A COMPARATIVE STUDY OF PREDICTING TEACHING SCORE BY USING CLASSIFICATION ALGORITHMS

Myint Myint Than[1]

[1] *Lecturer, University of Computer Studies, Kalay, Burma*

ABSTRACT

Today is Educational Age. Education is an essential element for the progress of country. *Educational institutions/ colleges have goals to deliver quality education so, one way to improve the level of education quality in higher education is that measure good performance of teachers' with higher teaching skill. Teacher training, learning time of related area and teaching experience are very important for qualified teacher. Data mining is used in higher education for analysis, predicting and evaluating the performance of student, teacher and other that have a role in the educational system. Mining in educational environment is called Educational Data Mining (EDM). Educational data mining is concerned with developing new methods to discover knowledge from educational database. There are varieties of popular data mining task within the educational data mining e.g. classification, clustering, outlier detection, association rule, prediction etc. The proposed system can able to predict teaching skill of teachers' using classification model under the data set of a colleagues and students evaluation and also identify factors that affect the qualification of teachers. This paper emphasizes decision tree (J48), Random Forest and Naive Bayes classification algorithms for classification model using Weka tool. Weka is data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data or called from the Java code. Weka is a collection of tools for regression, clustering, association, data pre-processing, classification, and visualization. In this paper, data set is collected information and results of a survey about 65 teachers and questionaries 200 students on teacher's behaviors in classroom, university of Computer Studies, Kalay (Burma).*

Keyword: - *Data Mining, Educational Data Mining (EDM), Decision Tree, Classification, Weka*

---

1. INTRODUCTION

Most researchers have worked on the evaluations and prediction of teacher performance in higher education using data mining techniques. Data mining helps to evaluate and predict the performance of teachers in higher education. They proposed a model to evaluate and predict the performance of teachers using different machine learning algorithm. Data mining techniques can be functional on the different data repositories including databases, data warehouses, and multimedia data. Advances in computer hardware and data mining software have made data mining systems accessible. It can be applied in any area, including education, banking and insurance, medicine, security and communication and affordable to many businesses. This paper provide basic knowledge of teaching and learning process for operative education. The main goal of any educational institutions/colleges are to deliver quality education to theirs students. One way to achieve the highest level of quality for higher institution/college education system is to pay teacher training for the related subject. Teacher training is essential to raise the performance of teachers' skill. The problems in higher education are the evaluation of teachers' performances in the related subject and other areas. The teaching experience is also important for the performance of teachers' skill. The experienced teacher knows how to teach their pupils to understand the lectures more easily and which teaching methods are more effective for them. The most widely applied tool to evaluate the performance of teachers' in a course is through surveying students' responses about the course and the teacher questionnaires. And also, the teacher may have the best performance on research to find out the solution for community problems, have good performance in community service, social life inside and outside the university. In that study 200 students' data were used and the relationship between the university exam results and achievements was examined .This paper analyses how to improve some aspects of educational quality with data mining algorithms and techniques by taking a specific teaching skill as target audience in academic study. This paper is organized information and results of a survey about

65 teachers in University of Computer Studies, kalay. It is proceeded to analyze and predict acceptation of a teacher for continuing the teaching in faculty. There are new rules and relations between selected parameters such as Teacher's-degree, Training, Teaching-Experience, Study-Time, Family-member, Evaluate-Score. This paper applies classification algorithm such as J48 decision tree induction, Random Forest and Naïve Bayesian classification algorithms by the use of WEKA 3.9.3 data mining tool.

