=1}^{N} w_t(i) \exp\big(-\alpha_t\, y_i h_t(x_i)\big)$$

In equation (4) it can be seen that the weighting of training data that is incorrectly classified will have a value of more than one while the weighting of training data that is correctly classified will be less than one. In equation (4) Z_t is a normalized constant so that $w_{t+1}(i)$ will be the distribution. The Adaboost algorithm will generate weak classifiers by practicing the next lesson based on the errors previously obtained. After the above steps continue until the T iteration, a strong classifier will be produced with the formula presented in equation (5)

$$H(x_i) = sgn(\sum_{t=1}^{T} \alpha_t h_t(x_i)) \quad (5)$$

In equation (5) $H(x_i)$is a combination of classifiers which is calculated as the sum of the weighted voting signs of T steps. Adaboost must be run for a long time at least as many as 1,000 steps, in order to obtain an increasingly convergent error rate. In this study, the steps taken in the Adaboost algorithm are as many as 1,000 steps; this is based on the opinion. (Vanwesel, 2005).

Boosting

The boosting algorithm is an iterative algorithm that gives different weights to the distribution of training data at each iteration. Each boosting iteration adds weight to the misclassified examples and decreases the weights to the correct classification samples, thereby effectively changing the distribution of the training data Machova (2006). Boosting method (AdaBoost) proposed Berry (2004). Selective costing ensemble can be a more effective solution to the class imbalance problem and allows to improve the identification of difficult minority classes and maintain the classification ability of the majority class. The approach to overcoming this problem can be made by several methods Weiss (2017). Namely over-sampling, under-sampling, and cost-sensitive. The boosting technique scheme can be seen in Figure 2.

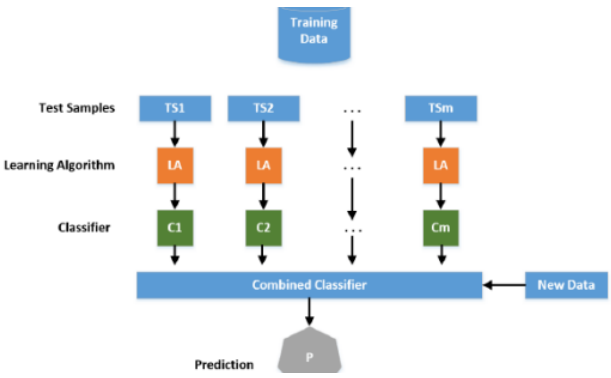

Figure 2
Schematic of the boosting technique
Source: Zhou, 2012

Calculation of Accuracy

Evaluation is the process of testing the performance of the classification algorithm used. In general, the classification algorithm performance evaluation. using confusion matrix Max (2007). Evaluation with the confusion matrix will produce accuracy, sensitivity, specificity, and G-mean values. Accuracy in classification is the percentage of accuracy of data records that are classified correctly after testing the classification results Han (2006).The specificity of the proportion of negative cases correctly identified. Recall or sensitivity is the proportion of positive cases correctly identified Powers (2011). In comparison, the geometric mean (g-mean) is one of the most comprehensive measures to evaluate the performance of classification algorithms, especially in the problem of class imbalance in the dataset. Gmean can show the overall accuracy of minority class accuracy and majority class accuracy Pristyanto (2018). The following is the equation for calculating accuracy, specificity, sensitivity, and G-mean (Han, 2006).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad Sensitifity = \frac{TP}{TP+FN}, \quad Specitifity = \frac{TN}{TN+FP}$$

$$G - Mean = Sensitifity * Specitifity$$

$$Total\ accuracy\ rate\ (1 - APER) = \frac{TP+TN}{N} \times 100\%$$

Table 1
Confusion Matrix for Accuracy

Actually Class	Prediction Class	
	On-time	Not on time
On-time	TP	FN
Not on time	FP	TN

Classification and Regression Tree

Classification and Regression Tree (CART) is one of the decision tree algorithms. CART was developed to perform classification analysis on response variables, both nominal, ordinal, and continuous. CART generates a classification tree if the response variable is categorical and generates a regression tree if the response variable is continuous. The main objective of CART is to obtain an accurate data set as a characteristic of classification Timofeev (2004). According to CART has several advantages, namely, the results are easier to interpret, more accurate, and faster in the calculation; besides that CART can be applied to data sets that have large numbers, very many variables with a mixed variable scale through the binary sorting procedure. (Lewis, 2000).

