

Student academic performance prediction using improved classification Technique

Kolin Sukhadia^{*1}, Dr. Ayman Alahmar²

^{*1}Graduate Student, Department of Electrical and Computer Engineering, Lakehead University, Thunder Bay, Ontario, Canada

²Assistant Professor, Department of Software Engineering, Lakehead University, Thunder Bay, Ontario, Canada

Abstract :

The fundamental Objective of student academic performance prediction is to enhance the processes of education and learning. Research objectives, such as gaining a deeper understanding of the teaching and learning phenomena, identifying weaker students at an early stage, recommending them extra Module/Course, identifying the factors affected to student success the most, and analysing student behaviour in e-learning system. Using prediction of students performance, we can identify the factors which need improvement to increase the university success rate as well as student success rate, which is a major objective of any educational institutions.

1. Introduction :

Educational data mining is a feature of data mining where the main concern is to construct models for extracting hidden knowledge from student's data which may help to improve student's academic performance. The fundamental Objective of student academic performance prediction is to enhance the processes of education and learning. Research objectives, such as gaining a deeper understanding of the teaching and learning phenomena, identifying weaker students at an early stage, recommending them extra Module/Course, identifying the factors affected to student success the most, and analysing student behaviour in e-learning system. Using prediction of students performance, we can identify the factors which need improvement to increase the university success rate as well as student success rate, which is a major objective of any educational institutions. This knowledge helps educational institutes to improve their teaching quality. Different attributes have an impact on student's academic performance. Behavioral and student absent in class has an effect on their academic performance.

To analyze data using classification technique, well known classification algorithms such as Decision tree (DT), Artificial neural networks (ANN), K-nearest neighbor (KNN), Bayes network, Naïve Bayes are being used for prediction purposes. By using classification technique based boosting technique help to improve classification accuracy and reduce errors. The major objective of the proposed methodology is

to build the ensemble classification model that classifies a students' performance as high, medium and low. Student dataset containing columns like studentid, raisehands, visitedresources, announcementsview, studentabsentdays, parentschoolsatisfaction, parentansweringsurvey, discussion, class . class are categorized into high, low and medium students as per students performance.

Nowdays there are many colleges and universities it is almost increased compare to previous years. At peak enrollment, 29.5 million students attended degree-granting institutions. According to college enrollment trend, for virginia state it is increased from 152K to 325K college enrollment from 1970 to 2000. For Alabama state, in 2000, 234,000 students were enrolled in college in Alabama. In 1970, 104,000 were enrolled. So by doing accurate prediction on large student dataset, it is very easy to identify students as high, medium and low. According to research paper "Predicting University Students' Academic Success Using Different Tree Classifiers And Ensemble Approaches To Suggest Suitable Program", other various classification techniques help to increase accuracy of student academic performance prediction. one of these technique show that classification accuracy is increased from 68.97% to 70.24%. According to "Student Academic Performance Prediction using Supervised Learning Techniques, iJET – Vol. 14, No. 14, 2019", Four classifiers oneR, MLP, J48, and IB1 were applied on the student data set these all classifiers give classification accuracy like 95.8%, 92.8%, and 93.4%. so proposed methodology of classification method based boosting technique helps in improving classification accuracy to approximately 98% to 99%. And reducing errors like RMSE, RAE and MAE One of the biggest challenges is to improve the quality of the educational processes so as to enhance student's performance. Instructors can update their teaching methodology to fulfill the requirement of poor performance students and can provide additional guidance to deserving students. The prediction results might help students develop a good understanding of how well or bad they would perform in a course and then can take steps accordingly. The further document is as follows : In section 2, there is methodology used in research, section 3 related work is there. Results and Discussion is there in section 4. In section 5, there is Conclusion and future scope. Section 6 indicates references.

2.Materials and Methodology :

Classification method based ensemble method boosting technique. Proposed methodology is as below.

Boosting boosts the performance of the weak classifier to a strong level. It generates sequential learning classifiers using resampling(reweighting) the data instances. Initially equal uniform weights are assigned to all the instances.

During each learning phase a new hypothesis is learned and instances are reweighted such that correctly classified instance having lower weight and system can concentrate on instances that have not been correctly classified during this phase having higher weights. It selects the wrongly classified instance, so that they can be classified correctly during the next learning step. This process continuous till the last classifier construction. Finally the results of all the classifiers are combined using majority voting to find the final prediction.

Adaboost :

Consolidating with numerous other learning algorithms, the meta algorithm is called as AdaBoost algorithm. This would improve the execution of classification. AdaBoost utilizes the settled administrator and it has a sub procedure. The better model is generated by sub-processor. More than one classifier is created by model. and generates a better model. By creating more than one classifier, the accuracy of classification is increased. This model leads to decision making by combining the results of their classification techniques. The accuracy of the given algorithm is improved by boosting method. combining with numerous other algorithms, the meta algorithm is called as AdaBoost algorithm. This would improve the performance of classification. AdaBoost uses sequential process. The better model is generated by sub-processor. By creating more than one classifier, the accuracy of classification is improved by the ensemble model. This model leads to decision making by combining the results of their classification techniques.

J48 algorithm :

The every aspect of the information is to split into minor subsets to base on a decision. J48 look at the standardized data gain that really the results the split the information by choosing an attribute. To summarize, the attribute extreme standardized data gained is utilized. The minor subsets are returned by the algorithm. The split strategies stop if a subset has a place with a similar class in all the instances. J48 develops a decision node utilizing the expected estimations of the class. J48 decision tree can deal with particular characteristics, lost or missing attribute estimations of the data and varying attribute costs. Here accuracy can be expanded by pruning

The Algorithm

Stage 1: The leaf is labeled with a similar class if the instances belong to similar class.

Stage 2: For each attribute, the potential data will be figured and the gain in the data will be taken from the test on the attribute.

Stage 3: Finally the best attribute will be chosen depending upon the current selection parameter.

proposed method description :**Proposed method Steps :**

Step1: instances of dataset $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Each instance has equal probability $D_0(i) = 1/n$.

Step 2 : T learning rounds

Step3: apply classifier J48 to student academic performance dataset.

Step 4 : Learn base learner from given dataset

Step5:Structural Learning:Apply J48 learning technique

Step 6: Parameter Learning of J48.

Step7: Measures for network analysis.model validation. cross validation is a standard way to obtain estimates of models.

Step8: Et error < 0.5 if yes then continue else stop.

Step9: get weak classifier ht : alpha_t=(weight associated with ht).

Step10 : update distribution Dt+1(i) for those examples for which learner classify them wrongly.

Step11:Linear combination of these different hypothesis weighted by alpha_t.

Step12: output result of adaboost with J48.

Proposed method AdaBoost with bayesnet classifier algorithm

Given : (x1,y1),(x2,y2)....(xm,ym)

Initialize weights D₁(i)=1/m

For t=1....T

Train weak learner using distribution D_t

Apply J48 base learner,which return base classifier ht with low weighted error

J48 : Learn J48 structure from given dataset

Model validation.cross validation is a standard way to obtain estimates of models.

