IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 7, July 2020
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

Prediction of Students' Performance Using A Combinational Classifiers

1 Dr. Yasir Mohamed Ahamed 1Assistant and 2 Dr. Fatima Dafallah Mohammed Elhassan

1Assistant Professor in Holy Quran University, The Republic of The Sudan, University of the Holy Quran and Islamic Sciences, Faculty of Computer Science and Information Technology, Computer Science Department

2 Professor in King Khalid University, Kingdom of Saudi Arabia, King Khalid University, Faculty of Science and Arts, Majardh, Computer Science Department.

Abstract

Research studies on educational data mining are on the increase due to the benefits obtained from the knowledge acquired from machine learning processes which help to improve decision making processes in higher institutions of learning.

The ability to predict a student's performance could be useful in a great number of different ways associated with university-level distance learning. There are many data mining classification techniques with different levels of accuracy In this study the main subjective Predict the grade of the final year of the students, using the GPA of the first three years of study using classification method with five algorithms and compare between these algorithms individual and combined by using WEKA tool.

Five classifiers are used individual To choice the best classifiers with highest accuracy Different performance measures are used to compare the results between these classifiers The results shows that logistic regression and smv has the highest accuracy among the other classifiers and using a combined model for a comparative performance analysis of the result of each of the combined two algorithms. A maximum accuracy of 90.277% was achieved, using combination of logistic regression and smv for performance validation. This indicates that indeed the graduating results of engineering students in University, in the fifth and final year of study can be reasonably predicted using their performance in the first three academic year with highest accuracy. there are some open problems in ensemble of classifiers, such as how to understand and interpret the decision made by an ensemble of classifiers because an ensemble provides little insight into how it makes its decision. For learning tasks such as data mining applications where comprehensibility is crucial

Keywords—data mining, machine learning, Educational data mining.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 7, July 2020
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

I. Introduction

Educational data mining is a machine learning process that has been applied for studying and predicting student for evaluating learning technologies integration and for identifying learning challenges. Relevant educational dataset that has accumulated overtime and depicts the operational process under evaluation must be available, and adequately processed to support a Data mining analysis toward achieving a reasonable accuracy.

Learning algorithms try to find a hypothesis in a given space H of hypotheses and in many cases if we have sufficient data they can find the optimal one for a given problem. But in real cases we have only limited data sets and sometimes only few examples are available. In these cases the learning algorithm can find different hypotheses that appear equally accurate with respect to the available training data, and although we can sometimes select among them the simplest or the one with the lowest capacity, we can avoid the problem combining them to get a good approximation of the unknown true hypothesis. Thus, there is a growing realization that combinations of classifiers can be more effective than single classifiers. Why rely on the best single classifier, when a more reliable and accurate result can be obtained from a combination of several? This essentially is the reasoning behind the idea of multiple Classifier systems. Bagging, and stacking, is basic parts of a data scientist's toolkit and a part of a series of statistical techniques called ensemble methods

In the first and second year of an engineering program, students are exposed to knowledge on sciences and basic introduction to engineering as a continuum of their secondary school education, and as an introduction to general engineering. In the third year, the curriculum is more focused on the core discipline of each engineering student, that is, electrical engineering, civil engineering, and so forth. By the end of the third year, engineering students are already grounded in the basics of their profession. The academic performance of engineering students from their first year to the third year is very vital in terms of acquisition of foundational knowledge, and its impact on their final graduation Cumulative Grade Point Average (CGPA). It is often said that beyond the third year it is very challenging for a student to move from the current class of grade (first class - 1st, second class upper division – 2|1, second class lower division – 2|2, and third class – 3rd) to a higher one due to the nature of academic courses at fourth year and fifth year which are more robust and touch core foundation of engineering disciplines.

In this study to predict the final year CGPA of an engineering student using the GPA of the first three years of study. The academic performance dataset of students that were admitted and graduated within across seven engineering programs in University was analyzed using a data mining model.

The main subjective Predict the class of grade of the final year of the students,
Using the GPA of the first three years of study using classification method with five algorithms (five) and compare between these algorithms individual and combined

The contraption of this is paper is to provide better result in terms of accuracy using the data mining algorithms individual and combined.

Thy could be deferent accurate Predicted Also, the performance evaluation of these algorithms varies according to specific characteristics. There may be an algorithm that outperforms another in some characteristics and fails in others. This requires taking advantage of all the features of all these algorithms combined

Section 2 Background presents the most well-known methods for Building ensembles that are based on a single learning and data mining prediction Algorithm, while section 3 Experiment results using a number data sets and Comparisons of the presented combining method, using different base classifiers, with other ensembles are presented in section 4. Conclusions and recommendations.

II. BACKGROUND

A. ENSEMBLES OF CLASSIFIERS

There are three main terms describing the ensemble (combination) of various models into one more effective model:

1. Bagging to decrease the model's variance;
2. Boosting to decreasing the model's bias, and;
3. Stacking to increasing the predictive force of the classifier.

The idea here is to train multiple models, each with the objective to predict or classify a set of results. Most of the errors from a model's learning are from three main factors: variance, noise, and bias. By using ensemble methods, we're able to increase the stability of the final model and reduce the errors mentioned previously. By combining many models, we're able to (mostly) reduce the variance, even when they are individually not great, as we won't suffer from random errors from a single source. The main principle behind ensemble modeling is to group weak learners together to form one strong learner. From many, together they emerge as one.

1. Bagging

The first term introduced, bagging, is shorthand for the combination of bootstrapping and aggregating. Bootstrapping is a method to help decrease the variance of the classifier and reduce over fitting, by resampling data from the training set with the same cardinality as the original set. The model created should be less over fitted than a single individual model.

A high variance for a model is not good, suggesting its performance is sensitive to the training data provided. So, even if more the training data is provided, the model may still perform poorly. And, may not even reduce the variance of our model. The simplest approach with bagging is to use a couple of small subsamples and bag them, if the ensemble accuracy is much higher than the base models, it's working; if not, use larger subsamples.that using larger subsamples is not guaranteed to improve your results. In bagging there is a tradeoff between base model accuracy and the gain you get through bagging. The aggregation from bagging may improve the ensemble greatly when you have an unstable model, yet when your base models are more stable been trained on larger subsamples with higher accuracy improvements from bagging reduces.Once the bagging is done, and all the models have been created on (mostly) different data, a weighted average is then used to determine the final score.

2. **Boosting**

Think about optimization of a function over its function space, where optimization can be solved using gradient descent. Vanilla gradient gradient descent is used to minimize a set of parameters. E.g. finding the weights of parameters for a linear regression, through updates from an error function. Parameter estimation seems trivial if we have a smooth convex parameter space, however not all problems provide such a simple plane to traverse over. Our problem, because there many categorial and binary variables it creates a complex gradient with many local minima to get stuck in during the optimization process. For these problems, we can use a different form of gradient descent called boosting.The main idea of boosting is to add additional models to the overall ensemble model sequentially. Previously with bagging, we averaged each individual model created. This time with each iteration of boosting, a new model is created and the new base-learner model is trained (updated) from the errors of the previous learners.The algorithm creates multiple weak models whose output is added together to get an overall prediction. This is ensemble modeling from earlier. The now boosted gradient shifts the current prediction nudging it to the true target, in a similar fashion to how gradient descent moves toward the true values. The gradient descent optimization occurs on the output of the varies models, and not their individual parameters. There are different methods to optimize boosting algorithms, but they are beyond the scope of this article. Unlike the bagging examples above, classical boosting the subset creation is not random and performance will depend upon the performance of previous models. As, each new subset which is iterated upon contains elements which could have been misclassified by previous models. We will also be using the same hard voting we used previously to ensemble the models together.

