SERVICES FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 8, NO. 2
ISSN: 2054 -7420

A Comparative Analysis of Classification Algorithms on Students' Performance

Nilar Aye
University of Computer Studies, Yangon
nilaraye@ucsy.edu.mm

ABSTRACT

Recently educational system, many features control a student's performance. Students should be well stimulated to study their education. Motivation leads to interest, interest leads to success in their lives. Appropriate assessment of abilities encourages the students to do better in their education. Data mining is to find out patterns by analyzing a large dataset and apply those patterns to predict the possibility of the future events. Data mining is a very critical field in educational area and it provides high potential for the schools and universities. In data mining, there are various classification techniques with various levels of accuracy. This paper focuses to make comparative evaluation of four classifiers such as J48, Naive Bayesian, Bayesian Network and Decision Stump by using WEKA tool. This study is to investigate and identify the best classification technique to analyze and predict the students' performance of University of Jordan.

Key words:—Data Mining, Error Measurement, Accuracy, Naive Bayesian, BayesNet, J48, Decision Stump

1 Introduction

Data mining is not just collection and handling data; it also constitutes an analysis of data for deriving interesting knowledge from it [11]. Data mining is taking part in an important role in educational associations and the most important fields of research area which has the objective of finding relevant information from the data stored in huge dataset [4]. Education is transferring knowledge to the students and developing their intuitive quality [1]. Educational data especially students' performance is needed to examine and analyze. Educational Data Mining (EDM) is the study concerned with mining educational data to identify interesting patterns and knowledge in educational associations [5]. The student's performance is one of most fundamental needs for the quality education. Analyzing students' data and information which is an interesting field of research enhances students' performance. The educational system is vital for any society, therefore, the data mining researchers aim attention at EDM and it has developed as a research area in recent years. Educational data mining is a useful technique for extracting the new knowledge of educational data and it can be used for improved prediction of the student's performance, academic achievement, subject importance etc. The education data have been organized from distinct educational surveys, school records and data mining technique such as classification can be applied for the enhancement of academic achievement.

It is necessary to apply data mining in education for the undergraduates and academics responsible. The data mining result provides knowledge and pattern, and students, academic responsible and

DOI: 10.14738/tnc.82.8267
Publication Date: 30th April, 2020
URL: http://dx.doi.org/10.14738/tnc.82.8267

educators employ it. In order to improve the performance of student, the academics responsible and educators worked upon the educational system [21]. The most simple and most efficient data mining technique which is used to classify and predict value is classification. Various different classification methods and techniques are used in Knowledge Discovery and Data Mining. There are own advantages and disadvantages of using every method or technique [5].

The objective of this study is to choose the best classifier for prediction of students' performance by comparing the basis analysis of accuracy and recall percentage of four classifiers such as J48, Naive Bayesian, BayesNet and Decision Stump by using WEKA tool.

2 Related Work

In this section, a literature review of numerous surveys and comparative analysis made on the classification techniques used on educational dataset are presented. R. Sumitha, and Vinothkumar [16] examined and made a comparative analysis for Classification algorithms on students' data set and discovered that J48 generates better accuracy of 97%. The researchers in [18] proposed a comparative analysis of various decision tree algorithms to classify the students' educational performance. They wanted to choose the perfect decision tree algorithm from among frequently used decision tree algorithms and give a standard for each of them. They examined that Classification and Regression Tree method generated better on the tested dataset, according to the resulted accuracy and precision using 10-fold cross validations.

AbdulHamid and Amin [3] examined and compared classification algorithms on students' enrolment approval and C4.5 generated high accuracy and lowest absolute errors. Trilok Chand Sharma and Manoj Jain [22] explained about the classification algorithms and discussed how the selected dataset was run by using those classifiers and found better performance and high accuracy are generated by using decision tree.

Nguyen N., Paul J. & Peter H., 2007 [15] made a comparative study on the accuracy of decision tree and Bayesian network algorithms to predict the students' performance of Under Graduate and Post Graduate students. They found that the decision tree classifier gives better accuracy than the Bayesian network classifier. Pandey A.K and Rajpoot D.S. [2] compared different classification algorithms on the school students who are alcohol consumption by applying the WEKA tool. Depending on the basis of the distinct parameters like correctly classified and incorrectly classified instances, accuracy etc., the classification has been examined. Decision stump algorithm gives the better classification than the other classifiers according to the generated result.

Md. N. Amin, Md. A. Habib, 2015 [14] compared decision tree, multilayer perception, and Naïve Bayes. The research work proved that the best algorithm is J48 with an accuracy of 97.61%, and the algorithm which is having lowest error rate with 27.91% is Naïve Bayes. Kotsiantis, S., C. Pierrakeas, P. Pintelas [12] compared five classification algorithms (Decision Tree, Perceptron-based Learning, Bayesian Net, Instance Based Learning and Rule-learning) to predict computer science students' performance from distance learning.