Calculate $\epsilon_t = \sum_{i=1}^m D_t(i) \delta(h(\mathbf{x}_i) \neq y_i)$

Set weight α_t based on the error

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t(h_t)}{\epsilon_t(h_t)} \right)$$

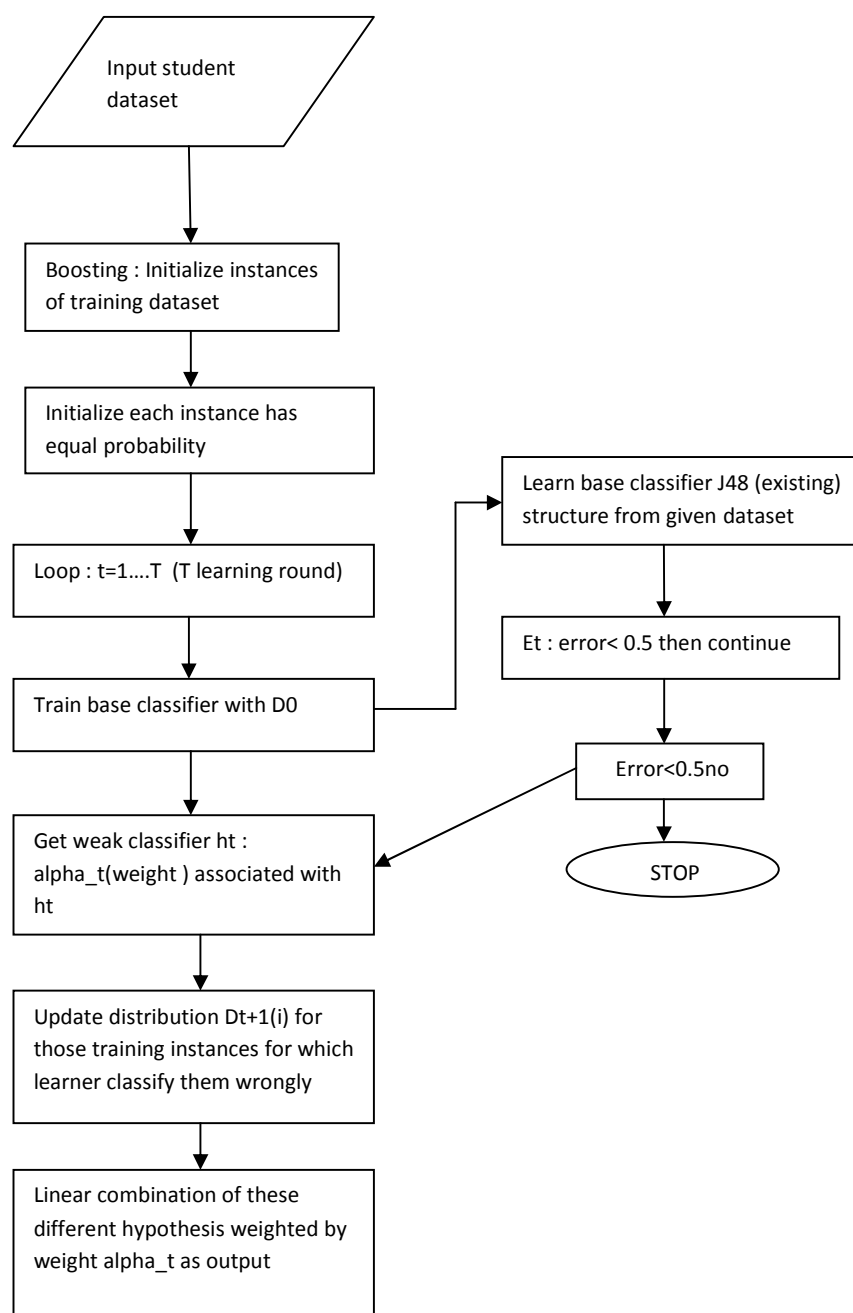
Update the distribution based on the performance

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Output the strong classifier :

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

Proposed method flowchart :



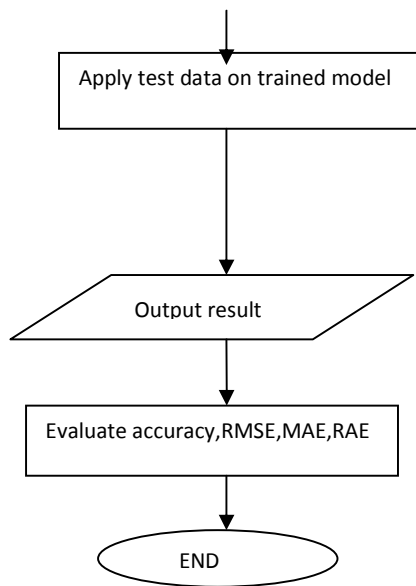


Figure : proposed classification method based boosting technique

3. Related work:

In[1], T.Thilagaraj, Dr.N Sengottaiyan does research on educational data mining. The excellence of educational institutions is viewed by the success rate of students and the skill set of the institutions will be measured by retained rate of students at risk. The different aspects like individual, social and psychological will be useful to measure student academic performance. This may lead to discover the students who are in risk and it help the management to take timely action. The student academic performance will be measured by their socio-economic and previous academic performances. This process will be performed by using educational data mining techniques. The determination of classes will be made before examining the data so it is also referred to as supervised learning. Based on the previous academic performances and Socio-economic situations the student performances were measured through Data mining techniques.

In [2], K.R.Kavyashree, LakshmiDurga has proposed that a country's growth is strongly measured by the quality of its education system. Education sector has witnessed sea change in its functioning. Today it is recognized as an industry and as an industry it is facing challenges. The challenges of higher education being decrease in students' success rate and their leaving the course without completion. An early prediction of students' failure may help the management provide timely counselling as well as coaching to increase the success rate and student retention. Data mining are widely used in educational field to find new hidden patterns from student's data which are used to understand the problem. Classification is one of the prediction type classifiers

that classifies data based on the training set and uses the pattern to classify a new data. Aim of the project is to develop an internetworking application that uses data mining technique to predict the students' performance based on their behaviour. This paper explored the link between emotional skills of the students along with the socio economic and previous academic performance parameters using Naive Bayes Classifier technique.

In research [3]Anjali B Raut and Ankita A Nichat (2017) present a student performance system which is based on the common classification technique, the Decision Tree. This system models an online examination, where students take up an exam and instant result (Pass/Fail) along with weak concepts are displayed. Generalization sequential pattern mining algorithm is used for result determination. The decision tree generated by C4.5 is used to analyze the performance of students and classify them based on their marks. The author's states that this analysis using data mining will help the management to identify weak students and provide extra coaching before the final exam.

