STUDY PROGRAM CLASSIFICATION SYSTEM INFORMATICS ENGINEERING OF UBHARA SURABAYA

[1]Wahyu Dyah Rizki Septiana, [2]Eko Prasetyo, [3]Rani Purbaningtyas, [4]Teddy Wishadi, [5]Emanuel Suprihadi

[1]Student Informatics Engineering, Bhayangkara University Surabaya
[2,3,4,5] Lecture Informatics Engineering, Bhayangkara University Surabaya

Jl. Ahmad Yani No 114 Surabaya

Email: wahyudyahrizky@gmail.com

ABSTRACT

One indicator to improve the quality of a university is the number of students who graduate on time. But the problem that often occurs at Bhayangkara University in Surabaya is the number of students entering and the number of students graduating unbalanced. Therefore this research was made to classify the period of study of students by using variables in the form of social studies semester 1-4, school origin, work status, morning / evening class status, and sex. This study aims to classify the length of time a student studies on time or late using the naïve bayes method. The results of this study indicate that the system is able to classify training data and test data on experiments conducted in each batch, the highest accuracy results are 59% and the lowest accuracy results are 56%.

Keywords: Classification, Naive Bayes, Student Study Period, Informatics Engineering

1. INTRODUCTION

Higher education is the last level for students to gain knowledge before entering the competitive world of work. At present higher education institutions are in a competitive environment. So that tertiary institutions, both public and private, are now required to have competitive advantages. Bhayangkara University in Surabaya is one of the private tertiary institutions in Surabaya with a high level of interest. With so many enthusiasts, UBHARA must improve its quality.

One indicator to improve quality is the number of students who graduate on time. But the problems that often occur at Bhayangkara University in Surabaya, the number of students entering and the number of students graduating is not balanced. This is due to the large number of students graduating with more than 4 years of study so that it affects the quality of UBHARA. The more important problem is the absence of a system that can classify the study period of students as early as possible at the Bhayangkara University in Surabaya. One method that will be applied to resolve the above problem is the Naïve Bayes method. Naïve Bayes method is one of the methods of calcification. Classification is the process of grouping things logically according to their common characteristics (Glossary). Classification techniques are approaches to carry out classification functions in data mining to classify data. This classification technique can also be used to make predictions or information that is not yet known. This certainly can be done to classify the period of study of students at the University of Bhayangkara. So the expected results are in the form of classification results which can later be used by Bhayangkara University in classifying the period of study of its students.

Based on the description above, this final project will conduct research to apply classification techniques with the Naïve Bayes algorithm. in order to estimate the classification of the study period of students of Bhayangkara University in Surabaya.

2. BASIC THEORY

Naïve Bayes algorithm to find the estimated time of student study (case study of information technology faculty of UNISBANK Semarang)[3]. The naïve bayes classification is a statistical classification that can be used to predict the probability of membership of a class. The naïve bayes classification is based on the Bayes theorem which has capabilities similar to decision trees and neural networks. The purpose of this study is to determine the prediction of the length of study of each student who has gone to college until a minimum of semester IV using the basis of

previous data. Data collection is then carried out data cleaning process. Data cleaning itself is intended to clean up problematic data which is then followed by data conversion while the data used is GPA <Gender, City of birth, School type and school city. The data is then entered into the database and used as training data. The naïve bayes process begins with the calculation of probability for each class followed by calculating the probability for each criterion and for each class the final step is to draw conclusions obtained from the comparison of criteria for each class. In this study the results obtained with an error rate of 20 records from a total of 66 testing data so the error rate was 34. The naïve bayes process begins with the calculation of probability for each class followed by calculating the probability for each criterion and for each class the final step is to draw conclusions obtained from the comparison of criteria for each class. In this study the results obtained with an error rate of 20 records from a total of 66 testing data so the error rate was 34. The naïve bayes process begins with the calculation of probability for each class followed by calculating the probability for each criterion and for each class the final step is to draw conclusions obtained from the comparison of criteria for each class. In this study the results obtained with an error rate of 20 records from a total of 66 testing data so the error rate was 34.

3. SYSTEM ANALYSIS AND DISCUSSION DESIGN

3.1 Informatics Engineering Study Program

With the development of the industrial world, both the service industry and the manufacturing industry, demands the availability of an educational forum that can meet the development of human resource needs. One forum that can accommodate these challenges is the Informatics Engineering Study Program which specializes in mastering applications.