3 Datamining definition and techniques

Data mining also called Knowledge Discovery in Databases (KDD) is defined to extract or "mine" knowledge from huge amounts of data [9].

Data mining is a process that transforms raw data into useful information. Figure 1 presents how raw data is transformed into useful information.

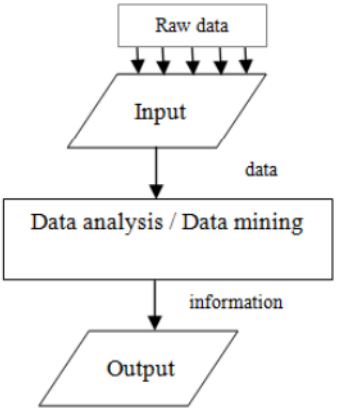


Figure 1. Knowledge discovery

The unprocessed data's necessary to be collected, stored and maintained properly before they are applied to any of the mining techniques. These mining techniques automatically searches huge store of data and discover patterns. Data mining relies on adequate data collection and warehousing as well as computer processing [10].

Knowledge Discovery process consists of different processes such as Data cleaning, transformation, data mining, pattern evaluation in extracting knowledge from data. Knowledge Discovery is employed in a multitude of tasks like association, clustering, classification, prediction, etc. [9].

3.1 Classifications

Classification is a supervised learning technique and it has the aim of generating a model, especially called classifier, which can classify the class label of unknown data. In another word, a training dataset is used to create a classifier and it is then applied to generate the class label of unknown data, into one of the existing classes. A two-step process is carried out during classification: learning phase and the classification phase. In the classification phase, one can classify any attribute vector applying the mapping function. To assess the classifier, it examines an already classified input and then its accuracy is calculated when the correct classification' percentage is obtained [17]. Several algorithms are used under classification tasks that have been applied to predict the students' performance. They are Decision tree, Bayesian Classifier, Artificial Neural Network, Support Vector Machine and K-Nearest Neighbor algorithms.

A. Naive Bayes Classifier

Naive Bayes is defined as a simple method which is used for developing up classifiers, also called the models that allocate the labels of class to problem instances, expressed as vectors of feature algorithm independent values, where some finite set produces the class labels. Classifiers is not only algorithm to do training, but a group of algorithms based on a common principle that every naive Bayes classifiers accept that a particular feature's value is of the value of any other feature, given the class variable. The Naïve Bayes means that a supervised learning methodology yet as a statistical method for classification. It assumes an underlying probabilistic model and it grants to get uncertainty regarding the model in a principled way by deciding probabilities of the outcomes. It will clarify diagnostic and predictive problems.

Bayes theorem is as follows: $P(H/X)=(P(X/H) . P(H)) / P(X)$ [19]

B. BayesNet Classifier

BayesNet or Bayesian Network is a graphical explanation of conditional probabilities. It means random attributes and conditional dependencies using a direct graph and nodes represent random variables. Bayesian Network generates computation process easier and better speed and accuracy for large database [8].

C. C4.5 (J48) Classifier

Depending on different attribute values of the possible data, J48 decision tree [23] also called predictive machine-learning model determines the target value (dependent variable) of a new pattern. In a decision tree, the inner nodes stand for the various attributes; the branches between the nodes are the possible values that these attributes will have within the detected patterns, whereas the terminal nodes are the final value (classification) of the dependent variable. The predicted attribute is assumed as the target value, because its value depends upon, or is determined by, the values of all the other attributes. The other attributes, which are used to predict the value of the dependent variable, are called as the independent variables in the dataset. It explores information gain for selecting an attribute to separate the data. To generate the decision, the attribute with maximum normalized information gain, i.e. the attribute that discriminates the various instances most clearly is applied.

D. Decision Stump Classification Algorithm

A decision stump [24], a machine learning model, consists of a one-level decision tree, i.e., it is a decision tree which has one internal node (the root) immediately linked to the terminal nodes (its leaves). A prediction is generated based on a single input feature's value which is also called 1-rules. Many variations are accessible depending on the type of the input feature. For simple features, one can develop a stump containing a leaf for each available feature value or a stump with the two leaves, one of which corresponds to some selected category, and the other leaf to all the other categories. These two schemes are same for binary features. A missing value may be considered as a yet another category. For extended features, usually, some threshold feature value is chosen, and the stump has two leaves — for values below and above the threshold. However, infrequently, various thresholds may be selected and the stump therefore consists of equal or more than three leaves.

4 Scope and Sample of the Study

Dataset, xAPI-Edu-Data is gathered from a Learning Management System Kalboard-360 of University of Jordan, and the dataset contains of 480 student records and 16 features. x-API-Data was a complete data [6]. A sample description of the attributes of data is given in Table 1 [7].

5 Methodology of the Study

The system makes a comparative analysis of four classifiers on the students' performance by using the WEKA data mining tool. University of Waikato, New Zealand developed WEKA (Waikato Environment for Knowledge Analysis) which is an outstanding suite of machine learning software developed using Java language. WEKA is a free software available under the GNU General Public License [9].