In [4], Ali Daud, A.S.M Badrudduza (2017) develop a prediction model based on learning analytics which predicts student's academic performance. Attributes like family expenditure, family income and family assets were included in this research work. To evaluate metrics like Precision Rate, Recall and F1- score were used. Five fold cross validation was used. Classifiers used are Naïve Bayes, SVM, C4.5 and Decision Trees in this work. Data was collected from graduate and undergraduate courses from different universities of Pakistan. Results reveal that SVM classifier outperformed other classifiers. Some interesting revelation of this research work is as follows; family expenditure affects student's performance, married students perform better than bachelor students and house condition of students influences the performance of the students..

In [5],TriptiDwivedi, Diwakar Singh has proposed that the survey of student background history which helps academic planners in institute to give right direction to student. If the class of student is predicted in midsession of institute in final year then it will be easy for the academic planner to plan some important workshop for the enhancement of performance of student which helps it in placement at the end of academic session. In educational institute data mining techniques plays an important role in each activities of institute whether it is academic, cultural, examination and training and placement etc. in which Educational Data Mining which is a field of data mining helps a lot to find the actual filtered data in various field of department in institute. Hidden knowledge through data mining techniques is extracted from large database which helps to predict the pattern in such activities. It plays a great role in predictions of student data for placement and performance.

In [6], Md Rifatul Islam Rifat, Abdullah Al Imran, A. S. M. Badrudduza has done research on Educational Performance Analytics of Undergraduate Business Students. In most of the Asian countries, it is a challenging task to perform EDM due to the diverse characteristics of the educational data. In this study, they have performed proposed framework.they have also conducted extensive experiments on real world dataset that has been prepared by the transcript data of the students from the marketing department.

The researchers Kostopoulos et al., [7] proposed an ensemble of classification and regression algorithms for

course and showed that their proposed model performed better than the base models even if the conventional ensemble methods. They combined the REPTree algorithm.

4. Results and Discussion :

Student academic performance dataset :

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
Nationality	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhands	VisitedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentSchoolSatisfaction	StudentAbsencesDays	Class
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	15	16	2	20	Yes	Good	Under-7	M
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	20	20	3	25	Yes	Good	Under-7	M
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	10	7	0	30	No	Bad	Above-7	L
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	30	25	5	35	No	Bad	Above-7	L
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	40	50	12	50	No	Bad	Above-7	M
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	42	30	13	70	Yes	Bad	Above-7	M
KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father	35	12	0	17	No	Bad	Above-7	L
KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father	50	10	15	22	Yes	Good	Under-7	M
KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father	12	21	16	50	Yes	Good	Under-7	M
KW	Kuwait	MiddleSchool	G-07	B	IT	F	Father	70	80	25	70	Yes	Good	Under-7	M
KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father	50	88	30	80	Yes	Good	Under-7	H
KW	Kuwait	MiddleSchool	G-07	B	Math	F	Father	19	6	19	12	Yes	Good	Under-7	M
KW	Kuwait	lowerlevel	G-04	A	IT	F	Father	5	1	0	11	No	Bad	Above-7	L
lebanon	lebanon	MiddleSchool	G-08	A	Math	F	Father	20	14	12	19	No	Bad	Above-7	L
KW	Kuwait	MiddleSchool	G-08	A	Math	F	Mum	62	70	44	60	No	Bad	Above-7	H
KW	Kuwait	MiddleSchool	G-06	A	IT	F	Father	30	40	22	66	Yes	Good	Under-7	M
KW	Kuwait	MiddleSchool	G-07	B	IT	F	Father	36	30	20	80	No	Bad	Above-7	M
KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father	55	13	35	90	No	Bad	Above-7	M
KW	Kuwait	MiddleSchool	G-07	A	IT	F	Mum	69	15	36	96	Yes	Good	Under-7	M
KW	Kuwait	MiddleSchool	G-07	B	IT	F	Mum	70	50	40	99	Yes	Good	Under-7	H
KW	Kuwait	MiddleSchool	G-07	A	IT	F	Father	60	60	33	90	No	Bad	Above-7	M
KW	Kuwait	MiddleSchool	G-07	B	IT	F	Father	10	12	4	80	No	Bad	Under-7	M
KW	Kuwait	MiddleSchool	G-07	A	IT	F	Father	15	21	2	90	No	Bad	Under-7	M
KW	Kuwait	MiddleSchool	G-07	A	IT	F	Father	2	0	2	50	No	Bad	Above-7	L
KW	Kuwait	MiddleSchool	G-07	B	IT	F	Father	0	2	3	70	Yes	Good	Above-7	L

Student dataset contains various parameters like gender, nationality, place of birth, grade-id, section-id, semester, relation, Raisedhands, VisitedResources, AnnouncementsView, Discussion, ParentAnsweringSurvey, ParentSchoolSatisfaction, StudentAbsencesDays and Class. Class are classified as medium, low and high student.

Performance Evaluation Metrics :

Accuracy : It is characterized as the general achievement rate of the classifier.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

Where , True Positive (TP) portrays the quantity of examples where system recognizes for a condition when it is truly present.

True Negative(TN) delineates the quantity of instances where system does not identify a condition when it is absent.

False Negative(FN) portrays the quantity of instances where system does not recognize a condition when really it is present.

False Positive(FP) portrays the quantity of instances where system distinguish a condition when it is truly absent.

MAE(Mean Absolute Error) : it is factor in insights which estimates the distinction between two continuous variables.

RAE(Relative Absolute Error) : This measure gives the total absolute error between variables.

Recall : $\text{true positive} / (\text{true positive} + \text{false negatives})$

Where, true positives : correctly identified instances.

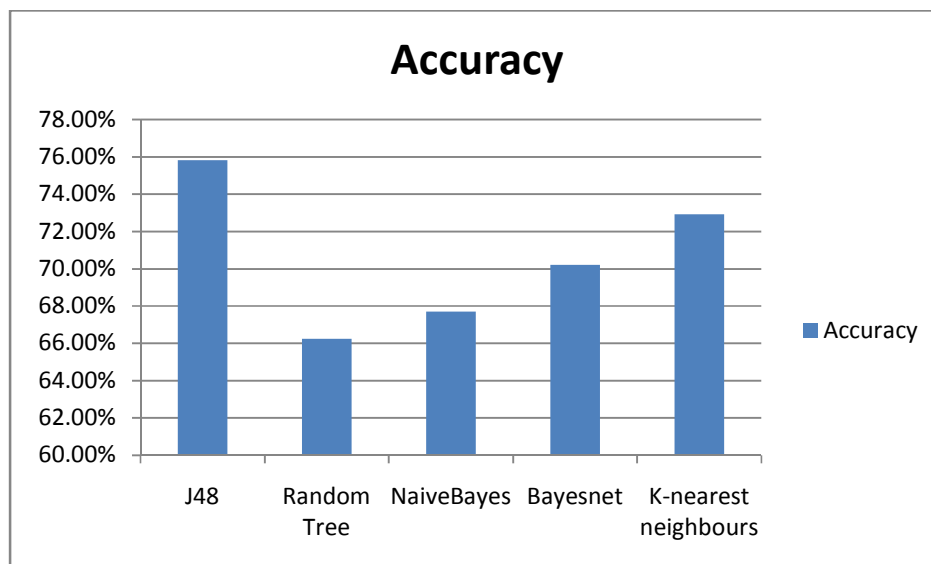
False negatives : Incorrectly classified instances.