3.2 Study period

The period of study is the time provided to complete study in a Bachelor's program for a maximum of 14 semesters (7 years). The cumulative length of study can be shorter, namely 7 or 8 semesters for the Bachelor program provided that they have completed a number of SKS specified in the curriculum of the study program for the Bachelor program with a minimum GPA of 2.00.

3.3 IP (Performance Index)

Achievement Index is the average credit score which is the final grade unit that describes the quality of the teaching-learning process each semester, or can be briefly interpreted: Large / Figures that state student achievement (success of teaching-learning process) in one semester. Or it can also mean the average credit score per semester determined based on the cumulative number divided by the number of SKS (Semester Credit Units) in one semester concerned and expressed in numbers two digits behind the formula.

3.4 Semester Achievement Index (IPS)

Is an achievement index calculation with all courses taken for each particular semester. So this calculation is done using the values obtained only in one semester.

3.5 Cumulative Performance Index (GPA)

It is a calculation of IP by combining all courses that have been taken up to a certain semester, the intention is the calculation of this value based on the accumulation of the value of all courses that have been taken up to a certain semester.

3.6 Student Status

Student status if seen from:

a) School Origin

Where the student's school comes from, this will also be one of the factors for choosing a faculty or department to be taken.

b) Job status

Students who only go to college but do not work will easily divide their time for other activities related to or not with college activities. While students who study while working will be more difficult to divide their time because the time is divided into two.

c) Morning / Night Class Status

Students can choose to attend class morning or evening. Usually students who are not working will prefer to attend morning classes, and vice versa students who go to college while working will choose to attend evening classes because they are morning for work and nights for college.

3.7 Data Mining

Data mining is a series of processes to explore the added value of a data set in the form of knowledge that has not been known manually. Data mining can also be interpreted as an activity to explore large amounts of data to find meaningful patterns and rules. One of the techniques created in data mining is how to trace existing data to build a model .

3.8   Data Mining Grouping

Data mining is divided into several groups based on the tasks that can be done, namely :

a) **Description**

Sometimes researchers and analysts simply want to try to find ways to describe patterns and trends contained in the data. Descriptions of patterns and trends often provide possible explanations for a pattern or trend.

b) **Estimate**

Estimates are almost the same as classifications, except that the target variable is more numerical than towards a category. The model is built using complete records that provide the value of the target variable as a predictive value. Then in the next review the estimated value of the target variable is based on the predictive variable values.

c) **Prediction**

Prediction is almost the same as classification and estimation. Except that in prediction the value of the results will be in the future.

d) Classification

In classification, there are target variable categories. For example, income classifications can be separated into three categories, namely high, medium and low income.

e) **Clustering**

Clustering is a grouping of records, observations or attention and form a class of objects that have similarities. A cluster is a collection of records that have similarities to one another and have dissimilarities with records in another cluster.

f) **Association**

The task of the association is to find attributes that appear at one time. In the business world it is more commonly called shopping basket analysis.

3.9 Classification

Classification is the work of assessing data objects to put them in a certain class with a number of classes that are available. In the classification there are two main tasks performed, namely the construction of the model as a prototype to be stored as memory and the use of the model to make an introduction / classification / prediction on another data object so that it is known in which class the data object is in the model that has been stored [8].

Classification is a data mining method that can be used for the process of finding a set of models (functions) that can explain and differentiate classes of data or concepts, the purpose of which is that the model can be used to predict class objects whose labels are unknown or can predict trends in data appear in the future. The classification method also aims to map data into pre-defined classes based on data attribute values .

3.10 Model in Classification

The model in the usual classification is defined as a place to receive input (training data), then be able to think about these inputs, and provide answers as outputs from the results of the thinker. The model is used to predict the class of test data. The work process in classification can be seen in Figure 3.1:

Figure 3.1 Work Process in classification In the construction of the model during the training process an algorithm is needed to build it, which is called the training algorithm. There are several training algorithms including naïve Bayes, artificial neuron networks, Support Vector Machines, and others[8].

3.11 Naïve Bayes

Naive Bayesianis a simple chance classification based on the application of the Bayes theorem with the assumption between independent explanatory variables. In this case, it is assumed that the presenceor the absence of a particular event from a group is not related to the presence or absence of other events.

$$P (= \mid Y =) = \exp X_i x_i y_i \frac{1}{\sqrt{2\pi\sigma_{ij}}} \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}$$

Naive Bayesian can be used for various purposes including document classification, spam detection or spam filtering, and other classification problems. In this case more highlighted regarding the use of the Naive Bayesian theorem for spam filtering.