2. RELATED WORK

In this paper is collected information and results of a survey about 65 teachers and questionaries of 200 students on teacher's behaviors in classroom, then focused on J48 decision tree algorithms for data analysis with an experimental approach and proved that J48 has more accurate result compared with Naïve Bayesian Classifier, and Random Forest classifiers. [1]. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. Now a day, large quantities of data is being accumulated. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Data could be large in two senses: in terms of size & in terms of dimensionality. Data Mining could help in a more in-depth knowledge about the data [2]. One of the first studies on data mining applied in education was published in 1995 by Sanjeev and Zytkow. Researchers gathered the knowledge discovery as terms like "P pattern for data in the range R" from university database [3]. Another study on data mining applied in education was published in 2000 by Becker and C. Ghedini ve E.L. Terra who are performed for defining and understanding the impact of changes in curriculum on students at a university in Brasil[4]. Many researchers have been already done in the field of educational data mining and performances of the teachers using data mining techniques For improving the performance of teachers as well as higher education.Many researchers have been given their review [5], [6],[7],[8],[9]. Maltepe University students identifying characteristics had been clustered using K-means algorithm in 2005 by Erdoğan and Timor. In that study 722 students' data was used and the relationship between the university entrance exam results and achievements was examined [10]. Vranić and Skoćır was examined how to improve some aspects of educational quality with data mining algorithms and techniques by taking a specific course students as target audience in academic environments [11].

3. METHODOLOGY

Here we used WEKA 3.9.3 tool for analyzing our dataset. WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. In a real-world datamining problems, WEKA is a state-of-the-art facility for developing machine learning (ML) techniques and their applications. It is a collection of different machine learning algorithms for data mining tasks and predictions. The algorithms are applied directly to a dataset. WEKA implements algorithms for data association, preprocessing, classification, clustering, regressing rules; it has another advantages as visualization tools. This package of WEKA used for developing new machine learning schemes. Under the GNU general Public License, WEKA is a open source software.

3.1 Data Preprocessing

The data collected from the real world is incomplete, inconsistent, inadequate and it consisting of noise, redundant groups. The Knowledge discovery using the training data with such an irrelevant, inconsistent and redundant data will reduce the mining quality. Data preprocessing is very important and a prerequisite step in the data mining process. Low quality data will lead to low quality mining results. Thus, the data can be preprocessed in order to improve the quality of the data. Data quality can be accessed in terms of accuracy, completeness and consistency. Preprocessing reconstructs the data into a format that will be very easy and effective for further processing.

To improve the quality of mining, the data preprocessing techniques are applied. The main objective of data preprocessing is cleaning of noise, filling up of missing values, reduce the redundancy and normalize the data. The preprocessing steps are data cleaning, data integration, data transformation, data reduction and data discretization. After preprocessing has been done the data will be complete, noise free and ready for classification. Any classification algorithm can be applied for classifying the data. There are various tools and techniques that are used for preprocessing which includes:

Data cleaning: It can be applied to remove noise and correct inconsistencies in data.

Data integration: It merges data from multiple sources into a coherent data store such as a data warehouse.

Data reduction: It can reduce data size by, for instance, aggregating, eliminating redundant features, or clustering.

Data transformation: It may be applied, where data are scaled to fall within a smaller range like 0.0 to 1.0.

Data discretization: It can also be useful, where row data values for attributes are replaced by ranges or higher conceptual levels. For example, raw values for age may be replaced by higher-level concepts, such as youth, adult, or senior.

3.2. Classification Analysis

Classification is a form of data analysis that extracts models describing important data classes. Such models, called classifiers, predict categorical (discrete, unordered) class labels.

Many classification methods have been proposed by researchers in machine learning, pattern recognition, and statistics. Most algorithms are memory resident, typically assuming a small data size. Recent data mining research has built on such work, developing scalable classification and prediction techniques capable of handling large amounts of disk-resident data. Classification has numerous applications, including fraud detection, target marketing, performance prediction, manufacturing, and medical diagnosis.

Classification analysis is used to map data sets into predefined groups and classes. As the classes are determined before examining the datasets, it is considered to be the supervised learning. The learning of classifier is " supervised" in that it is told to which class each training tuple belongs. It constracts with unsupervised learning (or) clustering, in which the class label of each training tuple is not known, and the number or set of classes to be learned may not be known in advance.

3.2.1 Bayes classification Method

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities such as the probabilities that a given tuple belongs to a particular class. Bayesian classification is based on Bayes' theorem. Naïve bayes are group of probabilistic classifiers built on the Bayes's theorem. The naïve Bayesian classifier is simple Bayesian classifier.