WEKA provides many typical data mining tasks, especially, data pre-processing, clustering, classification, regression, visualization, and feature selection. The researchers collected the students' performance data and then they transferred these data into excel sheet. And then they selected the

attributes which were required to analyze and changed these dataset into .csv file format. Pre-processing steps like cleaning and integration process are done on the gathered data. And then the

system transformed the cleaned and integrated data into ARFF formats and then classified the dataset by using J48, Naive Bayesian, Bayesian Network and Decision Stump. The classified data was visualized in graphs. According to the comparison result of the above four algorithms, the system generated the final decision and suggestion on different classifiers of the study [20].

Table 1. Attributes descriptions and domain values

Attributes	Description	Possible Values
Gender	Sex of students	{M,F}
Nationality	Student nationality	{KW, Lebanon, Egypt, SaudiArabia, USA, Jordan, Venzuela, Iran, Tunis, Syria, Morocco, Iraq, Palestine, Lybia}
Place of Birth	Birth place for the student	{KW, Lebanon, Egypt, SaudiArabia, USA, Jordan, Venzuela, Iran, Tunis, Syria, Morocco, Iraq, Palestine, Lybia}
StageID (Educational Stages)	Stage student belongs	{Lowerlevel, Middleschool, HighSchool}
GradeID	Student' grade	{G-01, G-02, G-03, G-04, G-05, G-06, G-07, G-08, G-09, G-10, G-11, G-12}
SectionID	Section name of Student	{A, B, C}
Topic	Couse Topic	{Math, English, IT, Arabic, Science, Quran, French, Spanish, History, Biology, Chemistry, Geology}
Semester	Semester Name	{First, Second}
Relation	Parent responsible for student	{Father, Mother}
Raised Hand	Raised Hand on class	Numbers of time
VisITedResources	Visited resources	Numbers of time
AnnouncementsView	Viewing announcements	Numbers of time
Discussion	Student Participation on learning process	Numbers of time
ParentAnsweringSurvey	Parents Participation on learning process	Yes, No
ParentschoolSatisfaction	Parent satisfaction level from school	{Good, Bad}
StudentAbsenceDays	Student absence days	{Above-7, Under-7}
Result	Result from mark calculation	{PASS, FAIL}

6 Measure Performance

Calculating the accuracy of different classifiers contributes analyzing their performance. However, since classification is usually a fuzzy problem, the correctness of the classifier may depend on the user. To calculate the classification accuracy, the percentage of tuples allocated in a right class can be used. This avoids the very fact that there also may be a cost correlated with an incorrect assignment to the wrong class [13].

6.1 Confusion matrix:

In a classification problem, a confusion matrix is used to calculate the accuracy of the solution. Suppose the dataset has n classes, a m x n matrix is generated as a confusion matrix, where Ci,j

represents the number of tuples from D that were assigned to class Ci,j but where Ci is the correct class. Evidently in outside the diagonal, if the value is only zero that will be the best solution [13].

A classification system generates a confusion matrix which contains information about actual and predicted classifications. System's performance is generally calculated using the data in the matrix.

In the confusion matrix, the elements in the matrix represent the following meanings.

1. ***a*** = the correct predictions' number that an instance is negative,

2. ***b*** = the incorrect predictions' number that an instance is positive,

3. ***c*** = the incorrect of predictions' number that an instance negative, and

4. ***d*** = the correct predictions' number that an instances positive [13].

Some standards and terms:

1. True positive (TP): If the result from a prediction is p and the actual value is also p, then p is called a true positive.
2. False positive (FP): If the actual value is n then n is said to be a false positive.
3. Precision and recall: Precision is the division of retrieved instances that are relevant, while recall is the division of relevant instances that are retrieved. Both of them are therefore based on an understanding and measure of relevance. Precision assumes as a measure of exactness or quality, whereas recall assumes as a measure of completeness or quantity. Recall is nothing but the true positive rate for the class [13].

In this paper, we have analyzed the students' performance dataset using WEKA tool for comparison of four classifiers. According to applying classifiers, this tool generates a relevant confusion matrix for each classifier and then depending on accuracy regarding correct and incorrect instances provide to calculate efficiency of each classifier.

7 Experimental work and result

In this study, we have executed classification using J48, Naive Bayesian, BayesNet and Decision Stump algorithms on x-API-Data.arff dataset in WEKA tool. WEKA tool provides inbuilt algorithms for them. Firstly, x-API-Data.csv dataset is transformed into x-API-Data.arff shown in Figure 2 and Data Pre-processor output and model visualization are shown in Figure 3 and Figure 4.

The students' performance dataset has three possible classes (Result) namely - M (Medium), L (Low) and H (High).