3.12 Naïve Bayes for Classification

The first step to calculate the probability value is to calculate the mean value () and the value of variance () with the formula:$\bar{x} s^2$

Information :

$$\bar{X} = \frac{x_{1+} x_{2+} x_{3\dots\dots\dots\dots} x_n}{n}$$

$\bar{x}$: average count (mean)

$x_{1+} x_{2+} x_{3\dots\dots\dots\dots} x_n$: data

n: amount of data

$$S^2 = \frac{x_i - \bar{x}}{n-1}$$

Information :

s^2 : variant value

x_i : x-i value

$\bar{x}$: average calculation result (mean)

n: amount of data

For features of numeric (continuous) type in Naïve Bayes the method is:

Where, the parameters obtained from the mean value, exp = 2.72, while the value = 3.14. For parameters taken from the number of classes in the column information while parameters are taken from the results of variants and parameters obtained from the values of variants that have been rooted.$\mu_{ij} \ \pi x_i \ \sigma_{ij}^2 \sigma_{ij}$

4. SYSTEM IMPLEMENTATION AND DESCISSION

4.1 System implementation

Flowchart illustrates the flow of the system at the classification stage with the Naive Bayes method that starts from entering student data. Then the data is processed using the Naive Bayes algorithm by calculating the amount of data for each class feature, calculating the probability value of each feature. Then do the initial class probability calculation, then calculate the final class probability to find out the final value of the probability results, then the results of the greatest probability value will be displayed on the system.
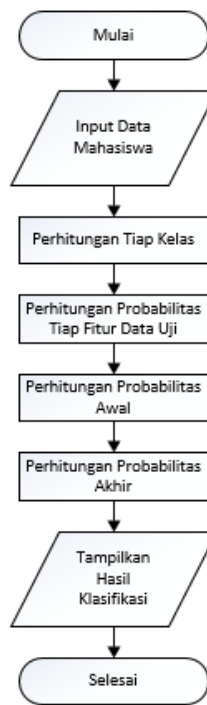
Figure 4.1.1 Classification Flowchart Stage

4.2 System Flow

The system flow is the stages that are used to calculate the naïve bayes classification, can be seen in Figure 2 as follows:

Figure 2 System Flow

4.2.1 Context Diagram

Context diagram is a general picture of the system. There is one entity that is directly connected to the system, the user. Diagram descriptionThe context can be seen in Figure 3:
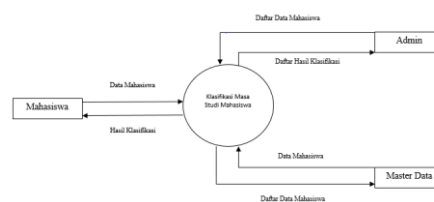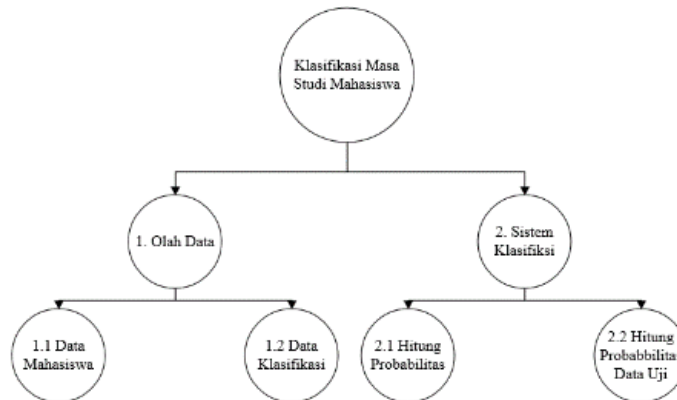

**Figure 4.2.2 Context Diagram**

4.3 Tiered Diagram

Figure 4.3.1 Tiered Diagram

Explain the flow of data in the classification system of the study period that is for processing data and classifying the study period. In the data processing function can process student data, while in the classification function can do the process of calculating the probability of training data and calculating the probability of test data.