Which states that: Consider X and H
X : is an evident, $X = x_1, x_2, \ldots\ldots x_n$
H: is the hypothesis. P(H\X) is the posteriori probability, of H conditioned on X.
Then $P(H/X) = (P(X/H) \times P(H)) / P(X)$

Naïve Bayesian Classification

The naïve Bayesian Classifier, or simple classifier, works as follows:

1. Let D be a training set of tuples and their associated class labels. As usual, each tuple is represent by an n-dimensional attribute vector, $X = (x_1, x_2, \text{---}, x_n)$, depicting n measurements made on the tuple from n attributes, respectively, $A_1, A_2, \text{---}, A_n$.
2. Suppose that there are m classes , $C_1, C_2, \text{---} C_m$. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is , the naïve Bayesian classifier predicts that tuple X belongs to the class Ci if and only if
   $P(Ci\backslash X) > P(Cj\backslash X)$ for $1 <= j <= m, j != i$.
   Thus, we maximize P(Ci\X). The class Ci for which P(Ci\X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem
   $P(Ci\backslash X) = (P(X\backslash Ci) \times P(Ci)) / P(X)$
3. As P (X) is constant for all classes, only
   $P(X\backslash Ci) \times P(Ci)$ needs to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is, $P(C_1)=P(C_2)= \text{----} = P(C_m)$, and we would therefore maximize P(X\Ci). Otherwise, we maximize $(P(X\backslash Ci) \times P(Ci))$. Note that the class prior probabilities may be estimated by
   $P(Ci)= |C_{i,D}| / |D|$, where $|C_{i,D}|$ is the number of training tuples of class Ci in D.

4. Given data sets with many attributes , it would be extremely computationally expensive to compute P(X\Ci) . To reduce computation in evaluating P(X\Ci), the naïve assumption of class-conditional independence is made. This presums that the attributes' values are conditionally independent of one another, given the class label of the tuple (i.e, that there are no dependence relationships among the attributes). Thus,

$$P(X\backslash C_i) = \prod_{k=1}^{n} P(X_k \backslash C_i)$$

$$= P(X_1\backslash C_i) \times P(X_1\backslash C_i) \times \text{----} \times P(X_n\backslash C_i)$$

We can easily estimate the probabilities $P(X_1\backslash C_i) \times P(X_2\backslash C_i) \times \text{----} \times P(X_n\backslash C_i)$ from the training tuples. Recall that here X_k refers to the value of attribute A_k for tuple X. For each attribute, we look at whether the attribute is categorical or continuous valued. For instance, to compute P (X\Ci) , we considered the following:

(a) If A_k is categorical, then P (Xk\ Ci) is the number of tuples of class Ci in D having the value X_k for A_k, divided by |C i,D|, the number of tuples of class Ci in D.
(b) If A_k is continuous-valued, then we need to do a bit more work, but the calculation is pretty straightforward. A continuous-valued attribute is typically assumed to have a Gussian distribution with a mean μ and standard deviation σ, defined by

$$g(x, \mu, \sigma) = 1 / (2 \pi \sigma)^{½} \times e^{-(x-\mu)2/2\sigma 2}$$

So that $P(Xk \backslash Ci) = g(x_k, \mu_{Ci}, \sigma_{Ci})$

These equations may appear daunting, but hold on! We need to compute μ_{Ci} and σ_{Ci}, which are the mean (i.e., average) and standard deviation, respectively, of the values of attribute A_k for training tuples of class Ci.

5. To predict the class label of X, P (X\Ci) × P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of tuple X is the class Ci if and only if
P(X \Ci) × P(Ci) > P(X\Cj) × P (Cj) for 1<= j<= m, j!=1

In other words, the predicted class label is the class Ci for which P (X\Ci) × P(Ci) is the maximum.

3.2.2 Decision Tree Induction

The decision tree is one of the classification algorithms. It is frequently used by the researchers to classify the data. The decision tree is very popular because it is easy to build and require less domain knowledge. Also the decision tree method is scalable for large database.

The first decision tree algorithm is developed in early 1980s is Iterative Dichotomiser (ID3). Quinlan and Kaufmann (1993) presented the C4.5 which is the successor of ID3. In 1984 Classification And Regression Tree (CART) is introduced. It is mainly support for the binary tree classification. All the three algorithms adopt the greedy approach and construct the decision tree in top down, recursive, divide and conquer manner.