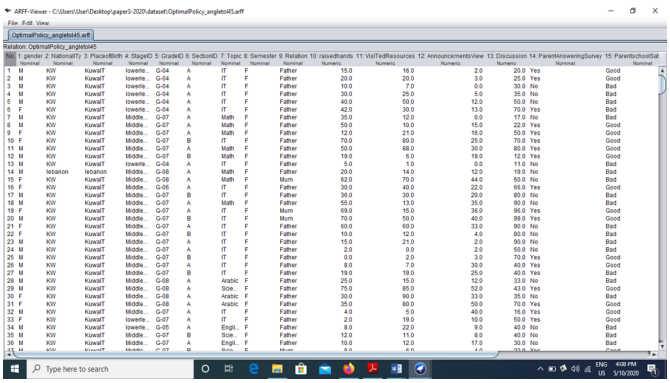


Figure 2. x-API-Data.arff dataset



Figure 3. Data Pre-processor output

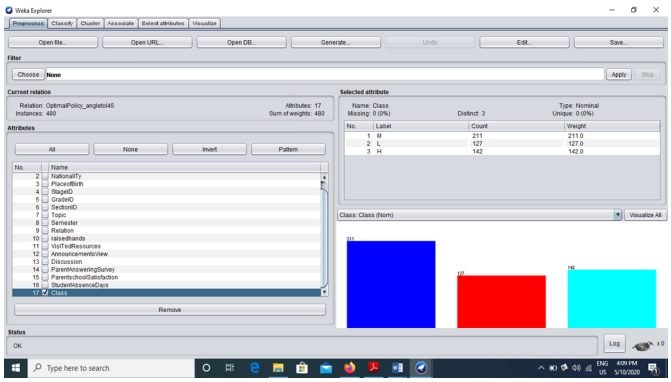
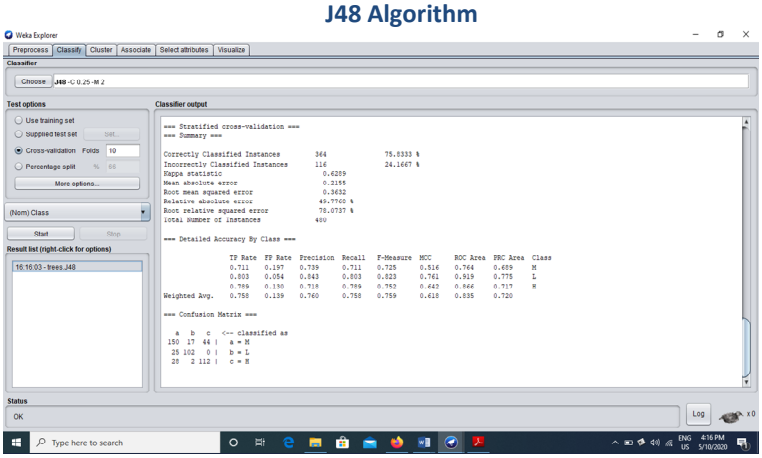


Figure 4. Model visualization

A. Result for classification using J48:

J48 Algorithm



When J48 is applied on the dataset, 364 instances are correctly classified out of 480 instances which shows success rate of75.8 % and the tool generates the confusion matrix for performance result with three possible values (M, L, & H).

Confusion matrix:

a	b	c	← classified as -----
150	17	44 |	a = M
25	102	0 |	b =L
28	2	112 |	c = H

For above confusion matrix, true positive for class $\boldsymbol{a} = \text{"M"}$ is 150 while false positive is 61 whereas, for class $\boldsymbol{b} = \text{"L"}$, true positive is 102 and false positives is 25 and also true positives for class $\boldsymbol{c} = \text{"H"}$

is 112 and false positive is 30 i.e. diagonal elements of matrix $150 + 102 + 112 = 364$ represents the correct instances classified and other elements 17+44+25+0+30= 116 represents the incorrect instances.

$$\text{True positive rate} = \frac{diagonal\ element}{sum\ of\ relevant\ row} \quad (1)$$

$$\text{False positive rate} = \frac{non-diagonal\ element}{sum\ of\ relevant\ row} \quad (2)$$

Hence,

TP rate for class M $= \frac{150}{150+17+44} = \frac{150}{211} = 0.711$

FP rate for class M $= \frac{25+28}{25+102+0+28+2+112} = \frac{53}{269} = 0.197$

TP rate for class L $= \frac{102}{25+102} = \frac{102}{127} = 0.803$

FP rate for class L $= \frac{17+2}{150+17+44+28+2+112} = \frac{19}{353} = 0.054$

TP rate for class H $= \frac{112}{112+28+2} = \frac{112}{142} = 0.789$

FP rate for class H $= \frac{44+0}{150+17+44+25+102+0} = \frac{44}{338} = 0.130$

$$\text{Precision} = \frac{\text{diagonal element}}{\text{sum of relevant column}} \quad (3)$$

Precision for class M $= \frac{150}{150+25+28} = \frac{150}{203} = 0.739$

Precision for class L $= \frac{102}{17+102+2} = \frac{102}{121} = 0.843$

Precision for class M $= \frac{44}{44+0+112} = \frac{44}{156} = 0.718$

$$\text{F-measures} = \frac{2*\text{Precision}*\text{recall}}{\text{precision}+\text{recall}} \quad (4)$$

