Journal of Theoretical and Applied Information Technology
28th February 2017. Vol.95. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTIVE EVALUATION OF PERFORMANCE OF COMPUTER SCIENCE STUDENTS OF UNNES USING DATA MINING BASED ON NAÏVE BAYES CLASSIFIER (NBC) ALGORITHM

[1]ENDANG SUGIHARTI, [2]SAFIT FIRMANSYAH, [3]FEROZA ROSALINA DEVI

[1,2,3]Computer Science Department, Universitas Negeri Semarang, Indonesia

E-mail: [1]endangsugiharti@mail.unnes.ac.id, [2]safitfirmansyah@yahoo.co.id, [3]ferozarosalina24@gmail.com

ABSTRACT

Predictive evaluation is essential in order to map the performance of students from the Department of Computer Science of Mathematics and Natural Sciences Faculty of Universitas Negeri Semarang (UNNES), a state university in Semarang, Indonesia and graduation of students in a timely manner can also be predicted. This predictive evaluation can be seen by making a system based on the algorithm of Naïve Bayes Classifier (NBC). The data were taken from the performance of students which is the GPA from the 1st semester up to the 4th semester. The problem is how to predict the success of students in Computer Science Department of UNNES to graduate on time based on the performance of students from the 1st semester to the 4th semester? The main purpose of this research is to produce a system based on NBC algorithm that is able to predict the success of students to finish the study on time based on the performance of the students which is the GPA of the 1st semester to the 4th semester. To resolve this problem, the research divided into two stages of completion. The first stage was the literature review. This stage has been conducted by the researchers. The second stage determined the prediction of Computer Science student achievement using the method of NBC. This stage including (1) Data Collection, (2) Build a data mining system, (3) Data Processing, (4) Conducting the process of prediction, and (5) Analysis of Results. Based on the calculations of NBC that has been carried out, it can be concluded that 85% of students will graduate on time. The use of NBC will be better when more training data.

Keywords: *Data Mining, Naïve Bayes Classifier, Predictive Evaluation, students' performance*

1. INTRODUCTION

Computer science department is the youngest department in mathematics and natural sciences faculty of Universitas Negeri Semarang (UNNES), a state university in Semarang, Indonesia. However, the existence of computer science department has been widely known to the public/high school students. As a result, although it is a new department but the demand is already as much as the other departments in the faculty of mathematics and natural sciences of UNNES such as the mathematics department and biology department.

Therefore, predictive evaluation is indispensable in order to UNNES especially mathematics and natural sciences faculty can find a map of students' performance from the computer science department of mathematics and natural sciences faculty of UNNES. The scientific significance which expected is that through the predictive evaluation of students' performance of Computer Science Department of UNNES will be known the prediction scientifically which can be accounted about the success rate of students to graduate on time. This research is based on the use of data mining. According to Fariz, A, Abouchabaka, J, and Rafalia, N [2], data mining is the process of finding relationships in data that does not known by the user and present it in a way that can be understood so that the relationship can be the basis for decision making. There is a lot of data mining techniques that can be used. According to Feng [3], analysis in data mining is an important area of research which has a unique position in a number of data analysis and processing. In this research, as the tool of predictive evaluation of students' performance from computer science department of UNNES used Naïve Bayes Classifier (NBC) algorithm.

The purpose of this research are as follows. (1) Generate a system based on the algorithm of Naive Bayes Classifier (NBC) to predict the performance of students of Computer Science of UNNES based

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

on the performance of students from the 1st semester up to the 4th semester. (2) Obtain a description of the process of prediction of success to graduate on time for students of Computer Science of UNNES using Naïve Bayes Classifier algorithm. (3) Getting the result of prediction of success of students of Computer Science of UNNES to graduate on time based on the performance of the students of the 1st semester to 4th semester. The results of this research are expected to become the input for institution in making policy, both for the recruitment of students as well as the recruitment of lecturer of computer science department.

The focus of this research is stressed on the predictive evaluation of students' performance by applying the algorithm of Naïve Bayes Classifier (NBC). This predictive evaluation of students' performance from computer science department of mathematics and natural sciences faculty of UNNES is chosen based on the performance of students of the 1st semester up to the 4th semester, by the following considerations. (1) the subjects in the first semester still mostly filled with general subjects. (2) the subjects in the second semester up to fourth semester has dominated by the subjects of the study program so that the students interest in the study program selected, how students learn the subject material, and the absorption of the material by the students are expected to become the basis in predicting the success of lecture.