The following are the basic steps used for decision tree algorithm:

Input: Data partition, D, which is a set of training tuples and their associated class labels;

Attribute list: The set of candidate attributes;

Attribute selection method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split point or splitting subset.

Output: A decision tree:

(1) Create a node N

(2) If tuples in ‘D’ are belongs to same class C, then return N as a leaf node labeled with the class C

(3) If attribute list is empty then return N as a leaf node labeled with the majority class in D;

(4) Apply Attribute selection method (D, attribute list) to find the “best” splitting criterion

(5) label node N with splitting criterion

(6) If splitting attribute is discrete-valued and multiday splits allowed then//not restricted to binary trees

(7) Attribute list ← attribute list-splitting attribute; //remove splitting attribute

(8) For each outcome j of splitting criterion //partition the tuples and grow subtrees for each partition

(9) Let Dj be the set of data tuples in D satisfying outcome j; //a partition

(10) If Dj is empty then attach a leaf labeled with the majority class in D to node N

(11) Else attach the node returned by Generate decision tree (Dj, attribute list) to node N;

End for

(12) Return N

The decision tree algorithm is very robust and learning efficiency with its learning time complexity of O(nlog2n). The outcome of a decision tree that can be easily represented as a set of symbolic rules (IF….THEN). This rule can be directly interpreted and compared with available knowledge and provide useful information. A decision tree comprises of node and leaves, where nodes represent a test on the values of an attribute and leaves represent the class of an instance that satisfies the conditions. The outcome is ‘true’ or ‘false’. Rules can be derived from the path starting from the root node to the leaf and utilizing the nodes along the way as preconditions for the rule, to predict the class at the leaf. The tree Pruning has to be carried out to remove unnecessary preconditions and duplications. The nodes represent genes and branches represent the expression conditions. The leaves of the tree represent the decision outcome. The brace under a leaf denotes the number of instances correctly and incorrectly classified by the leaf and the equivalent decision rules are derived from the decision trees.

3.2.3 Random Forest classifier

Random forests is a trademark term for an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Random forests are collections of trees, all slightly different. It randomize the algorithm, not the training data. How you randomize depends on the algorithm, for c4.5: don’t pick the best, pick randomly from the k best options. It generally improves decision trees decisions. Unlike single decision trees which are likely to suffer from high variance or high Bias Random Forests use averaging to find a natural balance between the two extremes. A random forest is a meta estimator that fits a number of classifical decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. Each decision tree is constructed by using a Random subset of the training data.

4. ARCHITECTURE OF PROPOSED SYSTEM

In this paper, it is done survey from students and teachers, University of Computer Studies, Kalay (Burma).

4.1 Data set

In this study, 65 records were used which is taken from University of Computer Studies, Kalay (Burma). The attribute of datasets have Teacher’s-degree, Training, Teaching-Experience, Study-Time, Family-member, Evaluation-Score. Training data are analyzed by a classification algorithm. In this article the class label attribute is Evaluation score. It has three attribute values, weak, good and excellent. This paper predicts the value of evaluate score by using the naïve Bayesian classification algorithm, J48 and Random Forest decision tree classification algorithms. The Weka software version 3.9.3 is used to predict the evaluation score of teacher.

Table -1: The list of attribute variables used in this Study

Variable name	**Data type**	**Description**
Evaluation score	Text	Evaluation score
Teacher's degree	Text	Teacher's degree
Training	Text	Teacher subject and teaching training
Teaching experience	Text	Teaching experience of teacher
Family-Member	Text	Number of family member
Study-Time	Text	Use time for Teacher study

Table – 2: The list of attribute values used in this study

Variable name	**Data type**	**Value**
Evaluation score	Nominal	{Weak ,Good, Excellent}
Teacher's degree	Nominal	{BA, MA, PHD}
Training	Nominal	{yes, no}
Teaching experience	Nominal	{ low, medium, high }
Family-Member	Nominal	{small, medium, large}
Study-Time	Nominal	{ low, medium, large}