F-measures for class M $= \frac{2*0.739*0.711}{0.739+0.711} = 0.725$

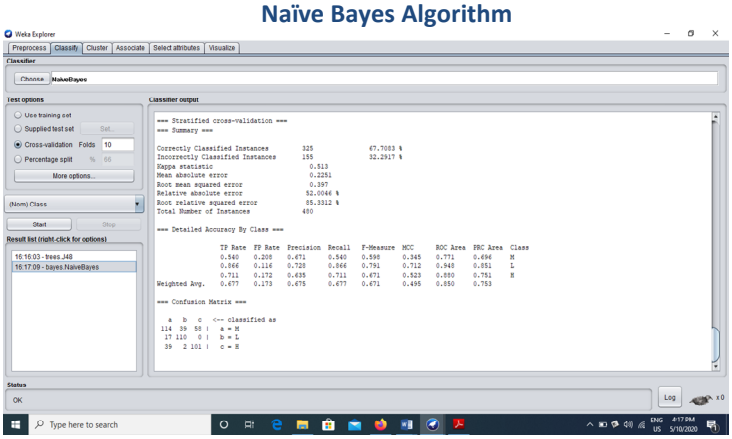
F-measure for class L $= \frac{2*0.843*0.803}{0.843+0.803} = 0.823$

F-measure for class H $= \frac{2*0.718*0.789}{0.718+0.789} = 0.752$

B. Result for classification using Naïve Bays:

Naive Bayes algorithm is chosen as a classifier and it is applied on the data set and the confusion matrix is generated for class = “Result” having three possible values (M, L and H). 325 instances are correctly classified and 155 instances are incorrectly classified among 480 instances applying this algorithm.

Naïve Bayes Algorithm



Confusion matrix:

a	b	c	← classified as -----
114	39	58 |	a = M
17	110	0 |	b =L
39	2	101 |	c = H

For above confusion matrix, true positive for class $\boldsymbol{a} = "M"$ is 114 while false positive is 97 whereas, for class $\boldsymbol{b} = "L"$, true positive is 110 and false positive is 17 and also true positives for class $\boldsymbol{c} = "H"$ is 101 and false positive is 41 i.e. diagonal elements of matrix $114 + 110 + 101 = 325$ represents the correct instances classified and other elements $39 + 58 + 0 + 17 + 39 + 2 = 155$ represents the incorrect instances.

Hence,
By using equation (1) and (2), the followings TP rate and FP rate of class label "M", "L" and "H" can be calculated as follow.

TP rate for class M = $\frac{114}{114+39+58} = \frac{114}{211} = 0.540$

FP rate for class M = $\frac{17+39}{17+110+0+39+2+101} = 0.208$

TP rate for class L = $\frac{110}{17+110} = 0.866$

FP rate for class L = $\frac{39+2}{114+39+58+39+2+101} = 0.116$

TP rate for class H = $\frac{101}{39+2+101} = 0.711$

FP rate for class H = $\frac{58+0}{114+39+58+39+2+101} = 0.172$

By using equation (3), the followings precision of class label "M", "L" and "H" are generated.

$$\text{Precision for class M} = \frac{114}{114+17+39} = 0.671$$

Precision for class L = $\frac{110}{39+110+2} = 0.728$

Precision for class H = $\frac{101}{58+0+101} = 0.635$

By using equation (4), the followings F-measures of class label "M", "L" and "H" are calculated as follows.

F-measures for class M = $\frac{2*0.671*0.540}{0.671+0.540} = 0.598$

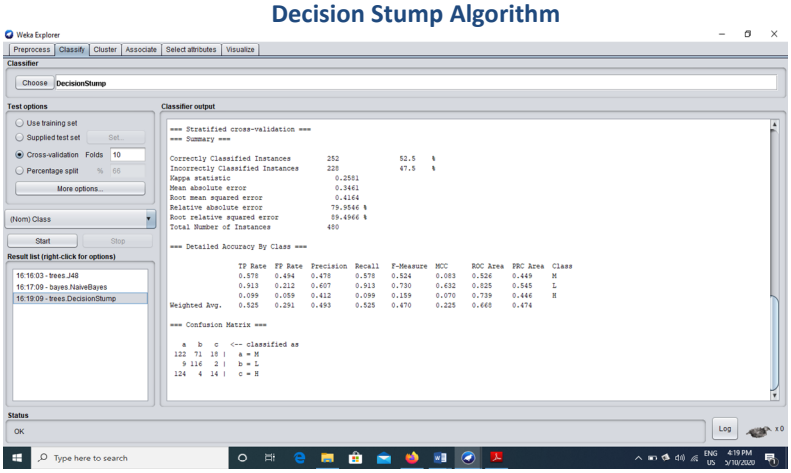
F-measure for class L = $\frac{2*0.728*0.866}{0.728+0.866} = 0.791$

F-measure for class H = $\frac{2*0.635*0.711}{0.635+0.711} = 0.671$

C. Results for classification using Decision Stump:

The tool is activated on Decision stump algorithm on the data set and generates the confusion matrix for performance result with class = "Result" having M, L & H. The number of correctly classified instances are 252 and 228 instances are incorrectly classified among existing instances 480.