Error Rate

Implementation :

Comparison of Accuracy of classification Techniques :

J48/Random Tree/NaiveBayes/Bayesnet/K-nearest Neighbours :



Accuracy for J48 : 75.8333%

Accuracy for Random tree : 66.25%

Accuracy for Naïve Bayes : 67.7083%

Accuracy for Bayesnet : 70.2083%

Accuracy for K – nearest neighbours : 72.9166%

	J48	Random Tree	Naïve Bayes	Byesnet	Knn
Accuracy	75.8333%	66.25%	67.7083%	70.2083%	72.9166%

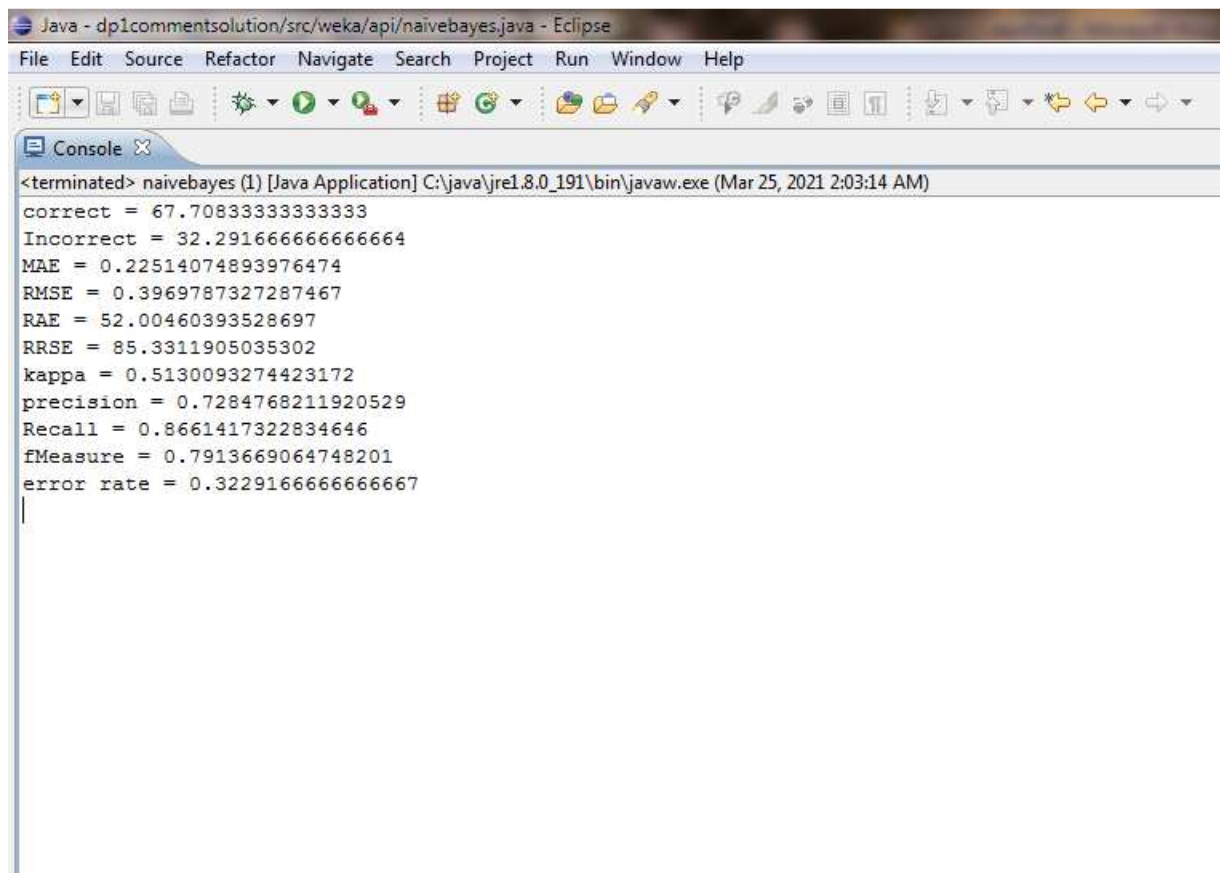
As above graph, I compare classification accuracy of prediction of Student academic performance. And improve classification accuracy by improved classification technique. In my research I choose J48 as my base classifier for my proposed technique to improve classification accuracy of student academic performance and reduce errors.

Classification Techniques output screenshot :

Randomtree output screenshot :

```
Java - dp1commentssolution/src/weka/api/randomtree.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
<terminated> randomtree [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 25, 2021 1:46:13 AM)
correct = 66.25
Incorrect = 33.75
MAE = 0.22544102161223492
RMSE = 0.46624231249425896
RAE = 52.07396304276946
RRSE = 100.21950373709048
kappa = 0.4813821805158166
precision = 0.7045454545454546
Recall = 0.7322834645669292
fMeasure = 0.7181467181467182|
error rate = 0.3375
```

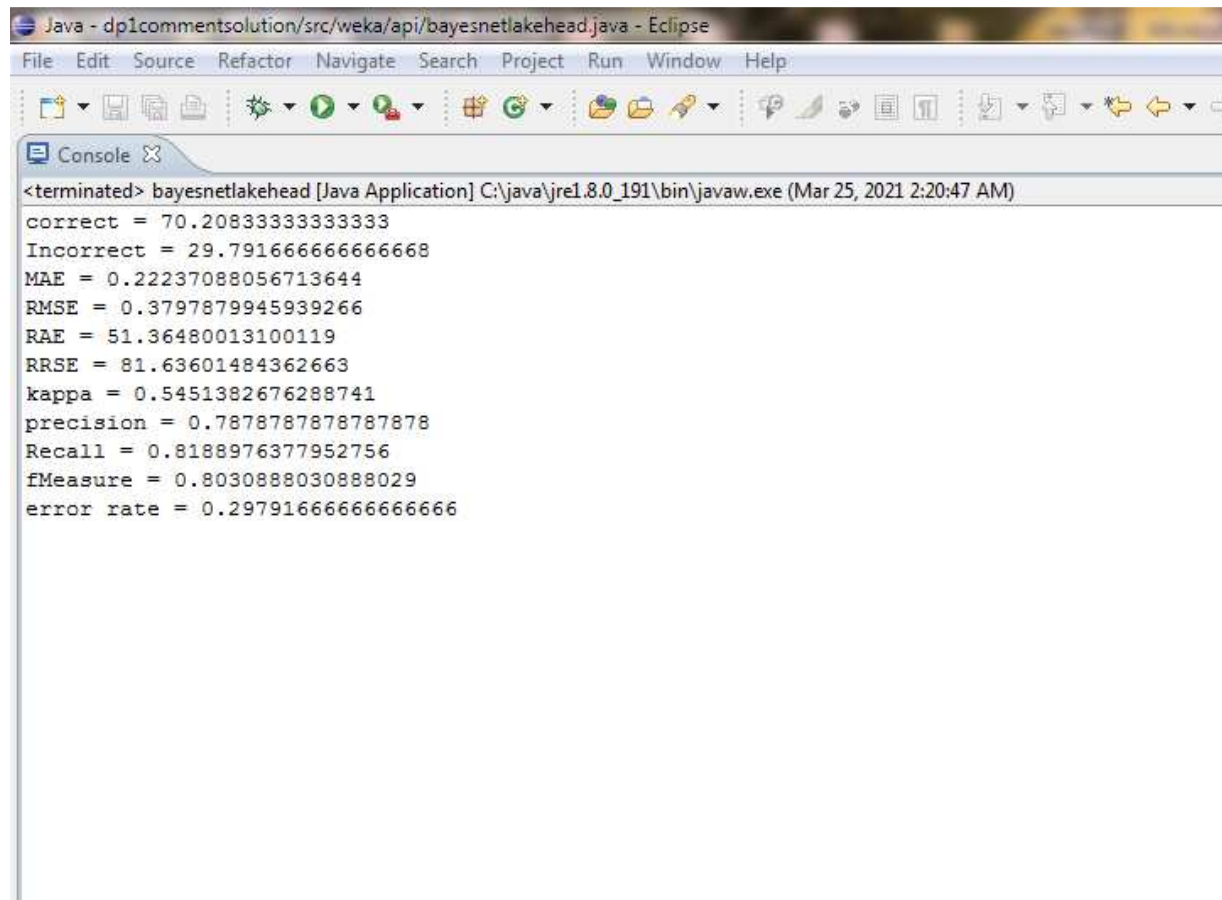
Naïve Bayes Output Screenshot :