The results of this research are expected to have contribution significantly to the literature, especially related to: (1) the development of the latest literatures in the Department of Computer Science of UNNES that can be applied to predict the successful of the students to graduate on time. Department of Computer Science take advantage of advances in science and technology, especially the way to use Data Mining based on NBC, in an effort to develop a system of renewable Information & Communication Technology; (2) The results of this research become a scientific reference for the similar department to develop the Information & Communication Technology system based on the current literature.

The problem is how the prediction of lecture success from students of computer science of mathematics and natural sciences faculty of UNNES based on GPA of the fourth semester?

2. LITERATURE REVIEW

It was defined by [8] that data mining is a process to use statistical technique, mathematics, artificial intelligence and machine learning to extract and identify related beneficial information and knowledge from any big database.

Subsequently, [4] and [12] wrote that data mining is an activity to find interesting patterns from large amounts of data, the data can be stored in the database, data warehouse, or other information storage. Data mining related to the other field of sciences, such as database systems, data warehousing, statistics, machine learning, information retrieval, and high-end computing. In addition, data mining is supported by other disciplines such as neural networks, pattern recognition, spatial analysis of data, image database, signal processing. Data mining is defined as the process of discovering patterns in data. This process is automatic or frequent on semiautomatic. The pattern that was found must be full of meaning and that pattern provides benefit, usually benefits economically.

Characteristics of data mining are as follows.

Data mining related to the discovery of something hidden and specific data patterns previously unknown.

Data mining usually use a very large data. Often the data are used to make the results more trusted.

Data mining is useful to make critical decisions, especially in the strategy.

Based on some understanding above, it can be concluded that data mining is a technique to dig valuable information buried or hidden in a data collection (database) which is so big to be found an interesting pattern that was previously unknown. The word mining itself means effort to get a little valuable from a large number of base materials. Therefore, data mining actually has long roots of disciplines such as artificial intelligent, machine learning, statistics and databases. Some methods which often mentioned in the literature of data mining include clustering, classification, association rules mining, neural networks, genetic algorithms, and others.

Data mining has been applied in various areas because of its ability to rapidly analyze vast amount of data. [5] built the graduates employment model using classification task in data mining and compare some of data mining approaches such as Bayesian method and the tree method. Data mining technique was also implemented by [6] for meteorological data analysis and data mining technique to perform market analysis was implemented by [9].

Pattern recognition is a discipline that studies the ways to classify objects into several classes or

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

categories and recognize the tendency of the data. Data mining, often called knowledge discovery in databases (KDD), is an activity that includes the collection, use historical data to find regularities, patterns or relationships in large data sets. Output of data mining can be used to improve decision making in the future.

Machine Learning is an area in artificial intelligence which related to the development of techniques that can be programmed and learning from past data. Pattern recognition, data mining and machine learning is often used to refer to the same thing. This field takes the science of probability and statistics sometimes optimization. Machine learning becomes a tool in the analysis of data mining. How these fields are related can be seen in Figure 1.

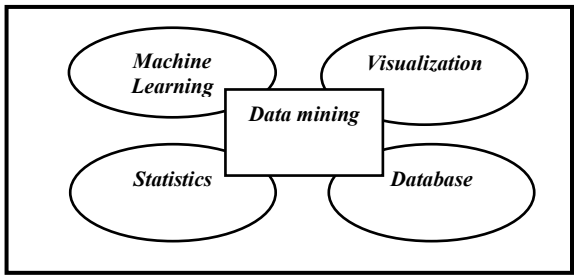

Figure 1. Relationship between data mining and other fields.

Naïve Bayes is a classification with probability method and statistics put forward by the British scientist Thomas Bayes, which predict future opportunities based on the experience in the past so it is known as Bayes' Theorem. [8] described the Naïve Bayes for each class of decisions, calculate the probabilities with the proviso that the class's decision was correct, given the object information vector. This algorithm assumes that the object attributes are independent. Probability is involved in producing the final estimate is calculated as the number of frequencies on the "master" of decision table. [11] also stated that Naïve Bayes Classifier is one of the simplest and the most effective method for classification. They provided the technique for the classification of diseased protein sequences using Naïve Bayes method. While [1] proposed an efficient decision support system for diabetes disease using Naïve Bayes Classifier.

