www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
Budapest Institute

Classification Analysis of Student Graduation Timeliness Using Decision Tree and Naïve Bayes Methods

Sri Nevi Gantini[1], Besse Arnawisuda Ningsi[2], Irvana Arofah[3]
[1]*Universitas Muhammadiyah Prof. Dr. Hamka Jakarta, Indonesia*
[2,3]*Universitas Pamulang, Indonesia*
nevi_gan@yahoo.co.id

Abstract

This study aims to determine the classification of student graduation timeliness by using the Decision Tree and Naïve Bayes methods. This study uses a quantitative method, where the approach used is the classification of various attributes that affect the timeliness of student graduation. The independent variables in the classification are mostly called attributes; In this study, the attributes of school of origin, gender, area of origin, profession of parents, study program and Grade Point Average (GPA) were used. While the dependent variable or in the classification is usually called a label, in this study the label used as a decision attribute is the timeliness of student graduation. In this study, two methods were used, namely using the nave Bayes method and a decision tree (decision tree) to determine the classification of the timeliness of student graduation and to determine the level of classification accuracy. Based on the results of the analysis, it can be concluded that the classification using the nave Bayes method obtained 36 predicted data according to the actual data and 7 different predicted data from the actual data. Meanwhile, in the 42 decision tree method, the predicted data is in accordance with the actual data and there is only 1 predicted data that is different from the actual data. Decision Tree method has a lower classification error rate than the Naïve Bayes method. The level of accuracy of prediction results using the Decision Tree method is higher than the Naïve Bayes method.

Keywords
naïve bayes; decision tree; graduation punctuality

I. Introduction

In the development of the world of education, especially after the rolling reforms, new phenomena have arisen in educational institutions, which are schools that use the term Integrated Islamic Schools (Titik, 2010: 42). The school is essentially aimed at helping parents teach good habits and add good character, also given education for life in society that is difficult given at home. Thus, education in schools is actually part of education in the family, which is also a continuation of education in the family (Daulay in Ayuningsih, W. et al. 2020).

Education is the key for a nation to improve the quality of human resources and the quality of the nation. One of the important sectors in supporting the progress of a nation is the education sector. A quality generation will be produced along with the development of higher quality education. The government places education in an important position and according to its function, namely to educate the nation's life and promote general welfare. Therefore, the government sets standards such as graduate competency standards, assessment standards,

educational process standards and several other standards in order to catch up with the quality of education from other countries.

Education is the foundation of a successful career, financial freedom, the ability to think and reason critically and to make informed decisions. Without education we will be limited to perform tasks and we will be ignorant to the things that are happening in and around our surrounding, and according to Martin Luther King, a people without knowledge is like a tree without roots. For education to be of great value, curriculums should be implemented. (Philips, S. 2020)

Through various policies, the government continues to strive so that the educational goals that have been set can be achieved properly. The diversity of subjects studied and the existence of levels in schools is one example of government policy in the world of education. College is the highest level of education after high school. One of the functions of higher education is the development of capabilities and the formation of a dignified national character and civilization in the context of the intellectual life of the nation.

Every university always strives to improve its quality, including the quality of its students. One of the benchmarks for the quality of the quality of students is their academic ability which can be seen from the value of the cumulative achievement index (GPA). The quality of student quality in addition to the learning achievement of each student and student competence, universities are also required to improve the quality of graduates, namely how to produce students who can graduate in accordance with the specified time. Basically, every student hopes to be able to fulfill the allotted educational time in the sense of being able to get a bachelor's degree on time. However, in reality there are still many cases of students who have not been able to complete their education on time.

The problem of not being on time in completing education also occurs in students at the University of Muhammadiyah Prof. Dr. Hamka (UHAMKA) Jakarta. The punctuality of student graduation is one of the important indicators in the accreditation assessment of both Study Programs and Universities, including the Pharmacy Study Program, Universitas Muhammadiyah Prof. Dr. Hamka (UHAMKA) Jakarta.

The timeliness of student graduation is certainly influenced by various factors including the Grade Point Average (GPA), number of credits taken, parental occupation, study program, gender, and type of school origin and so on. Several studies have been carried out including Suniantara et al (2017) using gender, study program, thesis length, GPA, 6th semester IP, and entrance exam scores to analyze the classification of graduation time, Yuniarti et al (2020) using entry pathway factors, school origin, major background, GPA, semester academic status, to non-academic data such as parents' occupations and parents' economic conditions to analyze graduation classifications. The science of statistics continues to develop along with the development of technology and the challenges of the role of technology in the era of the industrial revolution.

Data mining is a series of activities ranging from setting goals to evaluating results. Classification method is a process to obtain a model that describes and distinguishes the class of each data and predicts the class for data whose class is not known. Naïve Bayes and Decision Tree are one of the classification techniques in data mining that classify characteristics based on their probability values. One of the advantages of this approach is that the classifier will get a smaller error value when the data set is large (Berry, 2006). In addition, according to Han and Kamber (2006) the Naïve Bayes classification is proven to have high accuracy and speed when applied to a large number of databases.