Decision Stump Algorithm



Confusion Matrix:

a	b	c	← classified as -----
122	71	18 |	a = M
9	116	2 |	b =L
124	4	14 |	c = H

For above confusion matrix, true positive for class $a = "M"$ is 122 while false positive is 89 whereas, for class $b = "L"$, true positives is 116 and false positives is 11, and also true positive for class $c = "H"$ is 14 and false positive is 128 i.e. diagonal elements of matrix $122 + 116 + 14 = 252$ represents the correct instances classified and other elements $71 + 18 + 9 + 2 + 124 + 4 = 228$ represents the incorrect instances.

Hence,

By using equation (1) and (2), the followings TP rate and FP rate of class label “M”, “L” and “H” are generated.

TP rate for class M = $\frac{122}{122+71+18} = \frac{122}{211} = 0.578$

FP rate for class M = $\frac{9+124}{9+116+2+124+4+14} = 0.494$

TP rate for class L = $\frac{116}{9+116+2} = 0.913$

FP rate for class L = $\frac{71+4}{122+71+18+124+4+14} = 0.212$

TP rate for class H = $\frac{14}{124+4+14} = 0.099$

FP rate for class H = $\frac{2+18}{122+71+18+9+116+2} = 0.059$

By using equation (3), the followings precision of class label “M”, “L” and “H” are generated.

Precision for class M = $\frac{122}{122+9+124} = 0.478$

Precision for class L = $\frac{116}{116+71+4} = 0.607$

Precision for class H = $\frac{14}{18+2+14} = 0.412$

By using equation (4), the followings F-measures of class label “M”, “L” and “H” are generated.

F-measures for class M = $\frac{2*0.478*0.578}{0.478+0.578} = 0.524$

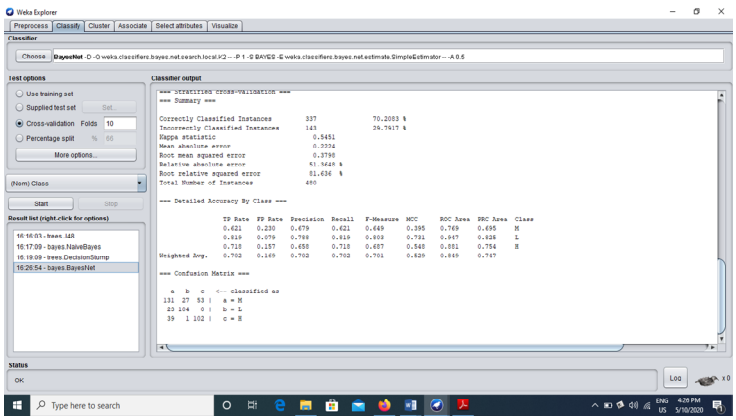
F-measure for class L = $\frac{2*0.607*0.913}{0.607+0.913} = 0.730$

F-measure for class H = $\frac{2*0.412*0.099}{0.412+0.099} = 0.159$

D. Results for classification using BayesNet

The tool selects the BayesNet algorithm and the algorithm is applied on the data set and the confusion matrix is generated for class= “Result” with three values (M, L & H). Among 480 instances, the above algorithm correctly classifies 337 instances and incorrectly classifies 143 instances.

BayesNet



Confusion matrix:

a	b	c	← classified as -----
131	27	53 |	a = M
23	104	0 |	b =L
39	1	102 |	c = H

For above confusion matrix, true positive for class $\boldsymbol{a} = "M"$ is 131 while false positive is 80 whereas, for class $\boldsymbol{b} = "L"$, true positive is 104, and false positive is 23 and also true positive for class $\boldsymbol{c} = "H"$ is 102 and false positive is 40 i.e. diagonal elements of matrix $131 + 104 + 102 = 337$ represents the correct instances classified and other elements $27 + 53 + 23 + 0 + 39 + 1 = 143$ represents the incorrect instances.

Hence,
By using equation (1) and (2), the followings TP rate and FP rate of class label "M", "L" and "H" are generated.

TP rate for class M = $\frac{131}{131+27+53}$ = $\frac{131}{211}$ =0.621

FP rate for class M = $\frac{23+39}{23+104+0+39+1+102} = 0.230$

TP rate for class L = $\frac{104}{104+23+0} = 0.819$

FP rate for class L =$\frac{27+1}{131+27+53+39+1+102} = 0.079$

TP rate for class H =$\frac{102}{39+1+102} = 0.718$

FP rate for class H= $\frac{53+0}{131+27+53+23+104+0} = 0.157$

By using equation (3), the followings precision of class label "M", "L" and "H" are generated.