RESULTS AND ANALYSIS

Data Selection

Data Selection data used came from the Academic Information System database at Klabat University. The attributes used are student identification number, grade point one to four semesters, number of credits taken in semesters one to four and length of study.

Pre-processing / Cleaning

Four data were deleted because there were empty values in the attribute. Data that has gone through the cleaning stage can be seen in Table 2.

  | ISSN: 2721-690X, E-ISSN: 2722-726X

Table 2
Cleaning data

No	NIM	SKS Semester				IPS				Masa Studi (Semester)
		1	2	3	4	1	2	3	4	
1	Students 205	9	-	10	13	2,17	-	1,85	1,96	13
2	Students 101	-	-	16	23	-	-	3,5	3,33	4
3	Students 61	16	3	10	8	2,53	0,00	2,25	2,94	13
4	Students 238	16	9	-	21	3,03	3,61	-	2,81	14

Transformation

Data is converted into an appropriate form so that the data mining process can be carried out. In the length of the study period, it is determined that for 8 and 9 semesters, the class is classified as "On time" which is written as classification class "A" and for those who are above 9 semesters are included in the "Not on time" class which is written in the class of classification "B". The data that has been transformed is divided into two parts, namely, training data and testing data. The training data amounted to 146 data consisting of students from class 2016-2018 and testing data totaling 49 students.

Algorithm CART with Bagging Techniques

The training data is divided into 10 bagging, where each data on the bagging is taken randomly from the initial training data and the amount is the same as the amount of data in the initial training data. For each bagging, the CART algorithm is calculated which then produces one decision tree for each bagging. In the data mining process, using the CART algorithm with this bagging technique produces 10 different decision trees. After obtaining 10 decision trees, the next step is to classify the testing data into each bag that contains the CART algorithm. Determination of the final classification by voting uses the class classification that appears the most in the testing data in the 10 bagging. The classification results can be seen in Table 3.

Table 3
The results of the testing data classification using the CART algorithm with the bagging technique

No	NIM	Bagging										Prediction class Bagging
		1	2	3	4	5	6	7	8	9	10	
1	Students 3	B	B	A	B	B	A	B	B	B	B	B
2	Students 4	B	B	A	B	B	A	B	A	B	B	B
3	Students 6	B	B	B	B	B	B	B	A	B	B	B
…	…	…	…	…	…	…	…	…	…	…	…	…
49	Students 105	A	A	A	A	A	B	A	A	A	A	A

  | ISSN: 2721-690X, E-ISSN: 2722-726X

Algorithm CART with Boosting Technique

First of all, the training data is weighted with the initial weight $D_1(i)\frac{1}{146} = 0.006849$ Then the data is taken randomly into boosting 1 which is taken from the training data. Boosting data 1 can be seen in the Table 4.

Table 4
Training Data

No	NIM	IP Semester				SKS Semester				Classification
		1	2	3	4	1	2	3	4	Class
73	Students 211	3,48	2,90	3,53	3,58	21	24	20	24	B
98	Students 23	2,57	3,17	3,02	3,33	21	21	23	24	A
38	Students 219	2,43	2,55	2,64	3,05	21	19	22	21	B
8	Students 11	2,71	1,50	2,94	2,45	21	21	16	20	B
25	Students 201	2,86	2,93	3,65	3,31	21	21	24	24	B
…	…	…	…	…	…	…	…	…	…	…
130	Students 233	2,36	1,78	3,03	2,75	21	18	15	24	B

Furthermore, the CART algorithm is calculated on boosting 1, and the boosting 1 decision tree is obtained. Then the initial training data test is carried out on the decision tree produced by the CART algorithm on boosting 1. The results of the training data classification on boosting 1 can be seen in Table 5.

Table 5
Results of training data classification using CART algorithm on boosting 1

No	NIM	IP Semester				SKS Semester				Actual	Classification
		1	2	3	4	1	2	3	4		
1	Students 1	2,79	2,98	3,48	3,00	21	21	21	24	B	B
2	Students 2	2,79	2,55	3,71	3,23	21	21	19	24	B	B
3	Students 3	2,36	1,17	1,65	2,91	21	18	13	16	B	B
…	…	…	…	…	…	…	…	…	…	…	…
146	Students 11	3,30	2,78	3,25	3,56	22	18	14	17	B	B