3. **Stacking**

Stacking is another ensemble model, where a new model is trained from the combined predictions of two (or more) previous model. The predictions from the models are used as inputs for each sequential layer, and combined to form a new set of predictions. These can be used on additional layers, or the process can stop here with a final result. In the example below I'm only going to use one layer for simplicity. Ensemble stacking can be referred to as blending, because all the numbers are blended to produce a prediction or classification. Keep in mind just by adding layers and more models to your stacking algorithm, does not mean you'll get a better predictor. There are no free lunches in machine learning.

B. DATA MINING PREDICTION USING CLASSIFIERS

1 Classification

In the field of Data Mining Classification is a main and most important technique. Class and category of a value is identified using classification based on the previously categorized values. Some of the important classification techniques are discussed. The Classification Algorithms used for Prediction

Naive Bayes Algorithm: Naive Bayes Algorithm is considered as very efficient and easy algorithm. The classification rate is considerably very high and most of the cases it predicts accurate results. This algorithm produces good results only when the data set is very large [1].

Random Forest Algorithm: To improve the predictive accuracy, Random Forest Algorith uses average values of the model. It is a meta-estimator, so it fits a number of decision trees on various sub samples of datasets. It also controls over fittings. The original sample size is always matches the sub-sample size but the samples are drawn with replacement [2].

Neural Network Algorithm: Neural Network Algorithm endeavors to recognize hidden relationships in a set of data through many processes like the way the human brain operates. A Neural Network Algorithm is a series of algorithms and it produces the biggest result without requiring redesigning the output criteria when it adapt to changing input [3].

J48 Algorithm: The ID3 algorithm's features are extended into J48 Algorithm. J48 has the following additional features such as, it accounts the missing values, decision tree pruning, continuous attribute value ranges, derivation of rules, etc. it is an open source java implementation of the C4.5 Algorithm in WEKA Data Mining tool [4].

C4.5 Algorithm: It is also a decision tree algorithm. Some of the features of C4.5 algorithm is as follows, it can be easily interpreted and very easy to implement. It accepts both continuous and discrete values. Some of the limitations are, when a small deviation in data can lead a completely different decision tree. It cannot work with small dataset [1].

C. CLASSIFICATION ACCURACY

Accuracy is defined as the proportion of correct classification from overall number of cases and it depends on confusion matrix. Table 2 shows the confusion matrix that illustrates the number of correct and incorrect predictions made by the classification model compared to the actual value.

1. Correctly classified instance:

The correctly classified instance show the percentage of test instance that were correctly and in correctly classified the percentage of correctly classified instances is often called accuracy or sample accuracy

Kappa statistics:

Kappa is chance –corrected measure of agreement between the classification and true classes.

Confusion matrix:

A confusion matrix,some times called classification matrix is used to assess the prediction accuracy of model . it measure whether amodel is confused or not, that is whether the model is making mistakes in it predictions . The confusion matrix can be obtained from asset of different scales to compare classifications , including accuracy, which is widely used

The classifiers are evaluated by a confusion matrix which is a combination of four outcomes. In binary classification, the output is either positive or negative. The four different classifications are:

True positives (TP)-accurate positive prediction

False positives (FP)-wrong positives prediction

True negatives (TN)-accurate negative prediction

False negatives (FN)-wrong negative prediction

The effectiveness metrics for classifier used in the research are:-

Precision (p):-

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

Number of true positives classifications divided by the sum of true positives and false positive classifications.

- Recall(R):

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

i.e number of true positives classifications divided by the sum of true positive and false negative classifications .

F1-SCORE-

F1-score is the harmonic mean of precision and recall

$$\text{F1} - \text{score} = \frac{2 * \text{P} * \text{R}}{(\text{P} + \text{R})}$$

Accuracy -

Accuracy is measured by dividing the number of correctly classified instances by the total number of instances.

$$\text{accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

$$\text{Error rate} = \frac{F_P + F_N}{T_P + F_P + T_N + F_N}$$

Mean Absolute Error (MAE):-

MAE measures the average magnitude of errors in asset of prediction . it is the summation of the differences between predicted and observation divided by the total number of test samples.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Root Mean Square Error (RMSE):-

It is the square root of the summation of the squared differences between predicted and actual observations, divided by the number of total test samples.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

ROC curve:

It is another way to evaluate the performance of the classification [12] where FP values are represented on the y axis and TP values on the y axis

$$T_{PR} = \frac{T_P}{T_P + F_N}$$

$$F_{PR} = \frac{F_P}{T_N + F_P}$$

Area under curve (AUC):

Another utility called (area under the curve) helps analyze the overall performance of the classification and the ideal classification has AUC.

The ROC is a good visualization tool to identifying the performance of classifier, we times need a numerical value for comparison purpose.

III. Related work

Many research studies have been done in educational data mining to predict the students' performance

In [5]. The key for success of ensembles is whether the classifiers in a system are Diverse enough from each other, or in other words, that the individual classifiers have a minimum of failures in common. If one classifier makes a mistake then the others should not be Likely to make the same mistake. Two of the most popular ensemble algorithms are bagging [6] and boosting [7]. There are two major differences between Adobos is a practical version of the boosting approach [7]. There are two ways that Ad boost can use these weights to construct a new training set to give to the base learning algorithm. In boosting by sampling, examples are drawn with replacement with probability proportional to their weights. The second method, boosting by weighting, can be used with base learning algorithms that can accept a weighted training set directly. With such algorithms, the entire training set (with associated weights) is given to the base-learning algorithm.

MultiBoosting [8] is another method of the same category that can be considered as wagging committees formed by AdaBoost. Wagging is a variant of bagging; bagging uses re resampling to get the datasets for training and producing a weak hypothesis, whereas wagging uses reweighting for each training example, pursuing the effect of bagging in a different way.

Melville and Mooney [9] present a new meta-learner (DECORATE, Diverse Ensemble Creation by Oppositional Re-labeling of Artificial Training Examples) that uses an existing “strong” learner (one that provides high accuracy on the training data) to build a diverse committee. This is accomplished by adding different randomly constructed examples to the training set when building new committee members. These artificially constructed examples are given category labels that disagree with the current decision of the committee, thereby directly increasing diversity when a new classifier is trained on the augmented data and added to the committee.

in [10] S**. Kotsiantis et al (2010)** A combinational incremental ensemble of classifiers as a technique for predicting students’ performance in distance education

It uses machine learning techniques to predict a student's exam score based on the student's data and grades in jobs and oral interviews, where it is based on the integration of three simple works: Naïve Bayes cumulative and the closest neighbor 1-NN and WINNOW linear classification using a voting method The majority, these classifications are used for easy adaptation to new labels for each case.

The database was taken from the University of Wanaya Open to separate informatics where each semester consists of units of study and students of the unit of information were selected INF10, as the database includes 1347 records containing student variables from the registration office and teachers records, with a variable for predicting the student’s success in the final exam or Sedimentation.

in [11] **P.Sittidechet al (2015)** Birth Asphyxia Classification Using AdaBoost Ensemble Method

The aim of this research is to predict new cases of asphyxia in newborns based on the application of AdaBoost technology on three simple classifications, namely BPNN neural networks, SVM vector vectors and DT decision trees, where they are compared using the confusion matrix.

The classifiers were trained and tested using the 10-fold cross validation technique applied to the database which consists of 1142 records where the under-sampling technique was applied, and the number of variables is 8.

in [12] **P. Mahato (2014)** Prediction of Stock Price Movement Using Various Ensemble Models

It studies various machine learning techniques and ways to integrate them to predict the direction of the stock price change in financial markets, and the study indicates that the Bagging method gives the best accuracy with regard to the data bases used,

Majority voting technique was applied for 3, 5 and 7 basic simple compilations, and a random subspace algorithm was applied with the Bagging technique.

In [**13**], the final CGPA of students was predicted using multiple linear regression and correlation to analyse the yearly GPA, and various inferential statistics were developed. The study determined the correlation between the first-year result and the final-year result of the student. With the aid of a regression plot, the students' GPA for the five years of study was fitted using multiple linear regressions in order to explain how the GPA for each year contributed to the variations in the final CGPA of the students at graduation.