4.4 Data Flow Diagrams (DFD)

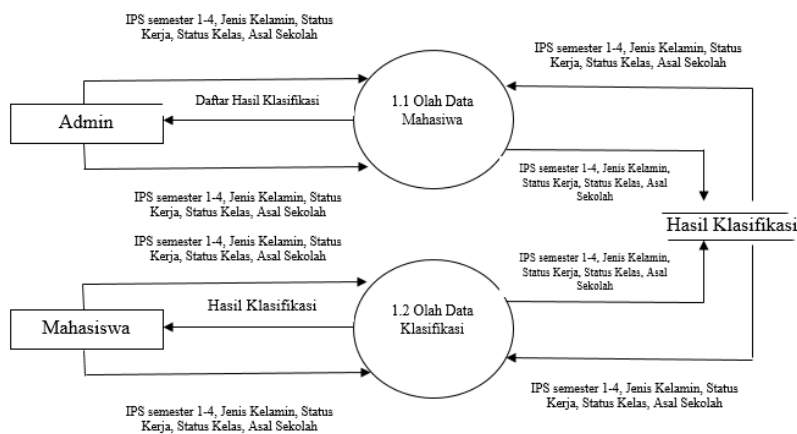### 4.4.1 DFD Level 1
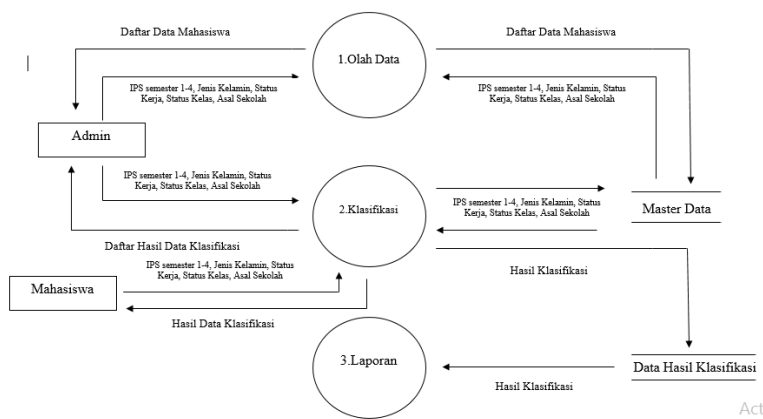
4.4.2 DFD Level 2

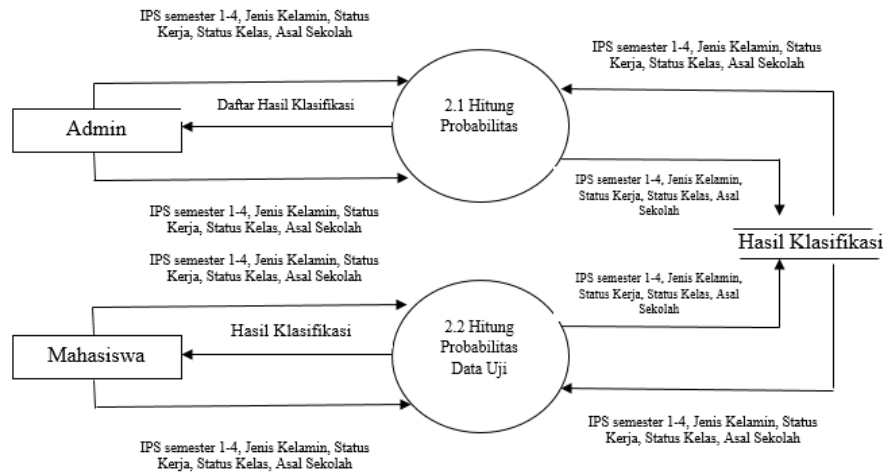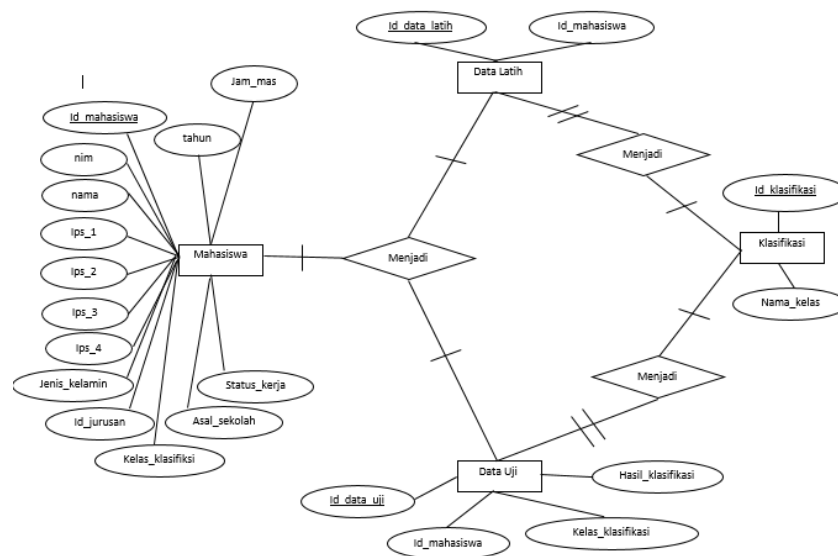**Figure 4.4.3 DFD Level 2 Process 1**