The Naïve Bayes classifier works very well compared with other classifier models. This is evidenced by [13] in his journal "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages "saying that" Naïve Bayes classifier has a level of accuracy that is better than the other classifier models "as shown in Table 1.

Table 1. Result of measure

Classifier	Accuracy	Precision	Recall	F-Measure
NB Classifier	95.20%	99.37%	95.23%	97.26%
DT Classifier	94.85%	98.31%	95.90%	97.09%

According to [4] The process of The Naïve Bayesian classifier, or Simple Bayesian Classifier, as follows:

Variable D becomes a training set tuple and label which associated with the class. As usual, each tuple is represented by n-dimensional vector attribute, $X = (x_1, x_2, ..., x_n)$, this describes that the n measurements made on a tuple of attributes n, respectively, $A_1, A_2, ..., A_n$.

Suppose there are m class, $C_1, C_2, ..., C_m$. Given a tuple, X, classifier will predict X which entry a group has the highest posterior probability, the conditions mentioned in X. That is, Naïve Bayesian classifier predicts that X tuple belongs to the class C_i, if and only if:

$$P(C_i|X) > P(C_j|X) \text{ for } 1 \leq j \leq m, j \neq i. \quad (2.1)$$

So maximize $P(C_i \mid X)$. C_i class which $P(C_i \mid X)$ is maximized called maximum a posteriori hypothesis. By Bayes theorem:

$$P(C_i|X) = \frac{P(X|Ci)P(Ci)}{P(X)} \quad (2.2)$$

Note :

$P(C_i|X)$ = probability of the hypothesis if given facts or record X (Posterior probability)

$P(X|C_i)$ = looking for value parameter that gives the greatest possibility (likelihood)

$P(C_i)$ = Prior probability of X (Prior probability)

$P(X)$ = Total probability tuple which appears

When $P(X)$ is constant for all classes, only $P(X \mid C_i) P(C_i)$ need to be maximized. If the probability of a previously class is unknown, it is generally assumed to be in the same class, namely $P(C_1) = P(C_2) = \bullet \bullet \bullet = P(C_m)$, so they will maximize $P(X \mid Ci)$. If not, it will maximize $P(X \mid C_i) P(C_i)$. Note that the prior probability classes can be estimated

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

by $P(C_i) = |C_i, D| / |D|$, where $|C_i, D|$ is the amount of tuple on training class C_i in D.

The dataset has many attributes, it will be very difficult in compute to calculate $P(X|C_i)$. In order to reduce the calculations in evaluating $P(X|C_i)$, Naïve independence conditional class assumptions made . Considered that the values of the attributes are conditionally independent of each other, given the class label from the tuple (namely that there is no relationship of dependence between attributes) thus:

$$P(X|C_i) = \prod_{k=1}^{n} X_k \mid C_i) = P(X1|C_i) \times P(X2|C_i) \times \ldots \times P(Xn|C_i) \quad (2.3)$$

So we can easily estimate the probability $P(x_1 | C_i)$, $P(x_2 | C_i)$,. . . , $P(x_n | C_i)$ of tuple training . Remember that here x_k refers to the value Ak attribute for the tuple X. For each attribute, it is judged whether the value of attribute is categorical or continuous. For example, to calculate $P(X | C_i)$ to consider the following matters:

If A_k is categorical, then $P(X_k | C_i)$ is the amount of tuple classes C_i in D has a value X_k to attribute A_k, divided by $|C_i, D|$, the amount of tuples classes C_i in D.

If Ak continuous-valued, it is necessary to do a little more work, but the calculations are quite simple. A continuous-valued attributes are usually assumed to have a Gaussian distribution with average μ and standard deviation σ, is defined by

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (2.4)$$

therefore:

$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma C_i) \quad (2.5)$$

After that, count μC_i and σC_i, which is a deviation of the mean (average) and standard each attribute value A_k for classroom training tuple C_i. Then use the second quantity in the equation, together with x_k, to estimate $P(x_k | C_i)$.