In [14] features such as student attendance, average scores, relevant course data, the level of student participation in class etc. were deployed in a data mining model for predicting the performance of 908 students.

In [15] a decision tree model was applied to predict the probability of failure of 1,547 students such that relevant knowledge can be acquired that will enable the management team to be able to deploy adequate and early intervention. In the study, the student grades were classified into five categories, and these are: excellent, very good, good, acceptable and fail. Ten input features that include the student's department, high school grades, level of participation in class, attendance, midterm scores, lab reports, homework grades, seminar score, completion of assignments and the overall grades were applied in the decision tree model developed

In [16] by using decision tree classifiers, the likelihood of a student to drop out of an institution was predicted through educational data mining.

In [17], association, classification, clustering and outlier detection data mining techniques were applied to analyse 3,314 graduate student performance records over a fifteen-year period. The dataset was analysed using Rule Induction, Naïve Bayesian classifier, K-Means clustering algorithm followed by density-based and distance-based outlier detection methods. 18 attributes of the student dataset were considered, and only 6 attributes: matriculation GPA, gender, specialty of the students, the city of the student, the grade and the type of secondary school attended were selected for the data mining analysis. The remaining 12 attributes were dropped due to their large variances and because some of the attributes are personal information that did not provide useful knowledge.

In [18] The unsupervised clustering analysis performed, identified four unique clusters in the dataset using k-means algorithm. Data mining method was applied by to evaluate student data towards identifying the key attributes that influence the academic performance of students. This provides an opportunity for improving the quality of higher education.

In [19], data mining technique was Applied to analyse student data at a Bulgarian university. The student dataset that was analysed, contained the personal and pre-admission attributes of each student. The Decision Tree Classifiers (J48), k-Nearest Neighbour, Bayesian, Naïve Bayes classifiers, the OneR, and the JRip Rule learners were applied to extract knowledge from the student dataset, and accuracy of 52e67% was achieved. The result showed that the number of courses failed in the first academic year and the admission score of the student are two major features among the very influential features in the classification analysis.

In [20] the authors used WEKA data mining software for the prediction of final student mark based on parameters in two different datasets. Each dataset contains information about different students from one college course in the past fourth semesters. The IBK shows the best accuracy among other classifiers

In [21] the authors represents a study that will be helped to the students and the teachers to improve the result of the students who are at the risk of failure. Information's like Attendance, Seminar and assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. The authors used Naïve Bayes classification algorithm that shows a highest accuracy compared to other classification algorithms.

The researchers in [22] conducted a comparative research to test multiple decision tree algorithms on an educational dataset to classify the educational performance of students. The study mainly focuses on selecting the best decision tree algorithm from among mostly used decision tree algorithms, and provides a benchmark to each one of them and found out that the Classification and Regression Tree method worked better on the tested dataset, which was selected based on the produced accuracy and precision using 10-fold cross validations

Researchers in [23] provided an overview on the data mining techniques that have been used to predict students' performance and also it focused on how the prediction algorithm can be used to identify the most important attributes in a student's data. Under the classification techniques, Neural Network and Decision Tree are the two methods highly used by the researchers for predicting students' performance.

In [24], predictive analysis was carried out to determine the extent to which the fifth year and final Cumulative Grade Point Average (CGPA) of engineering students in a Nigerian University can be determined using the program of study, the year of entry and the Grade Point Average (GPA) for the first three years of study as inputs into a Konstanz Information Miner (KNIME) based data mining model. Six data mining algorithms were considered, and a maximum accuracy of 89.15% was achieved. The result was verified using both linear and pure quadratic regression models, and R2 values of 0.955 and 0.957 were recorded for both cases. This creates an opportunity for identifying students that may graduate with poor results or may not graduate at all, so that early intervention may be deployed.

In [25] analyze and evaluate the university students' performance by applying different data mining classification techniques by using WEKA tool. The highest accuracy of classifier algorithms depends on the size and nature of the data. Five classifiers are used NaiveBayes, Bayesian Network, ID3, J48 and Neural Network Different performance measures are used to compare the results between these classifiers. The results show that Bayesian Network classifier has the highest accuracy among the other classifiers.

In [26] used J48, PART, Random Forest and Bayes classifiers to predict students‟ end semester grades on a data set of 300 students from different colleges and found that Random Forest classification algorithm gives the best results based on accuracy and classifier errors .

In [27] used Bayesian Network, J48, Naive Bayes, ID3, J48 and Neural Network classifiers to analyse and evaluate students‟ performance grades and found that Bayesian Network classifier has the highest accuracy among all classifiers. They concluded that the performance of the students of a university can be best classified using Bayesian Network classification methods. The dependency among random variables is depicted by using directed acyclic graphs, where the nodes in the graph represent the random variables. The dependency of random variables is depicted when a connection exists between a node and an arc .

- After studying the previous approaches, we developed an evaluation of these approaches and reviewed the most important recent works with regard to forecasting the use of machine learning algorithms and techniques for combining classifications, and we reached the following:
- P. Mahato's approach is one of the best available approaches because it provides a set of comparisons to the results of testing individual classifications and the method of merging the bagging on more than one database. Data used in training and testing
- S. Kotsiantis et al's approach is characterized by the fact that it uses a merger method based on majority voting technology, but did not consider improving the database before applying the classifications to it.
- The P.Sittidech ET approach uses Under-sampling technology to process the unbalanced database and therefore this process can drop a set of majority row situations that may contain important information for the classification process.

D. Experiments and Results

In this section, a relative report on the technique of extracting data from the classification and predicting the Students performance will be applied based on the Grade Point Average, of first three year of student, and then predict student performance based on first and second and third year before they reach the four year and choice the best classifier dependent on Accuracy, Error rate, F - measure, exactness and review. Precision.

1. Evaluation Metrics

1.1 Precision

Precision is a ratio of true positive tuples and all positive tuples in a dataset .

1.2 Recall

Recall is a ratio of true positive tuples against positive and negative tuples

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 7, July 2020
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

1.3 F-Measure

F - Measure is also called as F - Score. F - Measure is a mean of precision and recall. F- Measure value varies from 0 to 1. If the value of F-Measure is higher, then it is said to be a better classifier .

1.4 Accuracy

The classifiers accuracy is an important metric for evaluation. It is a ratio of positive tuples and negative tuples against all the tuples.

1.5 Error Rate

The error rate is an essential measure for evaluation. Lower error rate is said to be a better classifier. Error rate determines the error between the prediction and actual

2. Data set:

The data set name is full data .CSV this consist of the following 4 feature.[first year GPA, second year GPA ,third year GPA,class of degree] From University of Science & Technology Faculty of Engineering with 1841 instance see figure 4.1 Implementation algorithms of classification by using WEKA tools .