After obtaining the results of the training data testing. Then calculate the error value (ε_t) from the predicted class results that are not the same as the actual class The results of testing the training data using the CART algorithm results on boosting 1, there are 42 different data. $\varepsilon_t = \frac{42}{146} = 0.287671$ then determine the alpha value on boosting 1 $\alpha_t = \frac{1}{2} ln\left(\frac{1-0.287671}{0.287671}\right) = 0.453361$ After getting the Alpha value on boosting 1, then updating the weight of each data in the training data with the following equation $D_{t+1}(i) = \frac{D_t(i)}{Z_t} X \begin{Bmatrix} \exp(-\alpha_t)\ if\ h_t(x_i) = y_i \\ \exp(-\alpha_t)\ if\ h_t(x_i) \neq y_i \end{Bmatrix}$

The training data weight updates can be seen in Table 6.

  | ISSN: 2721-690X, E-ISSN: 2722-726X

Tabel 6
Pembaruan bobot setiap data pada data training setelah boosting 1

No	NIM	IP Semester				SKS Semester				$D_t(i)$	$D_{t+1}(i)$
		1	2	3	4	1	2	3	4		
1	Students 1	2,79	2,98	3,48	3,00	21	21	21	24	0,00685	0,004808
2	Students 2	2,79	2,55	3,71	3,23	21	21	19	24	0,00685	0,004808
3	Students 3	2,36	1,17	1,65	2,91	21	18	13	16	0,00685	0,004808
…	…	…	…	…	…	…	…	…	…	…	…
146	Students 11	3,30	2,78	3,25	3,56	22	18	14	17	0,00685	0,004808

Then the data is taken randomly to become boosting 2, which is taken from the training data, but the data that has a greater weight has a greater chance of being selected as well. Furthermore, repeating the same process as boosting 1 for 10 boosting so that 10 decision trees are obtained from the CART algorithm and 10 alpha values generated from each training data test for each boosting. After obtaining 10 boosting, it will be tested using data testing for each CART algorithm on each boosting that has been found. The results of testing data on the CART algorithm with the boosting technique can be seen in Table 7.

Table 7
Results of testing data testing on the CART algorithm with the boosting technique

No	NIM	Bagging										Prediction class
		1	2	3	4	5	6	7	8	9	10	Bagging
1	Students 3	B	B	B	A	B	A	B	B	B	B	B
2	Students 4	B	B	B	A	B	A	B	A	B	B	B
3	Students 6	B	B	B	A	B	B	B	A	B	B	B
…	…	…	…	…	…	…	…	…	…	…	…	…
49	Students 214	A	A	A	A	A	B	A	B	B	A	A

The final classification of the boosting technique uses the following equation.

$$H(X) = Sign\left(\sum_{t=1}^{T} \alpha_t h_t(X)\right)$$

The classification that results in A is positive, and B is negative—then multiplied by the alpha value of each boosting. Then add up all the products. If the final result H (x) is positive, then the final classification is worth A; if the final result H (x) is negative, then the final classification is worth B. The final classification of Boosting can be seen in Table 8.

Table 8
Final classification of CART algorithm with boosting technique

No	NIM	$\alpha_t H_t(X)$										H(X)	Prediction Class Boosting
		1	2	3	4	5	6	7	8	9	10		
1	Students 3	-0,45	-0,68	-0,39	0,56	-0,58	-0,68	-0,66	-0,40	-0,58	-0,45	-4,30	B
2	Students 4	-0,45	-0,68	-0,39	0,56	-0,58	-0,68	-0,66	-0,40	-0,58	-0,45	-4,30	B
3	Students 6	-0,45	-0,68	-0,39	0,56	-0,58	-0,68	-0,66	-0,40	-0,58	-0,45	-4,30	B
4	Students 8	-0,45	-0,68	-0,39	0,56	-0,58	-0,68	-0,66	-0,40	-0,58	-0,45	-4,30	B
5	Students 11	-0,45	-0,68	-0,39	0,56	-0,58	-0,68	-0,66	-0,40	-0,58	-0,45	-4,30	B
…	…	…	…	…	…	…	…	…	…	…	…	…	B
49	Students 214	0,45	0,68	0,39	0,56	0,58	0,68	0,66	-0,40	-0,58	0,45	3,46	A

Evaluation is done using confusion matrix method. The following confusion matrix from the classification results in testing data can be seen in Table 9.

Table 9
Confusion matrix classification results on testing data

	TP	FN	FP	TN
CART Algorithm	10	6	4	29
CART Algorithm with Bagging Techniques	10	6	3	30
CART Algorithm with Boosting Technique	13	3	3	30

The comparison of the evaluation results from the CART algorithm, the CART algorithm with the bagging technique, and the CART algorithm with the boosting technique using the confusion matrix can be seen in Figure 3.