To predict the class label x, $P(X|C_i)P(C_i)$ is evaluated for each class C_i. Classifier predict the class label of tuple x is the class C_i, if,

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ for } 1 \leq j \leq m, j \neq i, \quad (2.6)$$

In other words, the class label predicted is C_i where $P(X | C_i) P(C_i)$ is the maximum Bayesian classifiers have a minimal error rate compared with other classification. However, in practice this is not always the case, because of inaccuracies in the assumptions made for its use, such as an independent class conditions, and lack of data available probability.

Bayesian classifier is also useful in providing a theoretical justification for other classifiers that do not explicitly use Bayes' theorem.

3. RESEARCH METHOD

The validity of each stage of this research was analyzed and evaluated. Thus, each of the next evaluation stage will be done if and only if the previous stage has been completely valid. For example, before the researchers used the testing data in step 3 stage II, the first evaluation performed by the training data until it has been known that the programs made and the results are valid. Similarly, before the training data was used, the rightness of the use of NBC algorithms has been evaluated first and the result was valid.

Stages of the research illustrated in Figure 2 below with the following explanation:

a. Stage I, Literature Review

This stage is the stage through the literature relevant to the research. This phase is carried out to obtain the relevant information related to the research. A literature study by gathering information from books and journals that are relevant to this research, study the documentation relating to the evaluation of the performance of students from the Department of Computer Science 1th Semester up to 4th semester, data mining, and Data Mining based on Naïve Bayes Classifier (NBC) Algorithm. Furthermore, an analysis of the data mining process at the Department of Computer Science is mainly concerned with the performance of students. Data Mining Algorithms based on Naïve Bayes classifier (NBC) is used to evaluate the predictive of performance from students of Computer Science Department of UNNES.

b. Stage II, Determining Data Mining Pattern.

This stage is explained as follows:

1) The Data Collection Stage

The data collection stage is the initial stage after the preparation phase is completed. The data collected is the data related to the process of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

conditions and the data relating to the predictive evaluation of student performance. The data collected are training data and testing data, which is the GPA of the 1st semester up to the 4th semester, associated with the process of data that will be used for checking the system created and for the predictive evaluation of students' performance in real. Training data and testing data are valid data that existing or already established to be used as the material of Bussiness Understanding of evaluation of students' performance before doing the analysis based on the research objectives.

2) Stage of Data Mining Systems Building

Conditions and the data obtained from the data collection phase are used as the foundation for building a data mining system. In this step, the researchers conducted process of information system development includes the development of a data base, application development, and testing of applications, as well as the implementation of applications.

3) Stage of Data Processing

Once the system is building, the data obtained in the data collection were processed on a system that has been built.

4) Conducting the process of predictive evaluation of students' performance.

This process is performed on the data which already entered to the system. The system was given NBC's algorithm in order to make the process of Data Mining.

5) Analysis of the results

This process is done to determine the pattern of predictive evaluation of student performance.

6) Documentation

At each step, the researchers did research documentation process in the form of photos, or documents related.

3.1 Technique of Data Collection

The data collected in this research are training data and testing data, which is the GPA of the 1st semester to the 4th semester, in the form of already established facts about the settlement process based on Data Mining Algorithm of Naïve Bayes Classifier (NBC) as a tool of predictive evaluation of the performance of students from the Department of Computer Science of UNNES. The collection of data in this research used a thorough documentation and observation when the research activities carried out. The process of collecting this data using manual documentation of data at each stage of the research conducted. How that is done is recorded through the academic system of UNNES.

3.2 Data Analysis

Data analysis was performed as the efforts made by working with the training data and testing data, organize data, sorted them into units that can be managed, synthesize, search and find what is important and what is learned and decide. Data analysis is the process of organizing and sorting data into patterns, categories, and unit basic description so it can be found the theme and can be formulated working hypotheses according tothe problems.

Based on the above data analysis definition, then the data from observations at each research step narrated, identified related to data mining using Naïve Bayes classifier algorithm. While Trial Data Mining documented and made categorization.

The data in this research were analyzed by using an analysis technical of the data model of [7], which includes: (1) data reduction; (2) data display; and (3) conclusion/verification.

4. RESEARCH RESULT AND DISCUSSION

4.1 Research Result

The results achieved are presented based on the issues and the order of research purposes so that the presentation is more conclusive. Researchers have carried out the steps of research according to the procedure that have been planned in order to achieve the research objectives. The conclusive presentation is as follows.