The screenshot shows the Eclipse IDE interface with the console window open. The title bar of the IDE is 'Java - dp1commentsolution/src/weka/api/naivebayes.java - Eclipse'. The menu bar includes File, Edit, Source, Refactor, Navigate, Search, Project, Run, Window, and Help. The toolbar contains various icons for file operations and development tools. The console window displays the output of a Java application, indicating it has terminated. The output text is as follows:

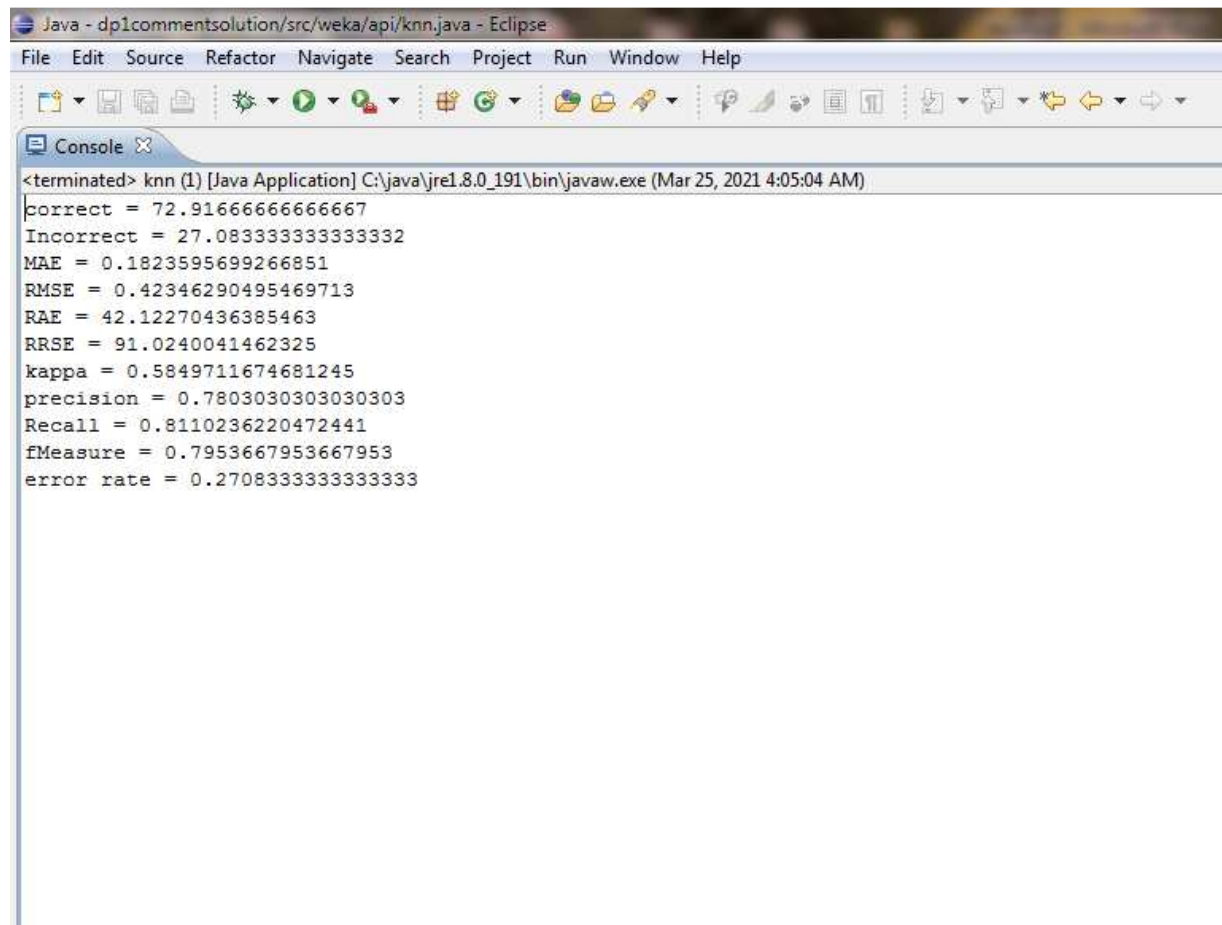
```
<terminated> naivebayes (1) [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 25, 2021 2:03:14 AM)
correct = 67.70833333333333
Incorrect = 32.291666666666664
MAE = 0.22514074893976474
RMSE = 0.3969787327287467
RAE = 52.00460393528697
RRSE = 85.3311905035302
kappa = 0.5130093274423172
precision = 0.7284768211920529
Recall = 0.8661417322834646
fMeasure = 0.7913669064748201
error rate = 0.3229166666666667
|
```

Bayesnet Screenshot output :