Table -3: The output variable (Evaluation score) used in the study

Score < 60	*Weak*
60 <= Score< 75	*Good*
75 <= Score < 100	*Excellent*

Table -4: The output variable (Teaching experience) used in the study

Years < 3	low
Years >= 3 and <= 10	medium
Years > 10	high

Table- 5: The output variable (Study-Time) for one week used in the study

Hours < 2	low
hours >= 2 and <= 4	medium
hours > 4	high

Table- 6: The output variable (Family member) used in the study

number <3	low
number >= 3 and <= 5	medium
number > 5	high

4.2 Loading Data set

Open the Weka software Version 3.9.3. Then Select Explorer > Open file. Then choose dataset file. The Explorer window should now look like this:

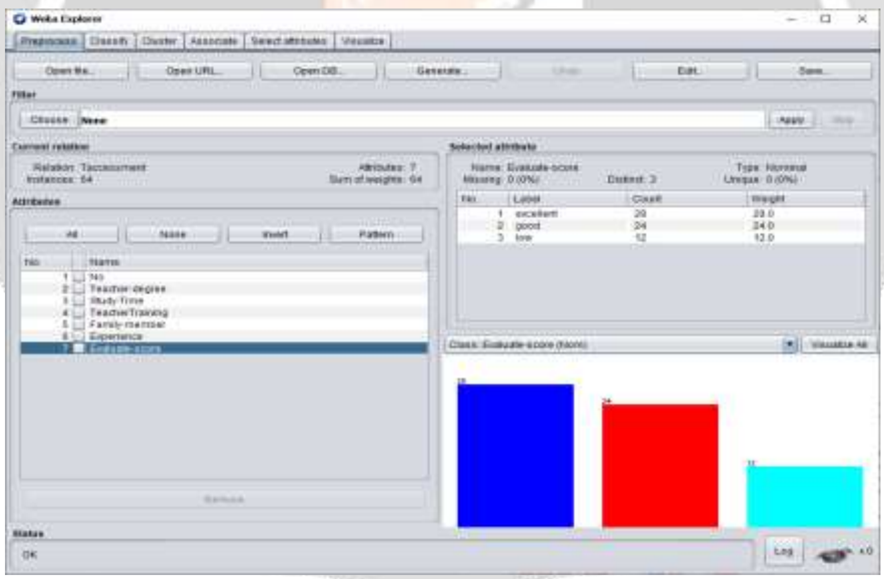

Fig - 1 : Opening data set in WEKA.

5 . CHOOSING THE EXPERIMENTAL ALGORITHMS

5.1 Experiment using Naïve Bayes classifier

Select Classify > Choose >Classifiers > Bayes > NaiveBayes. Then we set the Percentage Split of data, 70% of training and 30% for testing. Figure 3 shows the obtained results from applying Naïve classifier on the selected dataset which which shows 73.6842 % accuracy, 14 correctly classified instances, 5 incorrectly classified instances with Recall value of 0.737 and Precision of 0.739. Time taken to test model on test split is 0.01 seconds.

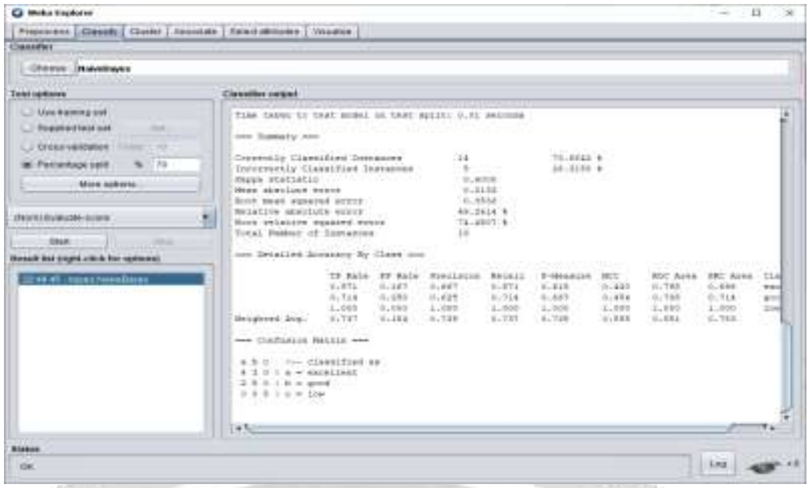

Fig - 2 : Naïve Bayes Classification Results

5.2 Experiment using J48 Decision Tree classifier

Select Classify > Choose > Classifiers > Trees > J48. Then we set the Percentage Split of data, 70% of training and 30% for testing. Figure 5 shows the obtained results from applying Decision Tree classifier on the selected dataset which which shows 89.4737 % accuracy, 17 correctly classified instances, 2 incorrectly classified instances with Recall value of 0.895 and Precision of 0.895. Time taken to test model on test split is 0.02 seconds.