1) Stage of Data Collection

The stage of data collection is the initial stage after the preparation phase is completed. The data collected is the data relating to the process conditions and the data relating to the predictive evaluation of student performance. These data are the material of business understanding of students' performance evaluation.

In this research the captured data is divided into two parts, namely the training data and testing data. The training data retrieved from the data of students who have graduated and have known whether the student graduates on time or not. This data are students' GPA of 1st semester GPA up to fourth semester. After the training data is rechecked and valid then it stored in the database system of the application. The training data is shown in figure 3.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The testing data is retrieved from the students of Informatics Engineering of Computer Science Department of Semarang state university. The testing data is taken from the student of 2013 and 2014. Prediction of graduation of Informatics Engineering students is based on the existing training data. After the testing data is rechecked and valid then it stored in the database system of the application. The testing data is shown in figure 4.

2) The stage of Data Processing using Naïve Bayes Classifier.

The stages of data processing are performed as follows.

The training data and testing data which processed using the numerical data so that it is needed the categorization of GPA. This categorization is based on the weights value as in the book of Academic Guidelines of Semarang State University which shown in Table 2.

Table 2. Categorization of GPA

Category	Range of GPA
A	3,51- 4,00
AB	3,01-3,50
B	2,51-3,00
BC	2,01-2,50
C	1,51-2,00
CD	1,01-1,50
D	0,51-1,00
E	0,01- 0,50

The next process is calculation of probability value per category of training data which is on time and not on time.

It is then continued by the calculation of probabilities of each semester in each training data. This is done by calculating the probability of occurrence of a grade in that semester. The probability data which produced are stored in the database systems of Department of Computer Science.

Furthermore, the calculation of the possibility of precise and imprecise possible value by multiplying the probability value of each term is multiplied by the probability values of on time or not on time. The results of this possibility are shown as follows.

This calculation shows that from the 80 students of Information Engineering there were twelve students included in the category of not on time, and 68 students included in the category of on time. This result shows that 85% of students of Computer Science are expected to graduate on time. In the system, it will be shown the student data that is on time and not on time as shown in figure 5.

The amount of training data in this research is very influential on the results of the classification of the NBC's algorithm as one form of classification algorithms [4]. Therefore, in this research used the training data from the students of Mathematics study program.

Based on the above description, then the results of this research: (1) It has been obtained the data to predict the success of students of Computer Science Department of Mathematics and Natural Sciences Faculty of UNNES based on GPA of semester 4. (2) It has been built the system of data mining exploration which based on the algorithm of Naïve Bayes Classifier (NBC) as predictive evaluation of performance of the students.

Based on the description above then the purpose of this research have been achieved, namely (1) It has been produced a system based on the algorithm of Naive Bayes classifier (NBC) which has been rechecked and valid to predict the performance of students of Computer Science of UNNES based on the performance of students from the 1st semester up to the 4th semester. (2) It has been obtained a description of the process of prediction of success to graduate on time for students of Computer Science of UNNES using Naïve Bayes Classifier algorithm. (3) It has been gotten the result of prediction of success of students of Computer Science of UNNES to graduate on time based on the performance of the students of the 1st semester to 4th semester.

4.2 Discussion

The data used in this research is divided into two categories namely training data and testing data which is the GPA of the 1st semester to the 4th semester. Training data are obtained by researchers from the documentation study on the database system at the State University of Semarang. This training data can be obtained in the integrated academic system in the form of already established facts. The training data obtained were subsequently analyzed using the algorithm of Naïve Bayes Classifier (NBC) to test the system in order to predict the students' performance of Computer Science of UNNES.

Initially, the system generated through this research was tested using the training data, that was the GPA of 1st semester up to 4th semester of students who have graduated. This stage resulted

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the predictions of success of study based on training data. Based on testing using the training data, the result was valid, the system is capable to predict in accordance with the facts of students' success of study.

Furthermore, after the system successfully tested using the training data, then the system is applied using existing testing data, that is data on students' GPA of 1st semesters up to 4th semester of the Information Engineering study program of computer science of UNNES. Based on the final results, it was obtained the information that the percentage of students of Information Engineering study program of Computer Science Department that will graduate on time is 85%. This is in accordance with Han and Kamber [4], which discusses that the Naïve Bayes Classifier can be applied effectively in order to perform a classification associated with predictions graduate on time or not on time.