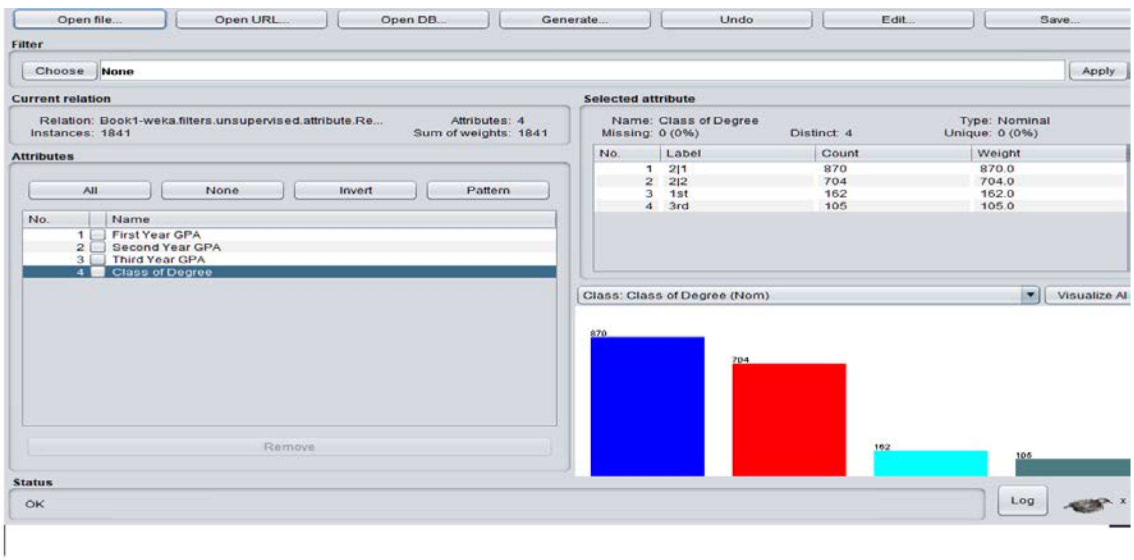

Figure.1.data set

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 7, July 2020
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