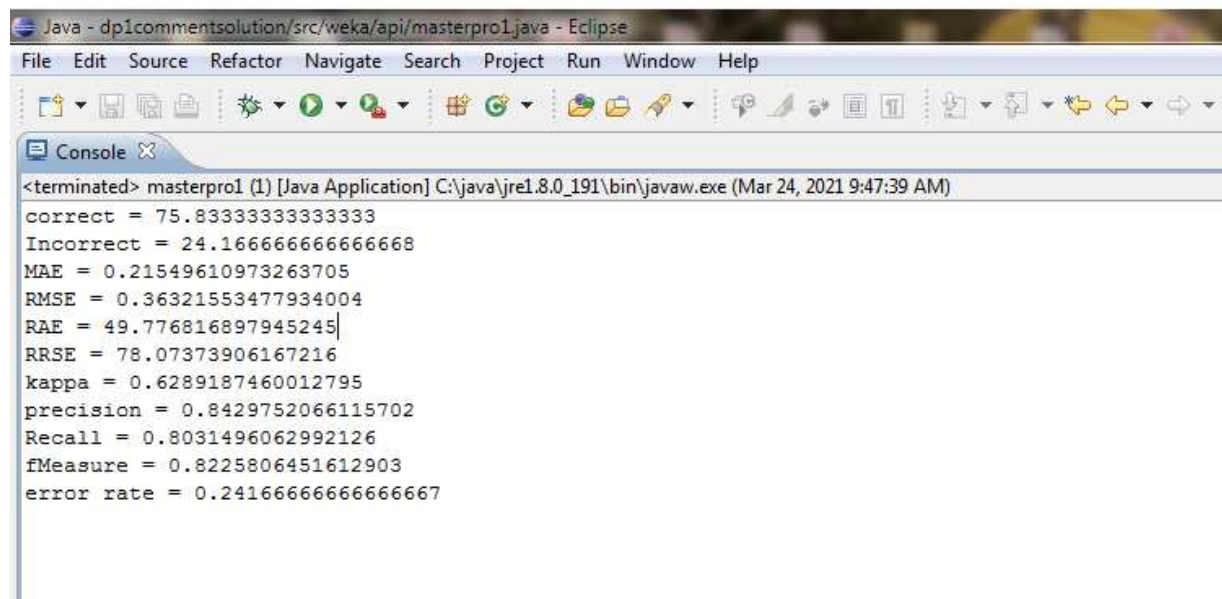
```
Java - dp1commentsolution/src/weka/api/bayesnetlakehead.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
<terminated> bayesnetlakehead [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 25, 2021 2:20:47 AM)
correct = 70.20833333333333
Incorrect = 29.791666666666668
MAE = 0.22237088056713644
RMSE = 0.3797879945939266
RAE = 51.36480013100119
RRSE = 81.63601484362663
kappa = 0.5451382676288741
precision = 0.7878787878787878
Recall = 0.8188976377952756
fMeasure = 0.8030888030888029
error rate = 0.29791666666666666
```

K-nearest neighbour screenshot output :



```
Java - dp1commentssolution/src/weka/api/knn.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
<terminated> knn (1) [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 25, 2021 4:05:04 AM)
correct = 72.91666666666667
Incorrect = 27.083333333333332
MAE = 0.1823595699266851
RMSE = 0.42346290495469713
RAE = 42.12270436385463
RRSE = 91.0240041462325
kappa = 0.5849711674681245
precision = 0.7803030303030303
Recall = 0.8110236220472441
fMeasure = 0.7953667953667953
error rate = 0.2708333333333333
```

Existing method output screenshot :



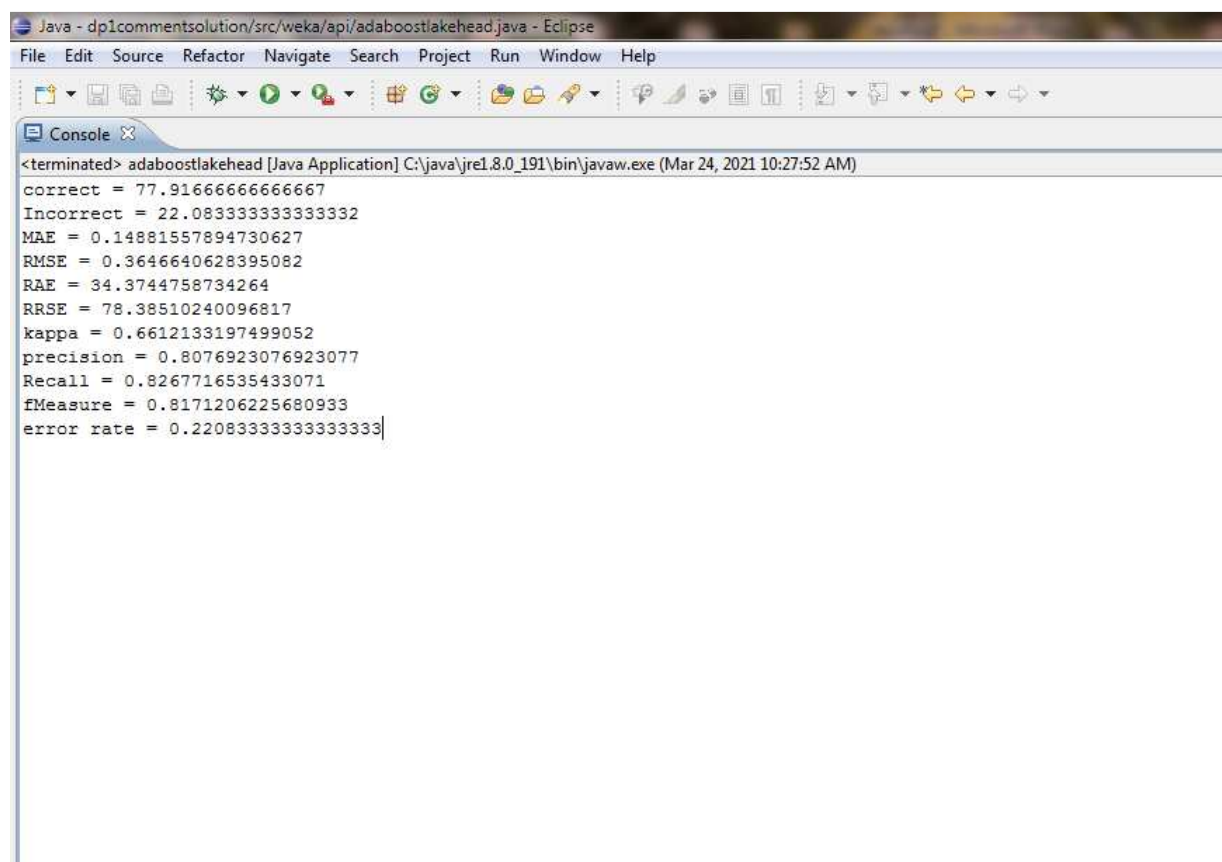
Java - dp1commentssolution/src/weka/api/masterpro1.java - Eclipse

File Edit Source Refactor Navigate Search Project Run Window Help

Console

```
<terminated> masterpro1 (1) [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 24, 2021 9:47:39 AM)
correct = 75.83333333333333
Incorrect = 24.166666666666668
MAE = 0.21549610973263705
RMSE = 0.36321553477934004
RAE = 49.776816897945245|
RRSE = 78.07373906167216
kappa = 0.6289187460012795
precision = 0.8429752066115702
Recall = 0.8031496062992126
fMeasure = 0.8225806451612903
error rate = 0.24166666666666667
```