Figure 4.4.4 DFD Level 2 Process 2

Figure 4.4.5 ERD

4.5 Figure 8. Entity Relations Diagram

Figure 8 explains the relationships between entities in the classification database of the study period of students, where in the database there are three entities namely students, admin and the results of classification. Starting from the admin must log in first, then can input some student data, while one user can input one student data.

5.   TESTING AND RESULTS

Experiments or experiments conducted on the classification system of the Bhayangkara University student study period is doing the classification of students with the naive bayes method, to prove whether the results of the program are the same as the original results of the classification of the study period of Bhayangkara University students in Surabaya or even far different from the original results, therefore need to be tested and compared the results of the program with the real data. Of course in testing will be done with training data and test data

Testing is done by comparing training data and test data as follows:

a.) Testing Force 2010   : training data totaling 165 data and test data totaling 165 data

b.) Force testing 2011   : training data totaling 190 data and test data amounted to 190 data
c.) Class of 2012 testing   : training data totaling 150 data and test data totaling 150 data
d.) Class of 2013 testing   : training data totaling 125 data and test data totaling 125 data

This is done so that we can find out the parameters that have high accuracy if applying the Naive Bayes method in the classification case of the study period of students.

Based on the tests performed by each batch, the highest accuracy results obtained in the 2012 batch testing with an accuracy value of 0.59 and the lowest accuracy results in the 2011 batch testing with an accuracy value of 0.56

6. Conclusion and Suggestion
6.1 Conclusion
The conclusions obtained in this study are:

The Classification System to classify the period of study of students of Bhayangkara University in Surabaya using the Naïve Bayes method has a fairly low level of accuracy to classify the period of study of students at Bhayangkara University in Surabaya, which is 59%. In the experiments conducted on each generation, the highest accuracy value was found in the class of 2012 with an accuracy value of 59% and the lowest accuracy value was found in the class of 2011 with an accuracy value of 56%.

6.2 Suggestion
a.) The system is only able to assess or classify the existing criteria, it is hoped that further research can add new criteria.
b.) Because the highest accuracy result is only 59%, it is expected to use other methods as a comparison with this study.

REFERENCES

[1] Firdaus Mahmudy, Wayan. (2014), Classification of News Articles Automatically Using the Modified Naïve Bayes Classifier Method, Universitas Brawijaya, Malang

[2] Izmi Andini, Taghsya. (2016), Predicting the Potential of New Product Marketing with the Naïve Bayes Classifier and Linear Regression Method, Achmad Jendral Yani University, Cimahi

[3] Jananto, Arif. (2013), Naïve Bayes Algorithm for Finding Estimated Study Time for Students, Unisbank, Semarang

[4] Kusuma Dewi, Sri. (2009), Nutrition Status Classification Using Naive Bayesian Classification, Indonesian Islamic University, Yogyakarta

[5] Laily Fithri, Diana. (2016), Data Mining Model in Determining Eligibility for Housing Selection Using the Naïve Bayes Method, Muria University, Kudus

[6] Murdianingsih, Yuli. (2015), Classification of Good and Problematic Customers Using the Naïve Bayes Method, UPN Veteran, Yogyakarta

[7] Nur Fais A, Sukma. (2016), Classification of Prospective Blood Donors by the Naïve Bayes Clasifier Method, Universitas Brawijaya, Malang

[8] Prasteyo, E. (2012), Data Mining, 1st edition, Andi, Yogyakarta

[9] Saleh, Alfa. (2015), Implementation of the Naïve Bayes Classification Method in Predicting the Amount of Use of Household Electricity, Potential Main University, Medan

[10] Syarli. (2016), Naïve Bayes Method for Graduation Prediction, Universitas Brawijaya, Malang

[11] Yusra. (2016), Comparison of Final Project Classification of Informatics Engineering Students Using the Naïve Bayes Classifier Method and K-Nearest Neighhbor, UIN Sultan Syarif Kasim, Riau