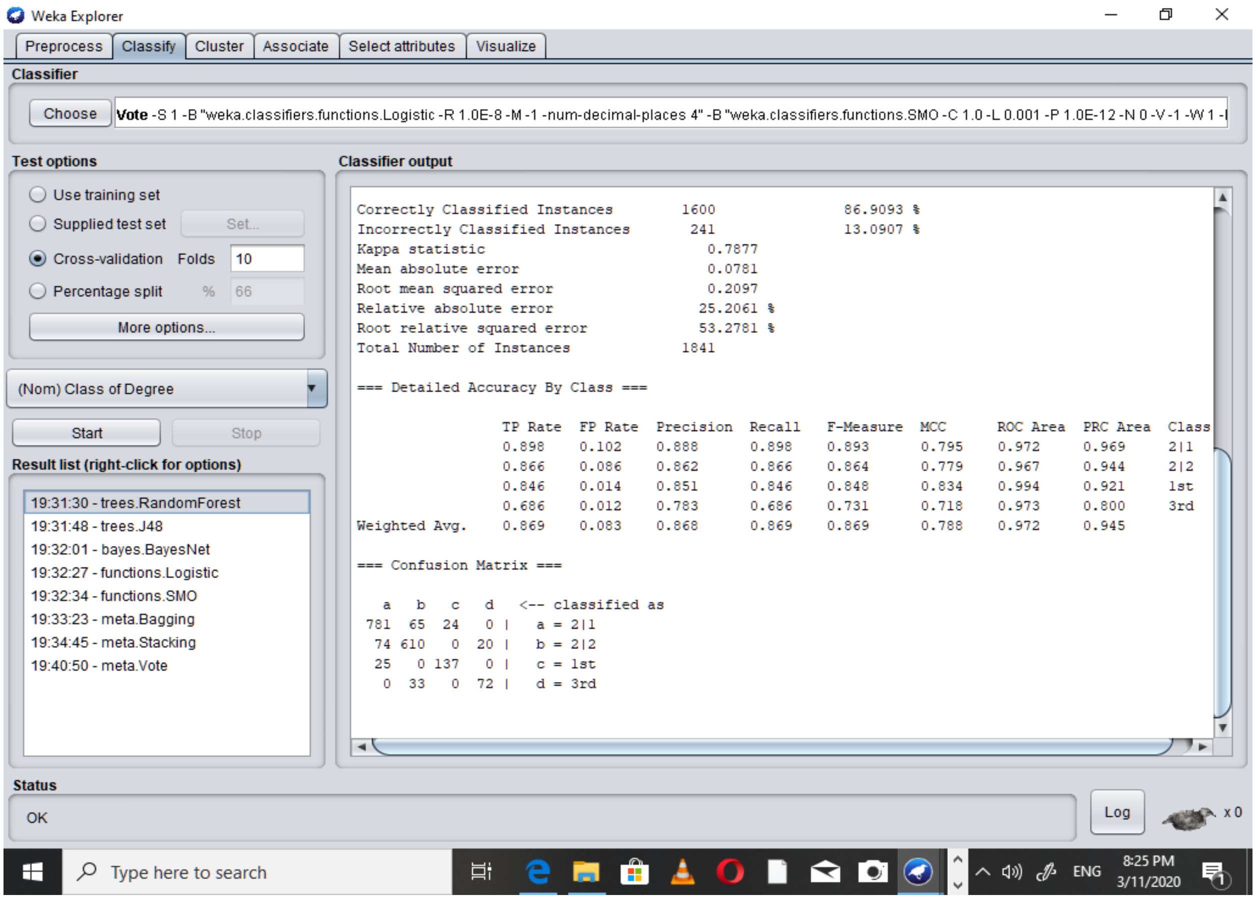

Figure.2.Result of classification model using randomforst algorithm

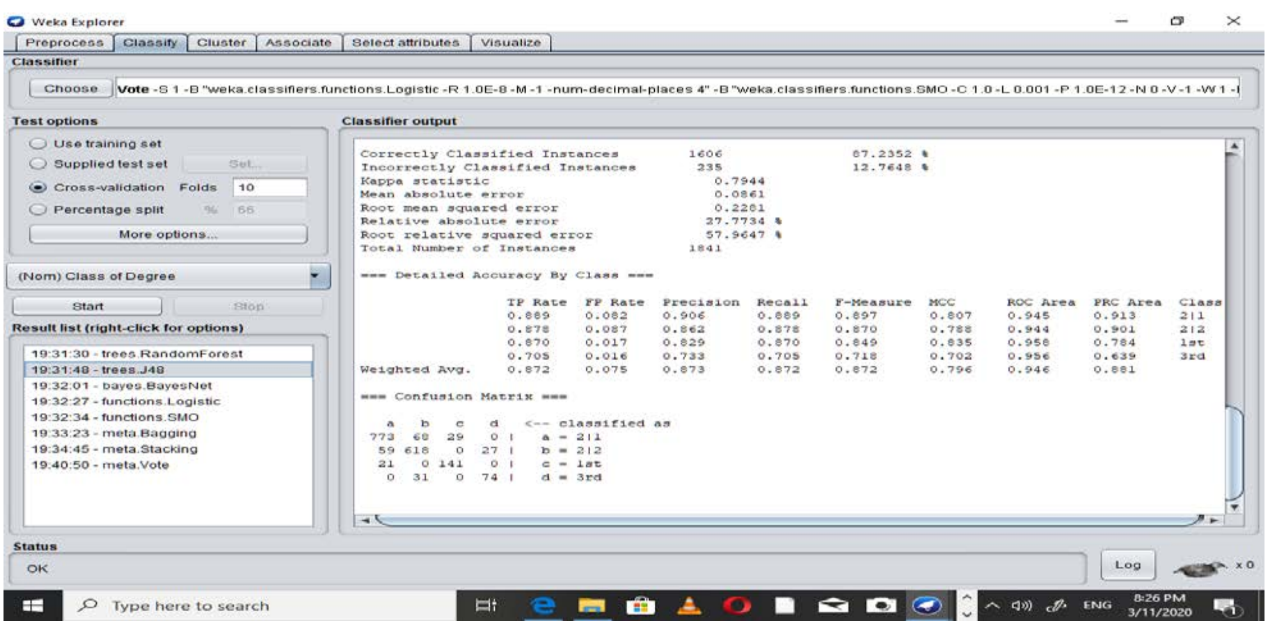

Figure.3.Result of classification model tree j48 algorithm

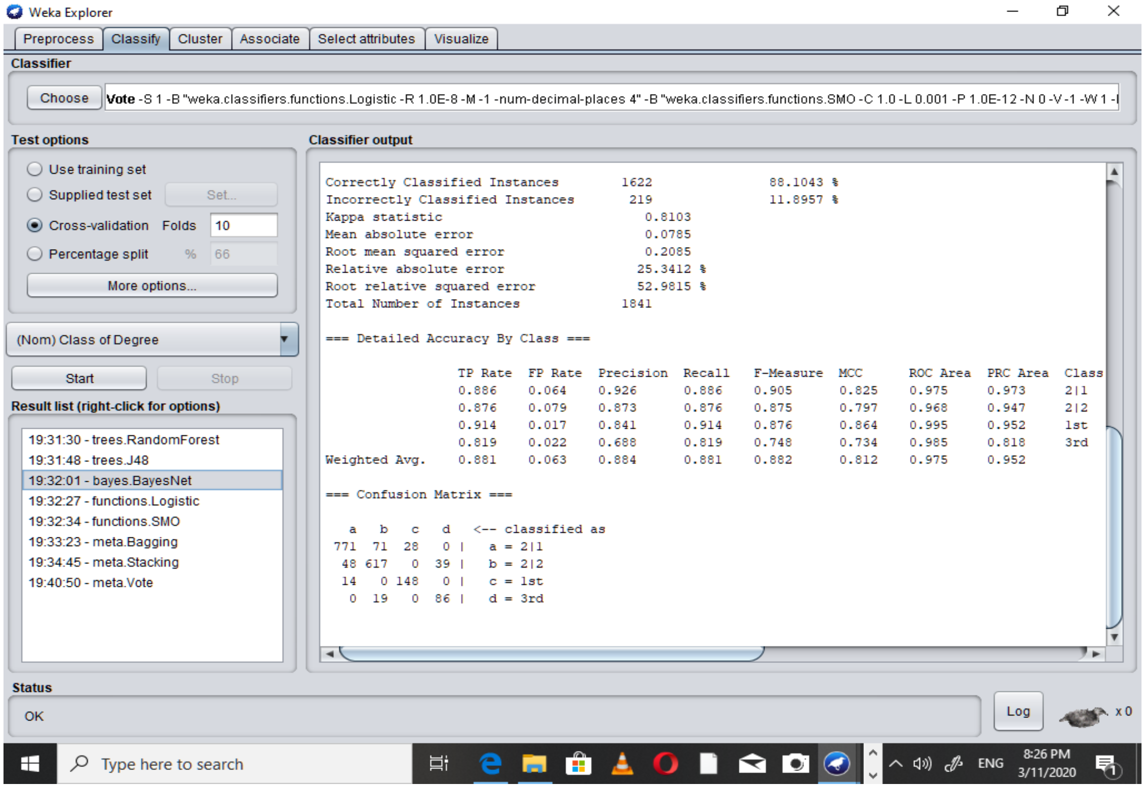

Figure.4.Result of classification model using Naïve Bayes algorithm

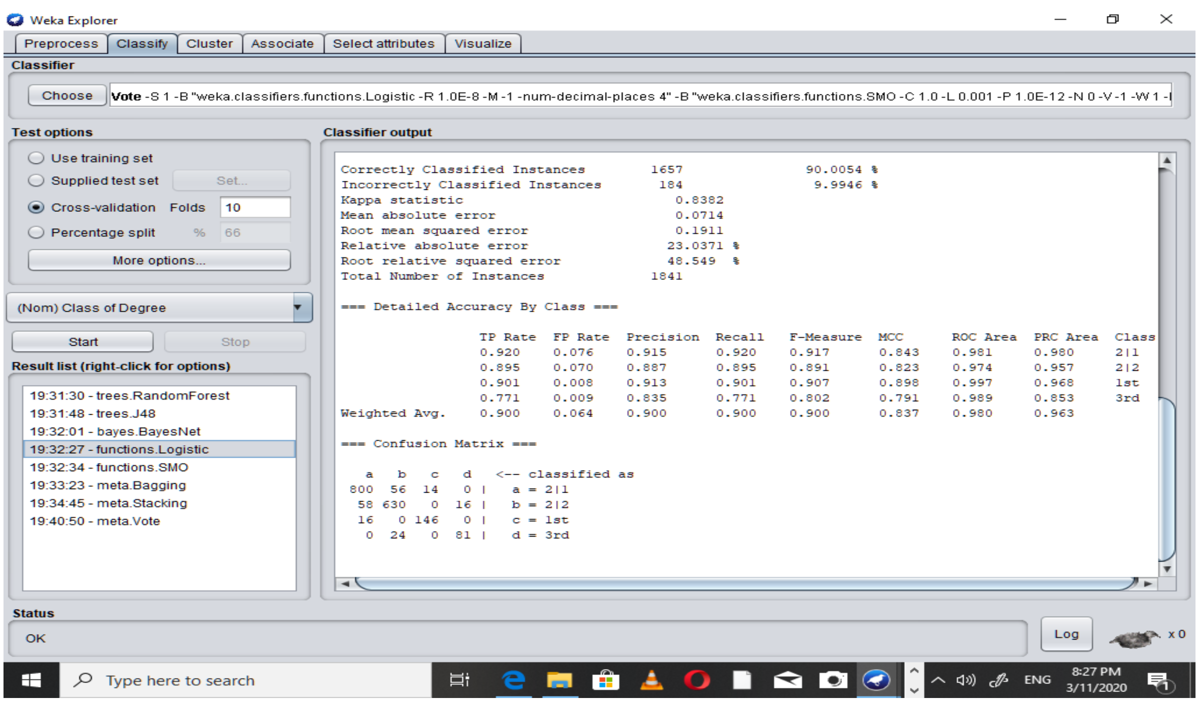

Figure.5.classification model using logistic algorithm.

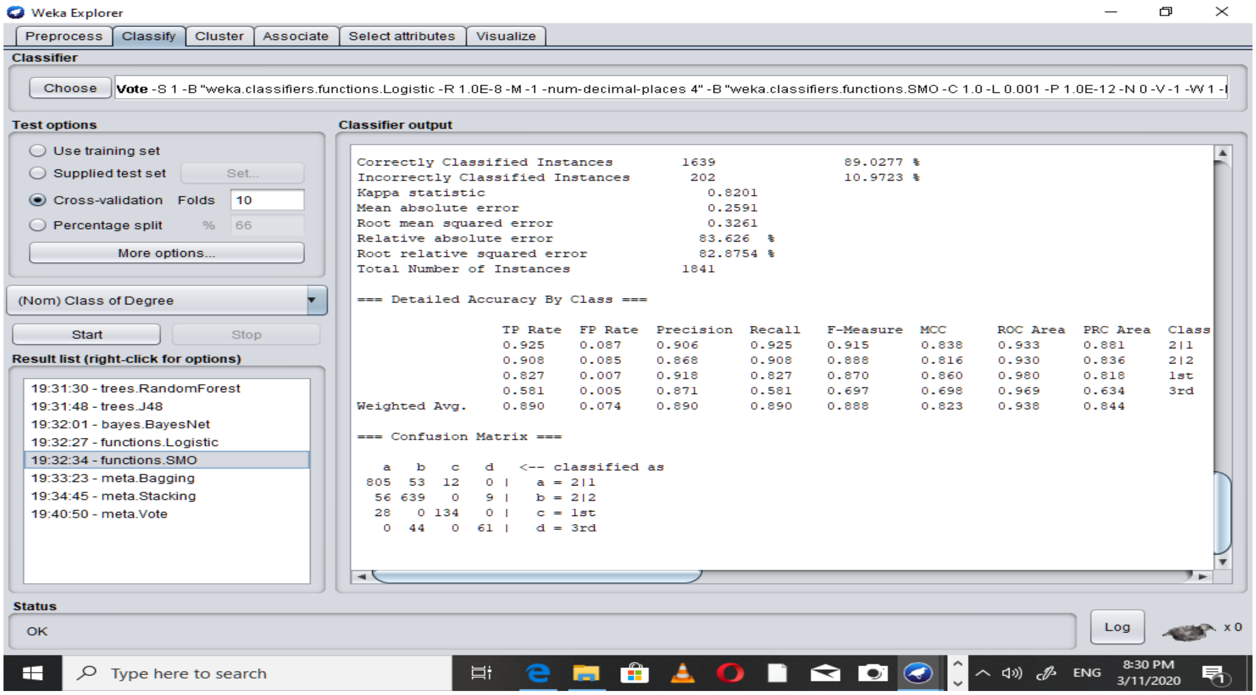

Figure.6.classification model using SVM algorithm.