Based on the description above, it is necessary to conduct research on the classification of the timeliness of student graduation, so that a policy framework can be made in controlling the quality of PT on the side of student graduation. In this study, two methods of data

analysis were used, namely: the nave Bayes method and the decision tree to determine the classification of the timeliness of student graduation and to determine the level of accuracy of the classification.

II. Research Method

This study uses a quantitative method, where the approach used is the classification of various attributes that affect the timeliness of student graduation. In addition, this research was also conducted using the literature study method by first determining the sample frame of the research respondents.

This study uses secondary data, with reference to the data of the Student of Pharmacy Study Program, University of Muhammadiyah Prof. Dr. Hamka (UHAMKA) Jakarta. The independent variables in the classification are mostly called attributes; In this study, the attributes of school of origin, gender, area of origin, profession of parents, study program and Grade Point Average (GPA) were used. While the dependent variable or in the classification is usually called a label, in this study the label used as a decision attribute is the timeliness of student graduation.

In this study, two methods were used, namely using the nave Bayes method and a decision tree (decision tree) to determine the classification of the timeliness of student graduation and to determine the level of classification accuracy. The stages of data analysis carried out are as follows:

1. Preparation of training data

In data classification using nave Bayes and decision trees, the existing research data is divided into two categories, namely training data and testing data. Training data is used to obtain a model that is used to test the testing data that will be classified later.

2. Determination of the probability value of the number of each label

After the training data has been read, then the probability of each label will be calculated, namely in this study the respective probabilities of the label on time and label will be calculated.

3. Determination of the probability value for each attribute $P(X|Ci) = 1,2,3…n$

We know the probability of each label, then calculate the probability of each attribute on each label. In this study, the attributes used were gender, type of school, regional origin, parental occupation, study program, and GPA predicate. For each of these attributes, the probability of each label being on time and the label not being on time will be calculated.

4. Multiply the attribute probability value for each label

It is already known each attribute probability on each label, then multiply all the attribute probabilities by the probability of each label (attribute multiplication based on the testing data to be tested).

5. Comparing the probability value of each label

The probability value that has been obtained from the multiplication of each label is the label on time and the label not on time before, and then compares the results of the two labels, which one has the greater probability.

6. Conclusion

Through known comparisons, conclusions are then drawn to determine whether the testing data tested is included in the label classification on time or the label classification is not timely. If the probability value of the label being on time is greater then it is included in the label classification on time and vice versa if the probability value of the label is not on time is greater then it is included in the label classification is not on time.

Test the level of accuracy is calculated:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\%$$

$$APPER = \frac{FP+FN}{TP+TN+FP+FN} \text{ x } 100$$

III. Result and Discussion

In nave Bayes analysis, the total number of data used is 270 students, where the data is divided into two parts, namely as many as 225 students as training data and 44 students as testing data. The probability of each label as shown in table 4 below:

Table 1. Probability of graduation time

Graduation Time Attribute Label	Amount	Probability
On time	139	0.618
Not on time	86	0.382
Total	225	1

It can be seen in the table above that the probability of graduation in the on time category is 0.618 and the probability of the graduation time label being not on time is 0.382. From the probability value, it shows that the probability of the student's graduation time being on time is still greater than the probability of not graduating on time. To classify using the Naïve Bayes and Decision Tree methods because the data used is large, this research uses Rapidminer software to make it easier to do prediction calculations.

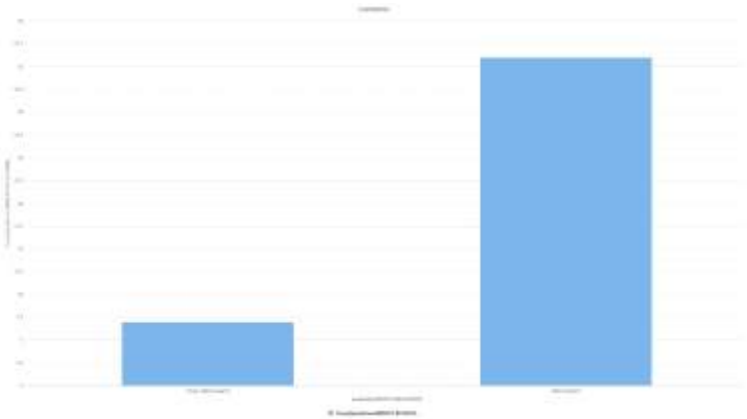

Figure 1. *Prediction Results of Testing Data MethodNaive Bayes*

Based on the figure, it is found that 36 predicted data are in accordance with the actual data and 7 predicted data are different from the actual data. After knowing the prediction results of nave Bayes calculation using rapidminer, then the accuracy level will be calculated using the confusion matrix table.

Table 2. Confusion Matrix Naive Bayes

		True Values	
		On time	Not on time
Prediction	On time	42	0
	Not on time	1	0

Based on the results of the calculations above, the precision value is 100%, which means that this value is a comparison of a true positive prediction with the overall predicted positive result. Then the recall value is 83.72%, which means that this value is a comparison of a true positive prediction with the overall positive data actually. Furthermore, in the calculation above, the accuracy results are 83.72% and the classification error rate is 16.28%.