Proposed Method Output Screenshot :



Java - dp1commentssolution/src/weka/api/adaboostlakehead.java - Eclipse

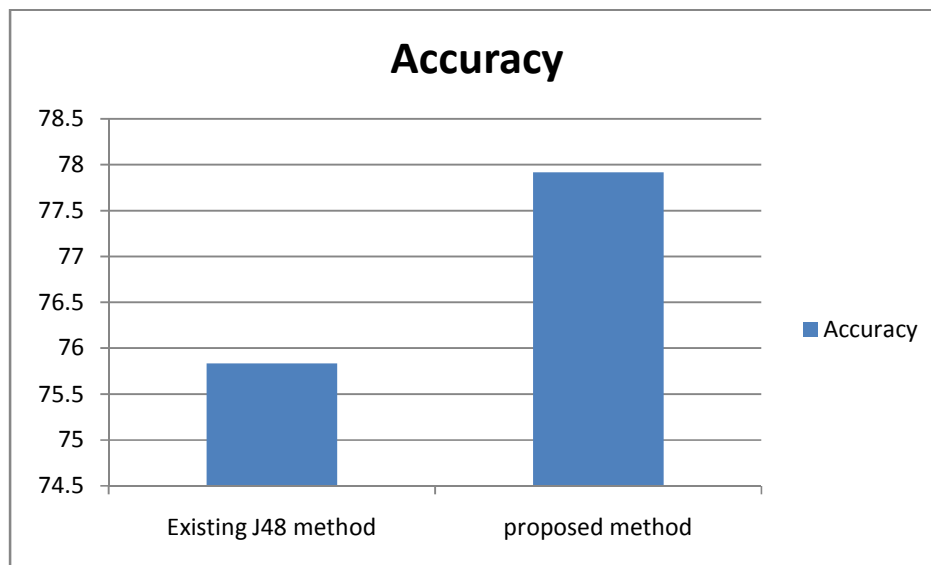
File Edit Source Refactor Navigate Search Project Run Window Help

Console

```
<terminated> adaboostlakehead [Java Application] C:\java\jre1.8.0_191\bin\javaw.exe (Mar 24, 2021 10:27:52 AM)
correct = 77.91666666666667
Incorrect = 22.083333333333332
MAE = 0.14881557894730627
RMSE = 0.3646640628395082
RAE = 34.3744758734264
RRSE = 78.38510240096817
kappa = 0.6612133197499052
precision = 0.8076923076923077
Recall = 0.8267716535433071
fMeasure = 0.8171206225680933
error rate = 0.22083333333333333|
```

Existing and Proposed Method Comparision :

Existing and Proposed Method Accuracy Comparison :



	Existing method	Proposed method
Accuracy	75.8333%	77.9167%

Accuracy for Existing J48 method : 75.8333%

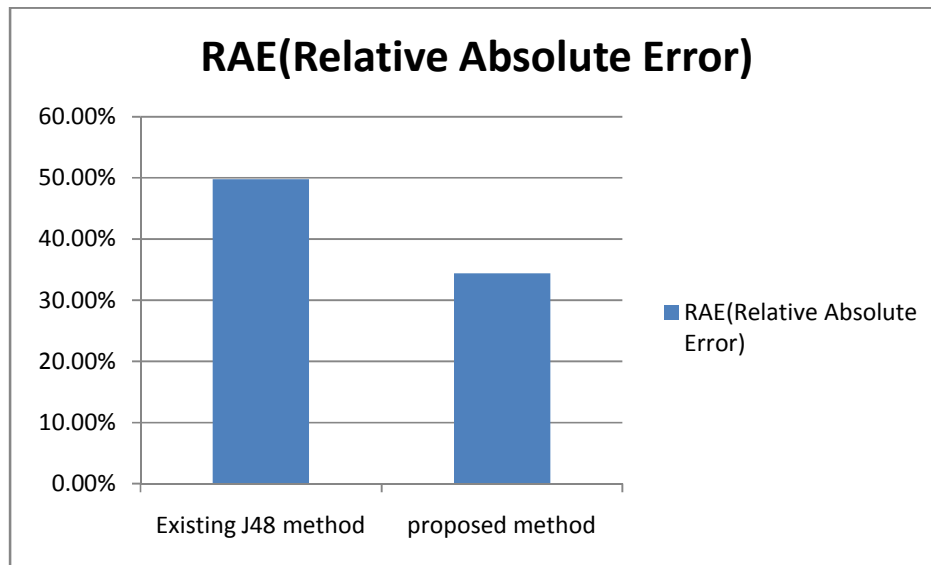
Accuracy for proposed method : 77.9167%

I applied existing method and proposed method on student academic performance dataset.

Proposed method of j48 based boosting technique improve classification accuracy.

Existing and Proposed Method RAE(Relative Absolute Error)

Comparison :



	Existing Method	Proposed Method
RAE	49.7778%	34.3744%

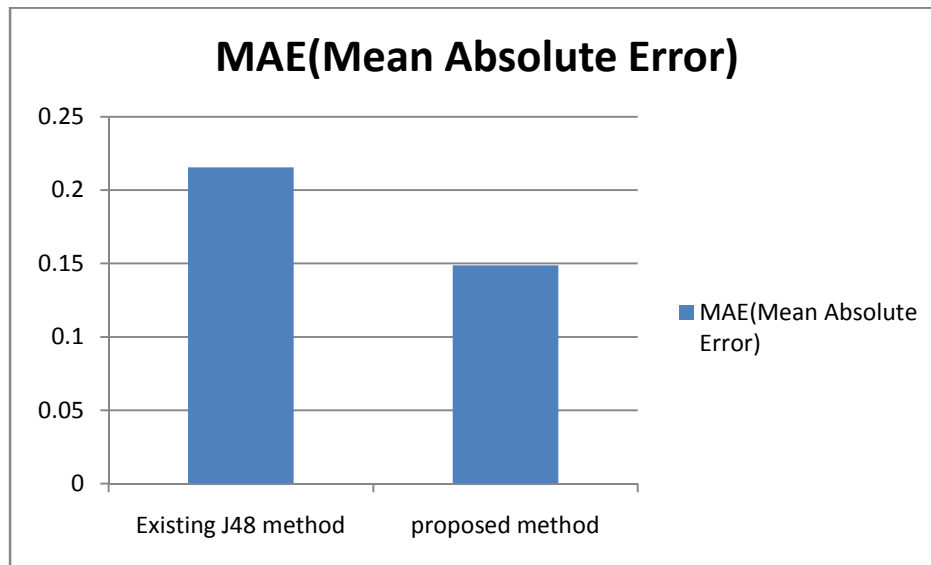
RAE for Existing J48 method : 49.7778%

RAE for proposed method : 34.3744%

Proposed method of j48 based boosting technique reduce Relative Absolute Error of student academic performance prediction.