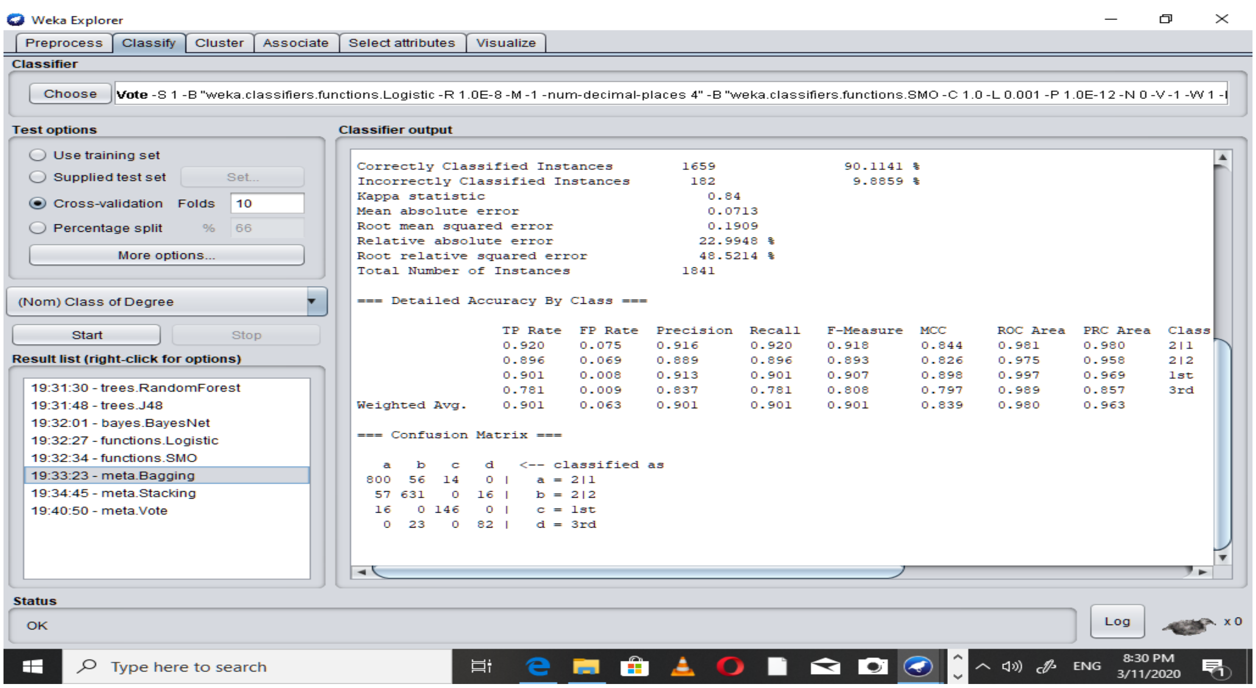

Figure.7.classification model using Bagging algorithm.

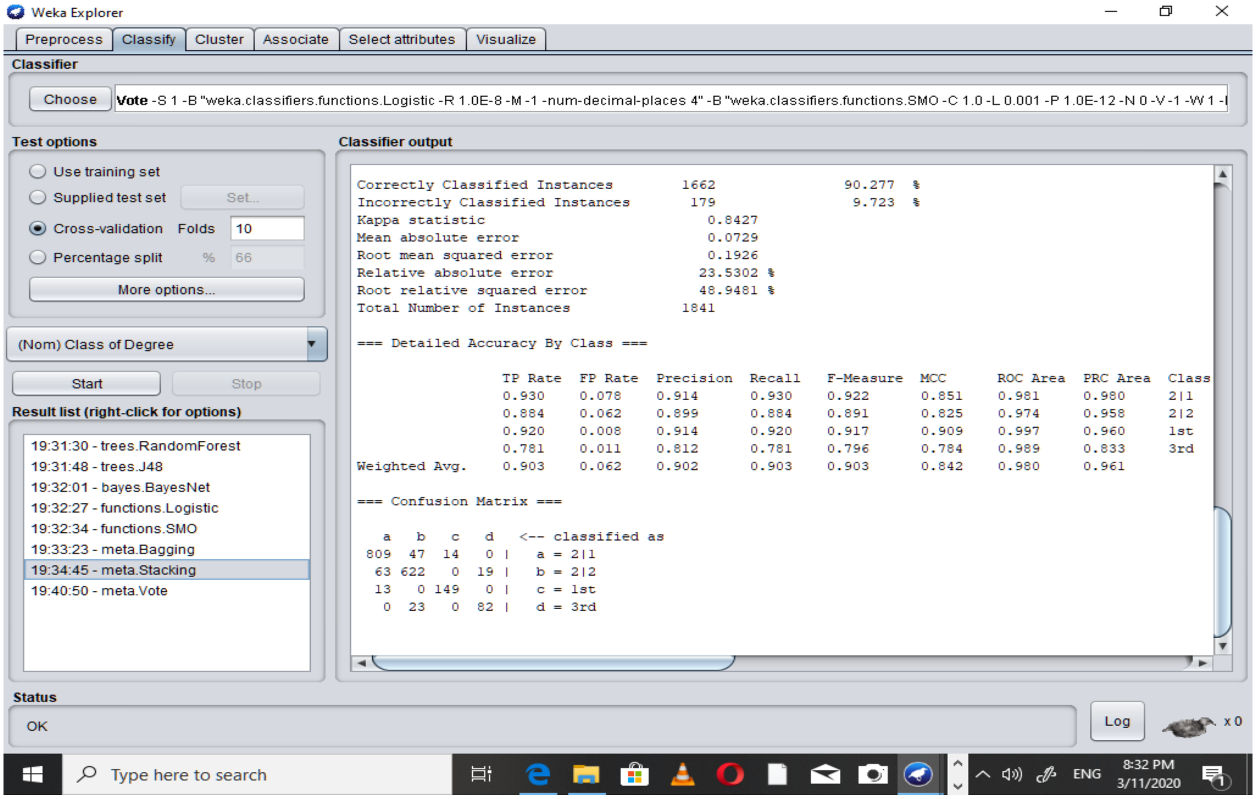

Figure.8.classification model using Stacking algorithm.