However, the results of this system have not entered the resistor element of the failure of students to graduate on time, for example, disturbances due to illness, economic difficulties, or due to other inhibiting factors.

Thus, the purposes of this research have been achieved, namely (1) The research team have been produced a system based on the algorithm of Naive Bayes classifier (NBC) so that be able to predict the state of graduation (on time or not on time) of students of Computer Science of UNNES based on the performance of students from the 1st semester up to the 4th semester. (2) A description of the process of prediction of success to graduate on time for students of Computer Science of UNNES using Naïve Bayes Classifier algorithm has been obtained (3) Based on the performance of the students of the 1st semester to 4th semester, the research team have obtained the result of prediction of success of students of Computer Science of UNNES to graduate on time.

Through the results of this study, it is expected that students of Computer Science of UNNES has a wider knowledge towards the use of an algorithm such as Naïve Bayes Classifier algorithm. The process application of an algorithm like this can be used as an initial reference for the development of science or literature.

Limitations in the process and the results of this research, (1) the results of the system are made based on the modification of the theory associated with Naïve Bayes Classifier algorithm is still a classical predicted results or a general nature. This system can be developed to predict the graduation of a student through the input individually; (2) the end result of this system has not led to the suggestion to students based on the results shown or written on the end result of this system, especially if the student has the information not graduate on time. Thus, this system still need to be developed further by giving a solution in accordance with the outcomes produced by these systems; (3) the results of the system was not incorporated elements of inhibiting the failure of students to graduate on time, for example, disturbances due to illness, economic difficulties, or due to other inhibiting factors.

5. CONCLUSIONS AND RECOMMENDATIONS

Based on the calculations of Naïve Bayes Classifier that has been carried out, the process by GPA of first semester up to fourth semester students, it can be concluded that 85% of students will graduate on time from the eighty students. For suggestions, the use of Naïve Bayes Classifier will be better when more training data. It is required the training data as precise as possible and the more the better.

REFERENCES:

[1] Ambica, A, Gandi, S, and Khotalanka, A. "An Efficient Expert System for Diabetes by Naïve Bayesian Classifier". Internationa Journal of Engineering Trends and Technology (IJETT). Vol 4 Issue 10 - Oct 2013.

[2] Fariz, A, Abouchabaka, J, and Rafalia, N. 2015. Using Multi-agents Systems in Distributed Data Mining: A Survey. *Journal of Theoretical and Applied Information Technology*. Vol. 73 No. 3.

[3] Feng, Yuhua. 2012. Analysis on Algorithm and Application of Cluster in Data Mining. *Journal of Theoretical and Applied Information Technology* . Vol. 46 No. 1.

[4] Han, J. and Kamber ,M. 2006. "Data mining: Concepts and Techniques, 2nd edition". The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor. 2006.

[5] Jantawan, B and Tsai, Cheng-Fa. "The Application of Data Mining toBuild Classification Model for Predicting Graduate Employment". International Journal of Computer Science and Informatio Security. Vol. 11, No. 10, October 2013.

[6] Kohail, SN and El-Halees, AM. "Implementation of data mining techniques for meteorological data analysis (case study for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Gaza street)". International Journal of Information and Communication Technology Research. Volume 1 No. 3, July 2011.

[7] Miles, M. B., & Huberman, A. M.. "Qualitative Data Analysis an Expanded Sourcebook". California: Sage Publishing, Inc. 1994.

[8] Orson & Belen. "Advanced Data Mining Techniques". Springer: USA. 2008.

[9] Sabitha, B, Amma, NGB, Annapoorani, G, and Balasubramanian, P. "Implementation of Data Mining Techniques to Perform Market Analysis". International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 11, November 2014.

[10] Turban, E, Aronson, JE, and Liang, TP. "Decision Support System and Intelligent Systems – 7 th ed". Pearson Education, Inc. Pearson Education, Inc. 2005.

[11] Vijayarani, S and Deepa, S. 2014. "Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences". International Journal of Computational Intelligence and Informatics. Vol. 3: No. 4, January- March 2014.

[12] Witten, I.H and Frank E. "Data Mining: Practical Machine Learning Tools and Technique, Second Edition". San Francisco: Morgan Kauffman. 2005.