Existing and Proposed Method MAE(Mean Absolute Error)

Comparison :



	Existing Method	Proposed Method
MAE	0.2154%	0.1488%

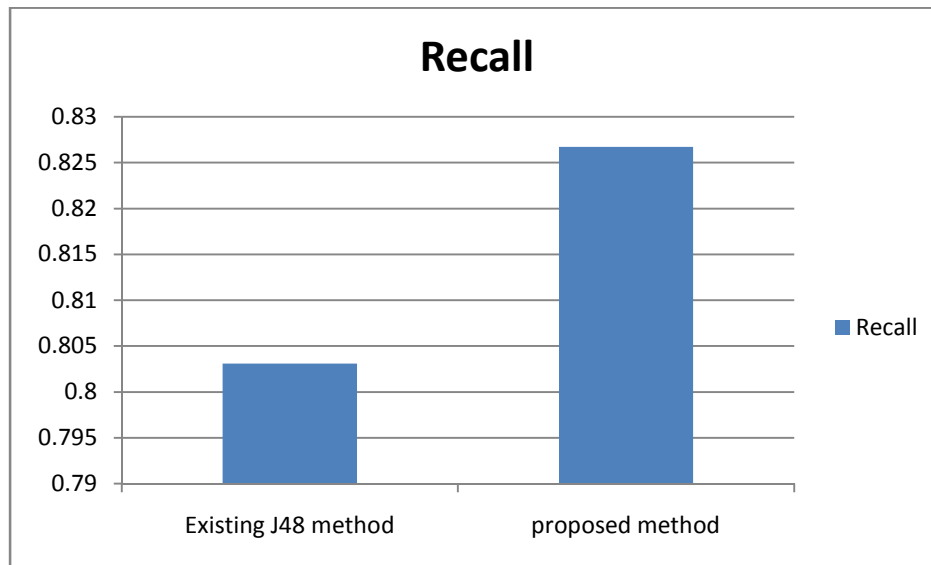
MAE for Existing J48 method : 0.2154%

MAE for proposed method : 0.1488%

Proposed method of j48 based boosting technique reduce Mean Absolute Error.

Existing and Proposed Method Recall :

Comparison :



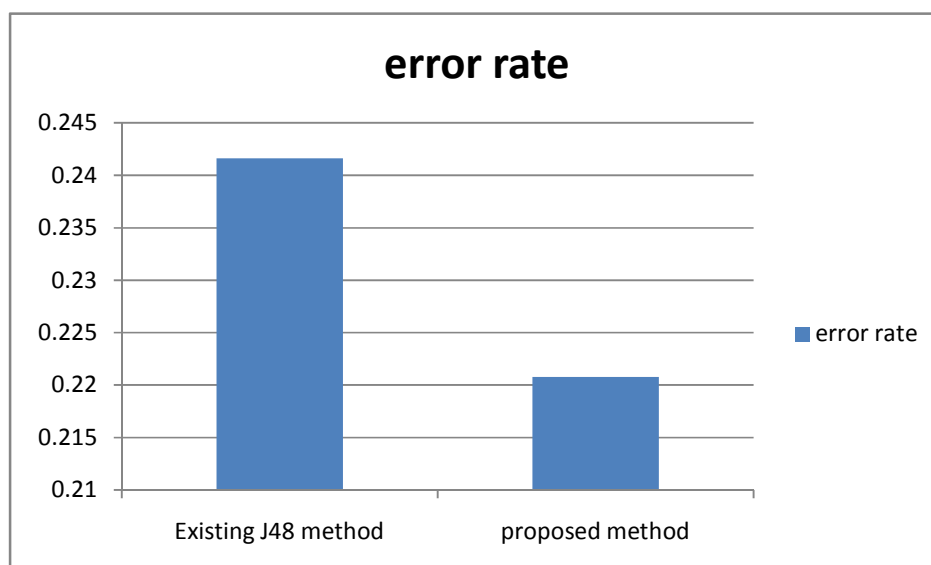
	Existing Method	Proposed Method
Recall	0.8031	0.8267

Recall for Existing J48 method : 0.8031

Recall for proposed method : 0.8267

Proposed method of j48 based boosting technique increase Recall.

Existing and Proposed Method Error rate :



	Existing Method	Proposed Method
Error Rate	0.2416	0.2208

Error Rate for Existing J48 method : 0.2416

Error Rate for proposed method : 0.2208

Proposed method of j48 based boosting technique reduce Error Rate.

In our research I compare various classification technique's accuracy and errors. Various classification techniques are like Random tree, J48, Naïve Bayes, Bayesnet and K-nearest neighbour. From all these, J48 classifier gives highest accuracy. To improve classification accuracy further, in our research there is proposed method improved classification technique is used that helps to boost the performance of classification technique. This proposed method improve the accuracy, Recall and reduce errors like MAE(Mean Absolute Error), RAE(Relative Absolute Error) and Error Rate. Proposed method boosting based J48 method increase accuracy from 75.83% to 77.91%. Proposed method boosting technique based J48 also increases Recall from 0.8031 to 0.8267. It reduces errors like RAE(Relative Absolute Error) from 49.77% to 34.37%. and MAE(Mean Absolute Error) from 0.2154% to 0.1488%. This proposed technique also helps to reduce error rate from 0.2416 to 0.2208. I use cross validation technique for our research. In **K Fold cross validation**, the data is divided into k subsets. Now the holdout method is repeated k times, such that each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set.

5. Conclusion and Future Scope :

Proposed Technique J48 based boosting technique improves the classification accuracy of student academic performance, Recall and reduce Errors like MAE(Mean Absolute Error), RAE(Relative Absolute Error). To increase further classification accuracy of student academic performance, parameter tuning of proposed method is done . It helps to classify correct instances more precisely. To improve classification accuracy, more dataset entries will help to improve classification accuracy and reduce errors.

6. References :

- [1].V. Sathya Durga, J. Thangakumar, "Students Performance Prediction through Educational Data Mining - An Uncomplicated Review", International Research Journal of Engineering and Technology (IRJET), 2019, pp. 1404-1406

[2].Vandna Dahiya, "A Survey on Educational Data Mining", International Journal of Research in Humanities, Arts and Literature, Vol. 6, Issue.5, pp. 23-30, 2018

[3].T.Thilagaraj, Dr.N Sengottaiyan "A Review of Educational Data Mining in Higher Education System" In Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering, pp. 349-358, 2017.

[4].Bindhia K. Francis & Suvanam Sasidhar Babu," Predicting Academic Performance of Students Using a Hybrid Data Mining Approach", Journal of Medical Systems ,springer,pp.1-15.

[5].Muhammad Imran, Shahzad Latif, Danish Mehmood, Muhammad Saqlain Shah, "Student Academic Performance Prediction using Supervised Learning Techniques",IJET–Vol. 14, No. 14, 2019.

[6].<https://educationdata.org/college-enrollment-statistics>

[7]. Md. Hasibur Rahman, Md. Rabiul Islam," Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques", 2nd International Conference on Electrical & Electronic Engineering (ICEEE), 27-28 December 2017, RUET, Rajshahi, Bangladesh.