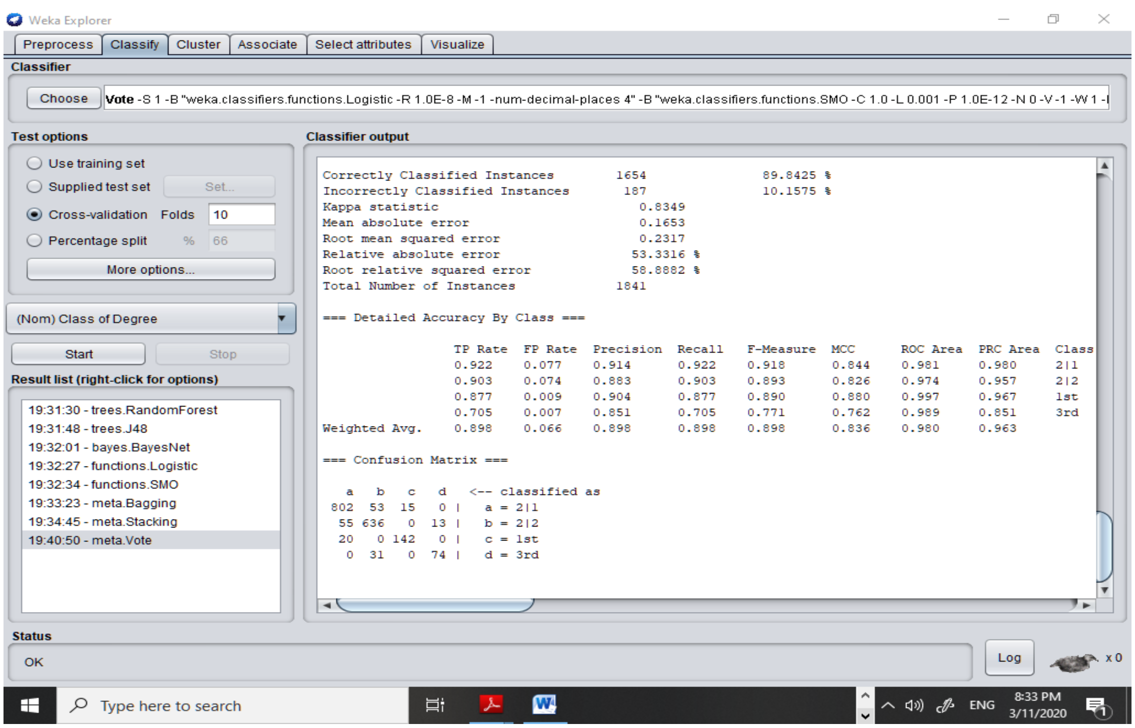

Figure.9.classification model using vote algorithm.

RESULTS DISCUSSION

The experimental results discussion has done on selecting 1841 instance. Five selected were executed individual. To choice the best tow classifiers with highest accuracy Different performance measures are used to compare the results between these classifiers algorithms and and combine to gather and compare with best accuracy in individual result in individual were used; Randomforst, J48, Naive Bayes, SVM, J48 and logistic and each one has its own characteristics to classify the data set. Table 1 shows performance results of individual classifiers by using WEKA, and Figure 10 shows the accuracy performance of individual classification techniques and Table 2 shows Error measures in weak of individual Classifiers and Table 4 shows the accuracy performance in Combining and best individual Classifiers

Table1 Comparison for Accuracy of individual Classifiers

Criteria	classifier				
	Randomforst	(J48)	Naïve Bays	SVM	logistic
Correctly classified instance	1600	1606	1622	1639	1657
Incorrectly classified instance	241	235	219	202	184
Accuracy (%)	% 86.9	% 87.2	% 88.1	% 89	% 90.0054

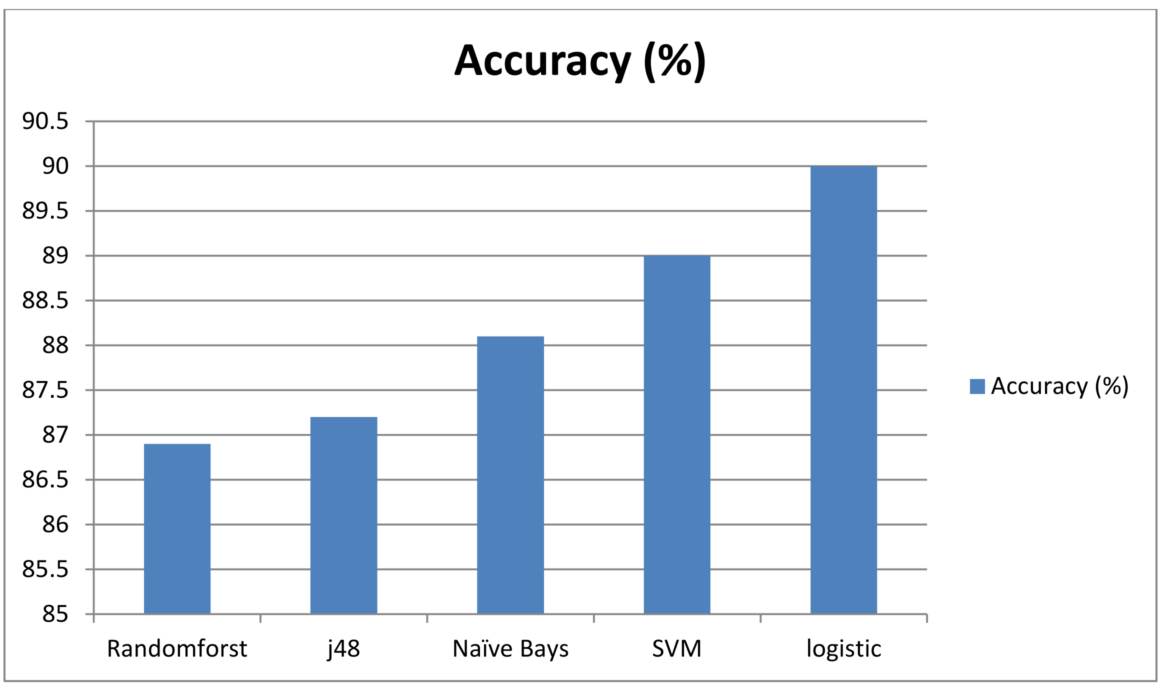

Figure.10. individual Classifiers Accuracy Performance

In table 1, the logistic regression and svm classifier has more correctly classified instances than other classifiers, which is usually referred to the best accuracy model. The graphical representation in Figure 7 shows that the best classifier of students' performance based on their dataset is the logistic regression and svm classifiers. In the result, logistic regression and svm has an efficient classification among other classifiers.

Table 2 Error measures in weak of individual Classifiers

Criteria	classifier				
	Randomforst	**(J48)**	**Naïve Bays**	**SVM**	**logistic**
Kappa statistic	0.7877	0.7944	0.8103	0.820	0.8382
Mean absolute error	0.0781	0.0861	0.0785	0.259	0.0714
Root mean squared error (RMSE)	0.2097	0.2281	0.2085	0.326	0.1911
Relative absolute error (RRSE)	25.2061 %	27.7734 %	25.3412 %	83.62 %	23.0371 %
Root relative squared error	53.2781 %	57.9647 %	52.9815 %	82.875%	48.549 %

In Table 2. Represents the error measures of individual classifiers, it shows that logistic regression and svm has a minimum error among other classifiers.

Table3 Weighted average of class label accuracy

classifier	**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
Randomforst	0.869	0.083	0.868	0.869	0.869	0.788	0.972	0.945	Weighted Age 2|1 2|2 1st 3rd
(J48)	0.872	0.075	0.873	0.872	0.872	0.796	0.946	0.881	
Naïve Bays	0.881	0.063	0.884	0.881	0.882	0.812	0.975	0.952	
SVM	0.890	0.074	0.890	0.890	0.888	0.823	0.938	0.844	
logistic	0.900	0.064	0.900	0.900	0.900	0.837	0.980	0.963	

Table 3 shows the performance accuracy of the five classifiers based on different classification metrics. These metrics are; (TP), (FP), Precision, Recall and F-measure measure are very important to determine the classifiers based on the accuracy. These metrics shows that logistic regression and avm classifier performs better than other classifiers.