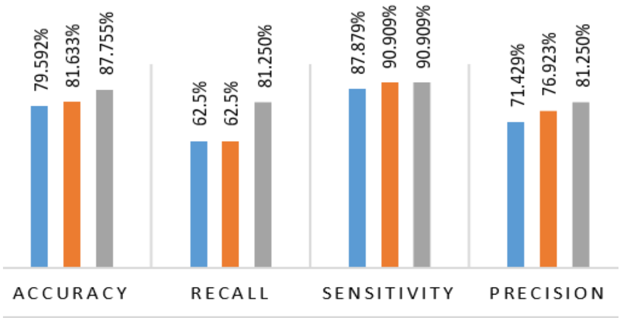

Figure 3
Percentage evaluation using confusion matrix
Source: Toms, 2014

  | ISSN: 2721-690X, E-ISSN: 2722-726X

CONCLUSION

The accuracy value of the overall classification of the student study period uses the CART algorithm of 79.592%, the CART algorithm with the bagging technique of 81.633%, and the CART algorithm with the boosting technique of 87.755%. The accuracy value for the minority class, namely class "A" or "On time" seen from the recall value, the CART algorithm classification model with the boosting technique is the best classification model among the three studied classification models to overcome the unbalanced class, namely 81.25 %.

REFERENCES

Berry G. S. (2004). *Data Mining Techniques for Marketings, Sales and Customer Relationship Management*, 2nd ed. Canada: Wiley Publishing, Inc, 2004.

Breiman L. (1984) *Classification and Regression Trees,* First. Florida: CRC Press.

Breiman L. (1996). *Bagging Predictors*, Mach. Learn., vol. 24, pp. 123–140.

Han J. (2006). *Data Mining: Concepts and Techniques*, 12th ed. San Fransisco: Morgan Kauffman.

Jiawei Han J.P., Micheline Kamber (2016) *Data mining: Concepts and techniques Transl. Tolkien* doi: 10.3726/978-3-653-01058-9/2.

Larose D.T. (2006). *Data Mining Methods and Models,* 1st. ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Lewis R. (2000). An introduction to classification and regression tree (CART) analysis," vol. 14.

Machová, K. (2006). *A Bagging Method using Decision Trees in the Role of Base Classifiers*. Acta Polytech. Hungarica, vol. 3, no. 2, pp. 121–132, 2006.

Maimon O. and L. Rokach. (2018). *Data Mining and Knowledge Discovery Handbook, 2nd ed*, no. January 2010.

Max B. (2007). *Principles of Data Mining*. London: Springer.

Powers D. (2011). *Evaluation: From Precision, Recall And F-Measure To ROC, Informedness, Markedness & Correlation,* vol. 2, no. 1, pp. 37–63.

Pristyanto I. (2017). *Hybrid Resampling to Handle Imbalanced Class on Classification of Student Performance in Classroom, Informatics Comput. Sci. (ICICoS 2017),* vol. 3, no. ICICoS, pp. 215–220.

Pristyanto I. (2018). Data level approach for imbalanced class handling on educational data mining multiclass classification, *Int. Conf. Inf. Commun. Technol. ICOIACT*, pp. 310–314.

Tibshirani R. (2008). *The Elements of Statistical Learning,* 2nd ed. New York: Springer.

Timofeev R. (2004). *Classification an Regresion Trees Theory and Application*. Humboldt University.

Toms C. (2014). *Exploration of classification using NBTree for predicting students performance, Int.* Conf. Data Softw. Eng., pp. 1–5.

Vanwezel R. (2005). *Improved Customer Choice Predictions using Ensemble Methods,* Eur. J. Oper. Res., vol. 18, no. 1, pp. 1–29.

Vedayoko L.G. (2008). *Expert System Diagnosis of Bowel Disease Using Case-Based Reasoning with Nearest Neighbor Algorithm, Sci. J. Informatics*, vol. 4, no. 2, pp. 134–142, 2017, doi: 10.15294/sji.v4i2.11770.

Weiss B. (2017.) *Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs*,Proc. First Int. Conf. Adv. Data Inf. Eng., vol. 6, no. 3, pp. 13–22.

Witten I., Frank E., and Hall M. (2011). *Data mining 2nd*, vol. 54, no. 2.

Zhou Z.H. (2012). *Ensemble Msethods Foundations and Algorithms*. New York: CRC Press.

  | ISSN: 2721-690X, E-ISSN: 2722-726X