Precision for class M= $\frac{131}{131+23+39} = 0.679$

Precision for class L =$\frac{104}{27+104+1} = 0.788$

Precision for class H=$\frac{102}{53+0+102} = 0.658$

By using equation (4), the followings F-measures of class label "M", "L" and "H" are generated.

F-measure for class M=$\frac{2*0.679*0.621}{0.679+0.621} = 0.649$

F-measure for class L =$\frac{2*0.788*0.819}{0.788+0.819} = 0.803$

F-measure for class H =$\frac{2*0.658*0.718}{0.658*0.718} = 0.687$

This experiments make a comparative analysis of four classifiers namely J48, Naïve Bayes, Decision Stump and BayesNet. After applying the WEKA tool on selected 480 instances, the performance measure of each classifier is generated. Each of these algorithms has its own characteristics to classify the data set with different measure. Performance results of all classifiers using WEKA tool are shown in Table 2.

Table 2. Experimental Result of four classifiers' performance

Criteria	Classifiers			
	Naive Bayes	Bayes Net	J48	Decision Stump
Accuracy (%)	67.71%	70.21%	75.83%	52.50
Correctly classified Instances	325	337	364	252
InCorrectly classified Instances	155	143	116	228

According to the generated values in Table 2, the J48 classifier is the best accuracy model because it has more correctly classified instances and the least incorrectly classified instances than other classifiers. The data in Table 2 is represented as the graphical demonstration shown in Figure 5. This figure proves that J48 classifier is the best classifier to predict the students' performance based on their dataset.

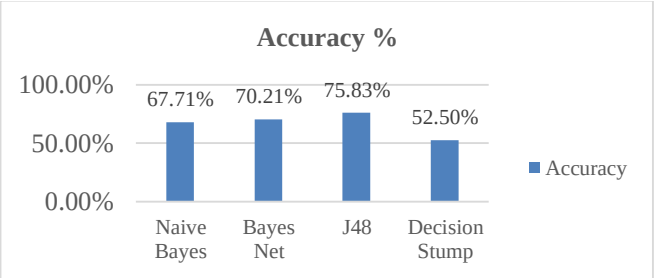


Figure 5. Accuracy performance of four classifiers

Table 3 represents the error measures of all classifiers; it shows that J48 has the lowest error based on all error measures such as MAE, RMSE, RAE and PRSE among other classifiers as shown in Figure 6.

Table 3. Error measures in WEKA

Criteria	Data Mining Classifiers			
	Naive Bayes	BayesNet	J48	Decision Stump
Mean Absolute Error (MAE)	0.2251	0.2224	0.2155	0.3461
Root Mean Squared Error(RMSE)	0.397	0.3798	0.3632	0.4164
Relative Absolute Error (RAE)	52.0046%	51.3648%	49.7768%	79.9546%
Root Relative Squared Error(RRSE)	85.3312%	81.6360%	78.0737%	89.4966%

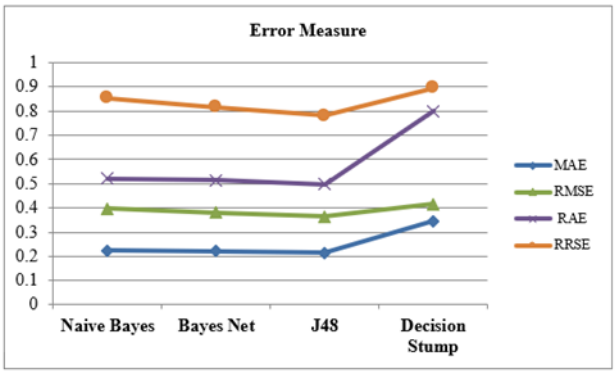


Figure 6. Error Measures

Table 4. Weighted average of class label accuracy

Criteria	Data Mining Classifiers				
	TP	FP	Precision	Recall	F-measure
Naive Bayes	0.677	0.173	0.675	0.677	0.671
BayesNet	0.702	0.169	0.702	0.702	0.701
J48	0.758	0.139	0.760	0.758	0.759
Decision Stump	0.525	0.291	0.493	0.525	0.47

In Figure 7 Precision, Recall and F-measures analyzed among all classifiers according to the generated result of WEKA tool shown in Table 4. It proves that the weighted average of J48 outperform other classifiers.



Figure 7. Classifiers performance metrics

8 Conclusion

The need of prediction over student performance is to provide teachers and parents to concentrating their students and children to improve their performance as well as researcher to select among the classification algorithms to find the best classifier for prediction the students' performance. Four classifiers are applied on the student data set and comparisons are analyzed based on the accuracy among these classifiers and different error measures are generated to determine the best classifier. Experimental results prove that J48 algorithm has the highest accuracy and lowest absolute error among other classifiers to predict the students' performance.