[13] Xhemali, Daniela dkk. 2009. "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages". International Journal of Computer Science Issues. Vol. 4 No. 1, 2009.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

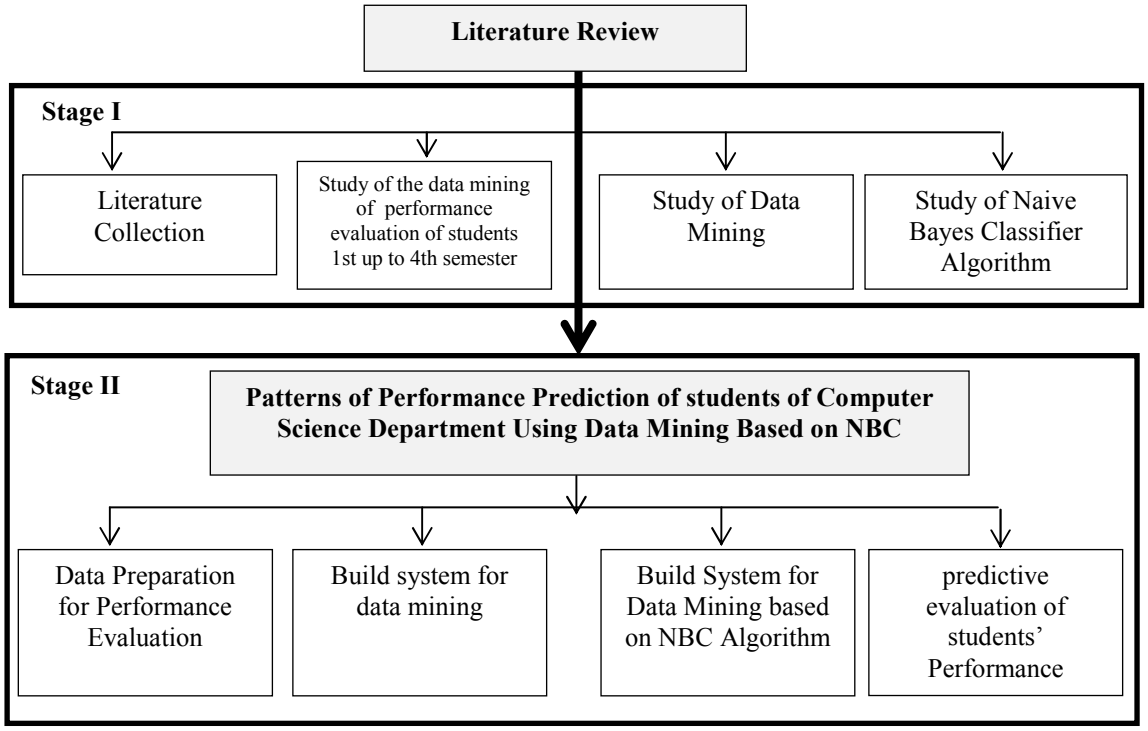

Figure 2. Stages of Research

Figure 2. Stages of Research

Figure 3. Training Data

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4. Testing Data

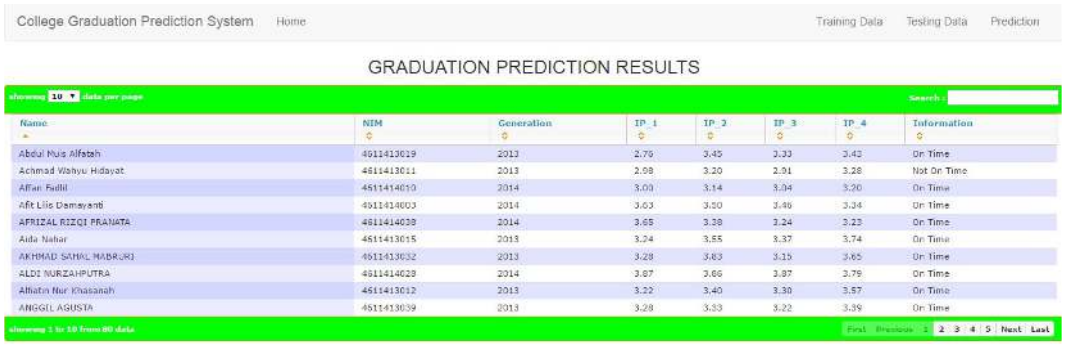

Figure 5. Graduation Prediction Result