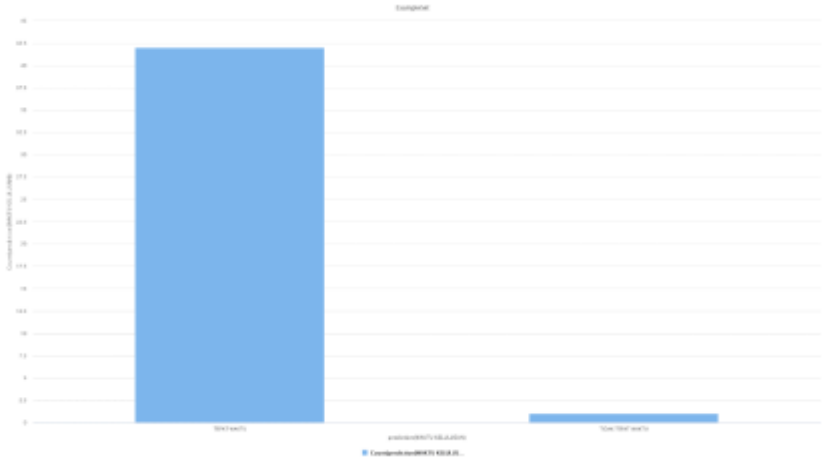

Figure 1. *Prediction Results of Testing Data Method Decision Tree*

Based on the picture, it is found that 42 predicted data are in accordance with the actual data and there is only 1 predicted data that is different from the actual data. After knowing the prediction results of Decision Tree calculations using rapidminer, then the accuracy level will be calculated using a confusion matrix table.

Table 1. Confusion Matrix Decision Tree

		True Values	
		On time	Not on time
Prediction	On time	42	0
	Not on time	1	0

Based on the results of the calculations above, the precision value is 100%, which means that this value is a comparison of a true positive prediction with the overall predicted positive result. Then the recall value of 97.67% was obtained, which means that this value is a true positive prediction comparison with the overall positive data actually. Furthermore, in the calculation above, the accuracy results are 97.67% and the classification error rate is 2.33%.

Based on the results obtained from the two methods, it is found that the Decision Tree method has a lower classification error rate than the Naïve Bayes method. The accuracy of the prediction results using the Decision Tree method is higher than the Naïve Bayes method. This means that the Decision Tree method is a better method in classifying student graduation times.

V. Conclusion

Based on the results of the analysis, it can be concluded that the classification uses the Nave Bayes method 36 predicted data were obtained in accordance with the actual data and 7 predicted data were different from the actual data. Meanwhile, in the 42 decision tree method, the predicted data corresponds to the actual data and there is only 1 predicted data that is different from the actual data.Decision Tree method has a lower classification error rate than the Naïve Bayes method. The level of accuracy of prediction results using the Decision Tree method is higher than the Naïve Bayes method. This means that the Decision Tree method is a better method in classifying student graduation times.

References

Arikunto. (2010). Suharsimi Arikunto.pdf. In Research Procedures A Practice Approach – Revised X.

Arofah, I., Ningsi, BA, & Masyhudi, L. (2020). Analysis of the factors that affect student academic achievement. scientific development media, 15(5), 4511–4522. http://ejurnal.binawakya.or.id/index.php/MBI/article/view/854

Ayuningsih, W. et al. (2020). mplementation of Islamic Education Curriculum Development in Al-Ulum Islamic School Medan. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal. P. 1033-1044.

Jiawei, H., & Micheline, K. (2006). Data Mining Concepts and Techniques. 2nd Edition. In Journal of Chemical Information and Modeling.

Kusrini, & Emha, T. (2015). Definition of Data Mining. Data Mining.

Larose, DT, & Larose, CD (2014). Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. In Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. https://doi.org/10.1002/9781118874059

Philips, S. (2020). Education and Curriculum Reform: The Impact They Have On Learning. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal. P. 1074-1082.

Rahayu, TM, Ningsi, BA, Isnurani, & Arofah, I. (2021). Classification of Student Graduation Timeliness with the Naive Bayes Method. Scientific Development Media, 15(10), 5097–5104. http://ejurnal.binawakya.or.id/index.php/MBI/article/view/1062

Sugiyono. (2013). Educational Research Methods Quantitative, Qualitative, and R&D Approaches Sugiyono. 2013. "Educational Research Methods, Quantitative, Qualitative, and R&D Approaches." Educational Research Methods Quantitative, Qualitative, and R&D Approaches. https://doi.org/10.1. In Educational Research Methods Quantitative, Qualitative, and R&D Approaches.

Xhemali, D., J. Hinde, C., & G. Stone, R. (2009). Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. International Journal of Computer Science, 4(1).

Yuliharyani, S. (2011). Decision Tree C4.5 Algorithm for Family Classification of Jamkesmas Participants Based on Poverty. Thesis for Undergraduate Program FMIPA Brawijaya University. Poor.