REFERENCES

[1] A. Dinesh Kumar, V.Radhika, "Mining Educational Data to Predicting Higher Secondary Students Performance", International Journal of Computational Intelligence and Informatics, Vol.6: No.2, September 2016, ISSN:2349-6363,pp:124-129

[2] A. K. Pandey and D. S. Rajpoot, "A Comparative Study of Classification Techniques by utilizing WEKA," IEEE, pp. 219–224,

[3] Abdul Hamid andAmin, “AComparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining”, Workshop on Interaction Design in Educational Environments,ISBN: 978-1-4503-3034-3,2014

[4] Agnik Dey, Abhirup Khasnabis, Ajeet Kumar, "Prediction and Analysis of Student Performance by Data Mining in WEKA", RCC Institute of Information Technology

[5] Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No.5, 2016

[6] Anoopkumar M, A.M.J.Md. Zubair Rahman, "Model of Tuned J48 Classification and Analysis of Performance Prediction in Educational Data Mining", International Journal of Applied Engineering Research ISSN 0973-4562 Volumn 13, Number 20 (2018) pp.14717-14727

[7] Elaf Abu Amrieh, Thair Hamtini and Ibrahim Aljarah, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods", International Journal of Database Theory and Application, Vol.9, No.8 (2016), pp.119-136, ISSN:2005-4270-IJDTA

[8] Hilal Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers", I.J.Modern Education and Computer Science, 2017, 8, 9-15

[9] Jai Ruby, Dr.K.David, "Predicting the Performance of Students in Higher Education using Data Mining Classification Algorithms- A Case Study", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volun 2, Issue XI, November 2014, ISSN:2321-9653

[10] Keshav Singh Rawat, "Comparative Analysis of Data Mining Techniques, Tools and Machine Learning Algorithms for Efficient Data Analytics", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN:2278-0661,p-ISSN:2278-8727, Volumn 19, Issue 4, Ver.III (Jul.-Aug.2017),PP 56-61

[11] Kishansingh Rajput, Bhavesh A. Oza, "A Comparative Study of Classification Techniques in Data Mining", 2017 IJCRT, Volumn 5, Issue 3 September 2017,2017 IJCRT,ISSN:2320-2882

[12] Kotsiantis, S., C. Pierrakeas, P.Pintelas, "Prediction of Student’s Performance in Distance Learning Using Machine Learning Techniques", Applied Artificial Intelligence,Vol.18,2004,No5,411-426

[13] Margaret H. Danham,S. Sridhar, "Data mining, Introductory and Advanced Topics”, Person education , 1st ed., 2006

[14] Md. N. Amin, Md. A. Habib, “Comparison of Different Classification Techniques using WEKA for Hematological Data”, American Journal of Engineering Research, vol. 4, PP 55-61, No. 3(2015).

[15] Nguyen N., Paul J., &Peter H., "A Comparative Analysis of Techniques for Predicting Academic Performance", In Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference. pp. 7-12, 2007

[16] R. Sumithaand Vinothkumar, “Predictionof Students Outcome Using Data Mining Techniques”, International Journal of Scientific Engineering and Applied Science,Volume-2, Issue-6, June 2016

[17] Ricardo Mendes And Joao P.Vilela, "Privacy- Preserving Data Mining: Methods, Metrics, and Applications”, IEEE, 2017.

[18] S.K Yadav, B. Bharadwaj, and S. Pal, 2012, "Data Mining Applications: A Comparative Study for Predicting Student's Performance", International Journal of Innovative Technology & Creative Engineering (ISSN: 2045-711), Vol. 1, No.12, December.

[19] Sharon Carl, Glaston D'souza, Linet Varghese, "Implementation of Classification Algorithms and their Comparison for Educational Dataset", IJISET- International Journal of Innovative Science, Engineering & Technology, Vol. 3, Issue 3, March 2016, ISSN 2384-7968

[20] Sohajbir Singh Ubha, Gaganpreet Kaur Bhalla, "Data Mining for Prediction of Students' Performance in the Secondary Schools of the State of Punjab", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 8, August 2016,ISSNV(Online): 2320-9801, ISSN(Print):2320-9798

[21] Sweta V. Parmar, Lokesh Kumar ,Sharma, "Comparative Study of Supervised Learning for Student Performance Evaluation", International Journal of Computer Engineering & Technology (IJCET), Volumn 9, Issue 2, March-april 2018, pp:32-38, Artical ID:IJCET_09_02_003

[22] Trilok Chand Sharma and Manoj Jain, "WEKA Approach for Comparative Study of Classification Algorithm", International Journal of Advanced Research in Computer and Communication Engineering,Vol. 2, Issue 4, April 2013.

[23] Wikipedia contributors, "C4.5_algorithm," Wikipedia, The Free Encyclopedia. Wikimedia Foundation, 28-Jan-2015.

[24] Wikipedia contributors, "Decision_stump," Wikipedia, The Free Encyclopedia. Wikimedia Foundation, 23-Aug-2012