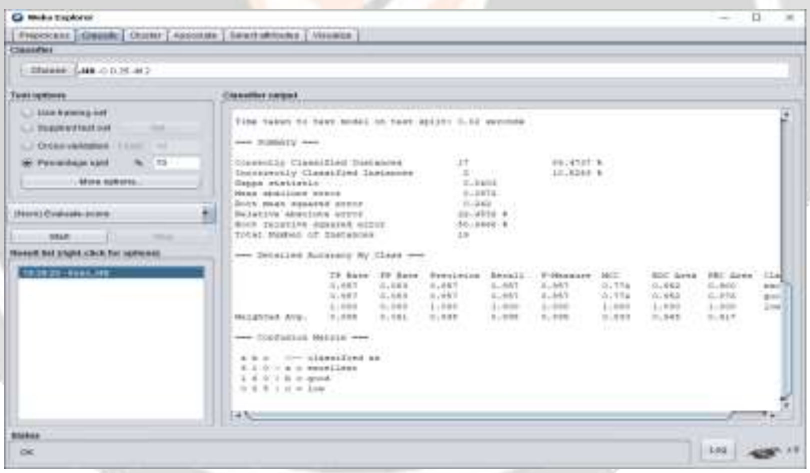

Fig- 3 : J48 Decision Tree Classification Results

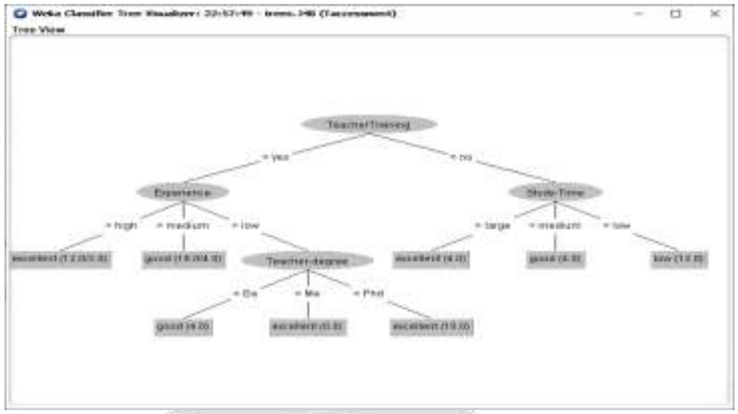

Fig -4 : Visualize J48 Decision Tree Classification Results

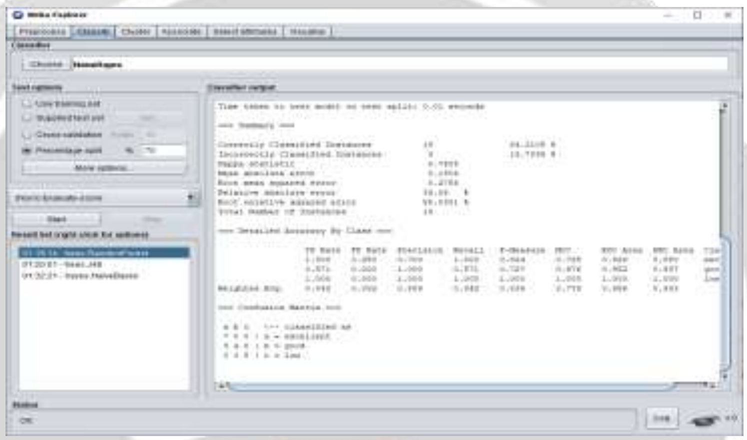

Fig – 5 : Random Forest Classification Results

5.3 Experiment using Random Forest classifier

Select Classify > Choose >Classifiers > Trees > Random Forest. Then we set the Percentage Split of data, 70% of training and 30% for testing. Figure 6 shows the obtained results from applying Random Forest Decision Tree classifier on the selected dataset which which shows 84.2105 % accuracy, 16 correctly classified instances, 3 incorrectly classified instances with Recall value of 0.842 and Precision of 0.889. Time taken to test model on test split is 0.01 seconds.