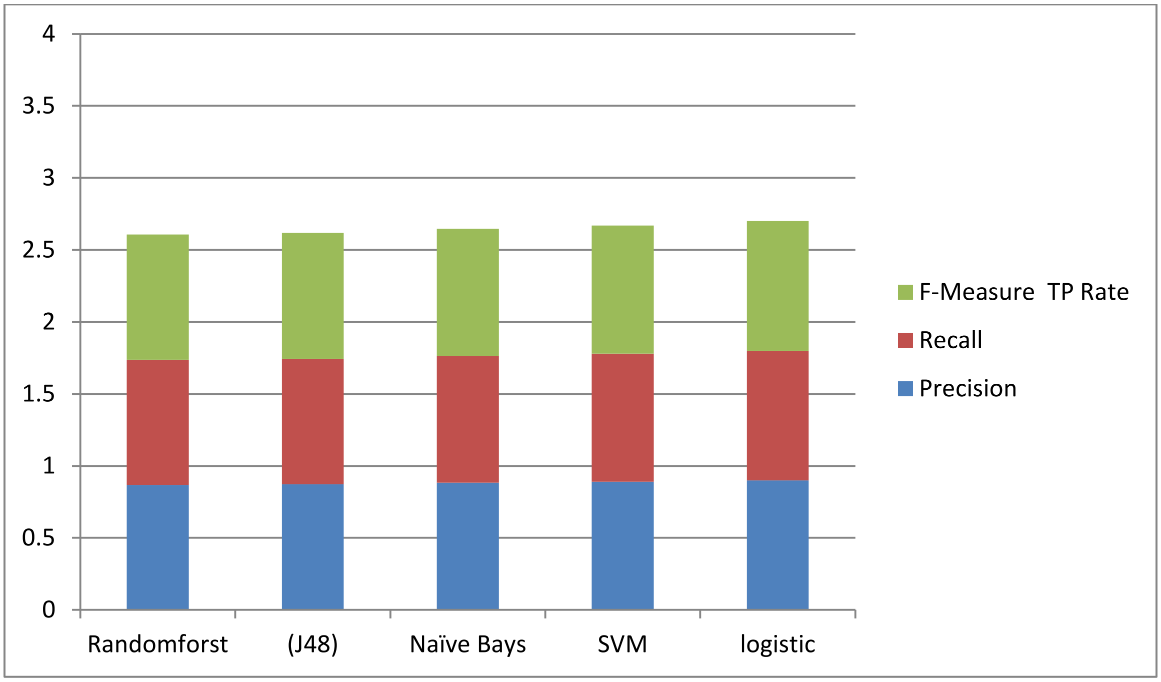

Figure.11. individual Classifiers Performance Metrics

In Figure 11 Precision, Recall and F-measures analysed among individual classifiers. It shows that the weighted average of logistic regression outperforms other classifiers.

Table4 Comparison for Accuracy of all Classifiers

Criteria	classifier			
	Logistic individual	vote	Bagging	stacking
Correctly classified instance	1657	1654	1659	1662
Incorrectly classified instance	184	187	182	179
Accuracy (%)	% 90.0054	%89.8425	%90.1141	%90.277

From above table we can conclude the performance of classifier with combination algorithm was the best in the comparison between performances of classifier without combination algorithm.

According to results, we have from this experiment, After comparing results obtained using boosting (stack) method for combination classifier and those reach without combination algorithm, the class of grade of a student's final, fifth-year graduation result can be reasonably predicted using the student's GPA for the first three years of study. The algorithm by using combination system chive high performance, show accuracy 90.277%, while the algorithm without combination (individual) chive low performance show the accuracy 90.0054%% in Logistic Regression algorithm,

Conclusion

An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples. One of the most active areas of research in supervised learning has been to study methods for constructing good ensembles of classifiers. The main discovery is that ensembles are often much more accurate than the individual classifiers that make them up. The main reason is that many learning algorithms apply local optimization techniques, which may get stuck in local optima

In this study, data mining approach was applied to evaluate the validity of this assumption by performing a predictive analysis to determine the final graduation CGPA and the class of grades of students in their final year using their GPA for the first three years of study. The program and the year of entry were applied as predictive inputs into a WEKA tool using five independent data mining algorithms that were executed individual and using a combined model for a comparative performance analysis of the result of each of the combined two algorithms. A maximum accuracy of 90.277% was achieved, using combination of logistic regression and svm for performance validation. This indicates that indeed the graduating results of engineering students in University, in the fifth and final year of study can be reasonably predicted using their performance in the first three academic sessions.

Finally, there are some open problems in ensemble of classifiers, such as how to understand and interpret the decision made by an ensemble of classifiers because an ensemble provides little insight into how it makes its decision. For learning tasks such as data mining applications where comprehensibility is crucial

References

1. Sagardeep Roy, Anchal Garg "Analyzing Performance of Students by Using Data Mining Techniques- A Literature Survey" 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)GLA University, Mathura, Oct 26-28, 2017.
2. https://www.analyticsindiamag.com/7-types-classification-algorithms.
3. https://www.investopedia.com/terms/n/neuralnetwork.asp.

4. Gaganjot Kaur, Amit Chhabra " Improved J48 Classification Algorithm for the Prediction of Diabetes" International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014.
5. Dietterich, T.G., Ensemble methods in machine learning. In Kittler, J.,Roli, F., eds.: Multiple Classifier Systems. Lecture Notes Computer Sciences, Vol. 1857, 2001, pp. 1-15.
6. Breiman L., Bagging Predictors. Machine Learning, Vol. 24, No. 3,1996, pp. 123-140.
7. Freund Y. and Robert E. Schapire. Experiments with a New Boosting Algorithm, Proceedings of ICML'96, pp. 148-156.
8. Webb G. I., MultiBoosting: A Technique for Combining Boosting and Wagging, Machine Learning, Vol. 40, 2000, pp. 159–196.
9. Melville P., Mooney R., Constructing Diverse Classifier Ensembles using Artificial Training Examples, Proceedings of IJCAI-2003, pp.505-510, Acapulco, Mexico.
10. S. Kotsiantis, K. Patriarcheas, M. Xenos, A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowledge-Based Systems 23 (2010).
11. Punnee Sittidech, Nipaporn Chanamarn, and Kanokwan Arunrudchadarom, Birth Asphyxia Classification Using AdaBoost Ensemble Method**.** Journal of Medical and Bioengineering (2015).
12. Pradeep Kumar Mahato, Prediction of Stock Price Movement Using Various Ensemble Models, College of Engineering, Pune (2014).
13. Oguntunde, Pelumi, Okagbue, Hilary, Oguntunde, Mooneye Adedoyin, Opanuga, A., 2018. Analysis of the inter-relationship between students' first year results and their final graduating grades. Int. J. Adv. Appl. Sci. 5 (10), 1e6.
14. Bucos, M., Dr_agulescu, B., 2018. Predicting student success using data generated in traditional educational environments. [Article]. TEM J. 7 (3), 617e625.
15. Ahmed, A.B.E.D., Elaraby, I.S., 2014. Data mining: a prediction for student's performance using classification method. World J. Comp. Appl. Technol. 2 (2),43e47.
16. Yadav, S.K., Bharadwaj, B., Pal, S., 2012. Mining Education Data to Predict Student 's Retention: A Comparative Study arXiv preprint arXiv:1203.2987.
17. Tair, M.M.A., El-Halees, A.M., 2012. Mining educational data to improve students' performance: a case study. Int. J. Inf. Commun. Technol. Res. 2 (2), 140e146.
18. Al-Radaideh, Q., Al-Shawakfa, E., Al-Najjar, M., 2006. Mining student data using decision trees. In: Paper Presented at the the 2006 International Arab Conference on Information Technology. ACIT'2006.

19. Kabakchieva, Dorina, 2013. Predicting student performance by using data mining methods for classification. Cybern. Inf. Technol. 13 (1).
20. I. Milos, S. Petar, V. Mladen and A. Wejdan, Students' success prediction using Weka tool, INFOTEH-JAHORINA Vol. 15, March 2016. 684.
21. P. Shruthi, B. Chaitra, Student Performance Prediction in Education Sector Using Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016.
22. 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.
23. A.Mohamed Shahiria,, W. Husaina , N. Abdul Rashida, "A Review on Predicting Student's Performance using Data Mining Techniques" Procedia Computer Science 72 (2015) 414 – 422, ELSEVIER.
24. Aderibigbe Israel Adekitan , Odunayo Salau , The impact of engineering students' performance
25. Hilal Almarabeh و I.J. Modern Education and Computer Science, 2017, 8, 9-15 Published Online August 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2017.08.02
26. Hussain S, Dahan N.A, Ba-Alwi F.M, Ribata N., " Educational Data Mining and Analysis of Students Academic Performance Using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 9, Issue 2, pp. 447-459, 2018
27. Almarabeh, H., "Analysis of Students‟ Performance by Using Different Data Mining Classifiers" I.J. Modern Education and Computer Science, Vol. 9, Issue 8, pp. 9-15, 2017