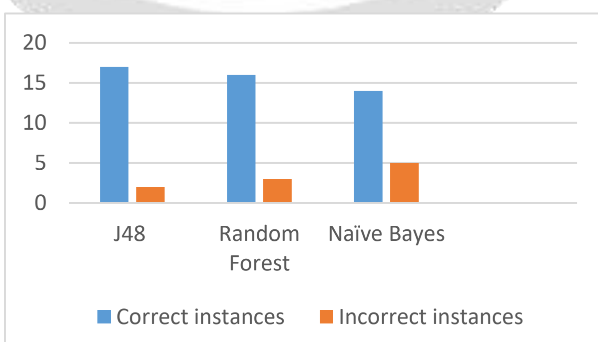

Chart – 1: Comparison of Classifiers

6. CONCLUSIONS

In this paper, a comparison for Naïve Bayesian, J48 and Random Forest decision tree classifiers are shown on the frequent item set is developed for the analysis of data. The data set is taken and analyzed to know which analysis will be better for prediction. According to the resulting factors, the performance of J48 Decision Tree classifier is more accuracy and more precision than Naïve Bayes and Random Forest classifiers. While evaluating teaching skill**,** evaluation's score of teaching is very important factor that many universities gather this information to improve the level of education quality of that university. By the analysis using data mining, J48 decision tree algorithm is the best result that education managers could use this rule in future decisions to submit new teachers and continue to accept selected old teacher, and then the education managers must give teacher training for qualified teacher. Like training support teaching skills as well as language skills, the trained teachers are better in all aspects of what a professional teacher should know about teaching and learning. Teachers are the key persons and they have knowledge of implementing curriculums and the use of materials to create effective learning.

7.REFERENCES

[1]. Jiawei Han, Micheline Kamber, JiamnPei, "Data Mining: Concepts and Techniques", 3nd edition.

[2]. Sunita B Aher, Mr. LOBO L.M.R.J. Data Mining in Educational System using WEKA, IJCA,2011

[3]. A. P. Sanjeev ve J. M, Zytkow. "Discovering Enrollment Knowledge in University Databases," 1th Conference on KDD (Montreal.20-21 August 1995), 246.

[4]. K. Becker, C. Ghedini ve E.L. Terra, "Using KDD to analyze the impact of curriculum revisions in a Brazilian university," SPIE 14th Annual International Conference (Orlando. April 2000), 412.

[5]. Asanbe, M. O., Osofisan, A. O., and William, W. F. 2016. Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique.

[6]. Kumar, S.2017.A Modern Data Mining Method for Assessment of Teaching Assistant in Higher Educational Institutions.

[7]. Wadkar, S.B., Dharmadhikari, S. C., and Dwivedi, S. K. 2017.A Survey on Performance Evaluation and Prediction of Teachers', Students Performance Using Data Mining Techniques.

[8]. Turhan, K., Yaris, F., and Nural, E. 200. Does Instructor Evaluation by Students Using a Web-Based Questionnaire Impact Instructor Performance.

[9]. Naeem, H.,Sang, S.,Abro, W. A., and Naeem, M. R. 2015. A generic feedback system for better evaluation of teacher performance.

[10]**.** Ş.Erdoğan, M. Timor, "A Data Mining Application in a Student Database," Havacılık ve Uzay Dergisi. Cilt No 2,Sayı 2: 57-64, (July 2005), 57.

[11]. M.Vranić, D. Pintar, Z.Skoćır, "The Use of Data Mining in Education Environment," ConTEL 2007 (Zagrep 13- 15 June 2007), 243.

[12]. International Educational Data Mining, available **at** **http://www.educationaldatamining.org/**

BIOGRAPHIES

Myint Myint Than is a lecturer at University of Computer Studies, Kalay city, Sagaing division, Burma.
She received the M.I.Sc. (degree in Master of Information Science) from Mandalay Computer University, Burma, in 2001.