IMPLEMENTATION OF ENHANCED DATA MINING ALGORITHM TO PREDICTION OF STUDENTS PERFORMANCE USING EDUCATIONAL SYSTEM

[1]GURPREET SINGH, [2]ER. HARJIT SINGH, [3]ER. HARAVTAR SINGH

[1]Professor, Department of Computer Sci. &Engg. ,St. Soldier Group of InstitutionsNear NIT, Jalandhar(Punjab)-144011, INDIA
[2]Assistant Professor, Department of Computer Sci. &Engg , St. Soldier Group of Institutions Near NIT,Jalandhar(Punjab)-144011, INDIA
[3]Network System Engineer, Department of CSE., St. Soldier Group of Institutions Near NIT, Jalandhar(Punjab)-144011, INDIA
E-mail: [1]saini3077@gmail.com, [2]er.harjeetsingh81@gmail.com, [3]er.haravtar@gmail.com

Abstract - In this exploration work, the execution of different element determination calculations was assessed on various arrangement calculation utilizing the understudies execution dataset created for the examination. The proposed examination made a few correlations with assess the adequacy of the element determination systems utilizing the measures including mistake and exactness parameters. The general point of the examination was to dissect the adequacy of different calculations to anticipate poor, normal and great students on premise of scholastic execution as well as different variables. The dataset for the investigation was alluded from the school that incorporated the subtleties of the understudies like QOS, HOCC. Different parameters were additionally viewed as utilizing sham information to assess the general execution of the understudy aside from the scholastic execution. Along these lines, it could be helpful to the instructive pioneers and the board of the universities, if the highlights in the as of now accessible information can be going about as the pointer for foreseeing the execution of the understudies. The real goal of this examination is to break down the understudy's information accessible in the school to distinguish a particular examples that may be valuable in the forecast of their execution. The particular goal of the investigation is to order understudies as indicated by their execution in the scholarly just as different factors separated from scholastic execution.

Keywords - ENBA, DT, EDM, DM, QOS, HOCC

I. INTRODUCTION

Information Mining is the programmed revelation of connections normally in substantial database and, in certain cases, the utilization of the disclosure results in basic leadership. This is a fundamental procedure where canny strategies are connected so as to separate information designs.

Two fundamental motivations to utilize information mining:

- Too much information and too little data
- There is a need to remove valuable data from the information and to decipher the information

can be either utilized in a robotized decision candidly steady system. This is the motivation to use data mining, especially in science and business areas which need to dismember a great deal of data to discover designs which they couldn't by and large find.

II. EDUCATIONAL DATA MINING

Instructive Data Mining is a rising order, worried about creating strategies for investigating the remarkable kinds of information that originate from instructive settings, and utilizing those techniques for better development of personnel. Information mining is extraction of fascinating examples or learning from immense measure of information. [1]As we know substantial measure of information is put away in instructive database, so as to get required information and to locate the concealed relationship, diverse information mining strategies are created and utilized. There are assortments of prevalent information mining undertaking inside the instructive information mining for example arrangement, bunching, exception identification, affiliation rule, forecast and so on. Characterization and expectation are two types of information investigation that can be utilized to remove models depicting vital information classes or to foresee future information patterns. Such investigation can help give us a superior comprehension of the information on the loose. Here we apply essential strategies for information arrangement, for example, how to assemble choice tree classifiers, Bayesian classifiers, a test set is utilized, comprised of test tuples and their related class marks. These tuples are arbitrarily chosen from the general informational index. They are free of the preparation tuples, implying that they are not used to build the classifier. The exactness of a classifier on a given test set is the dimension of test set tuples that are adequately portrayed by the classifier. The related class characteristic of each test tuple is differentiated and the informed classifier's class conjecture for that tuple. delineates a couple of procedures for surveying classifier accuracy[2].In the event that the exactness of the classifier is viewed as satisfactory. The

classifier can be utilized to order future information tuples for which the class mark isn't known. [3]Different characterization forms are information cleaning, pertinence investigation, characteristic subset determination, information change, and decrease.

III. BAYESIAN CLASSIFICATION

Bayesian classifiers are quantifiable classifiers. They can anticipate class investment probabilities, for instance, the probability that a given tuple has a spot with a particular class. Bayesian characterization depends on Bayes' hypothesis, portrayed underneath. Studies looking at order calculations have discovered a basic Bayesian classifier known as the guileless Bayesian classifier to be similar in execution with choice tree and chose neural system classifiers. Bayesian classifiers have additionally displayed high precision and speed when connected to extensive databases. Innocent Bayesian classifiers expect that the impact of a trait esteem on a given class is free of the estimations of alternate qualities. This presumption is called class contingent autonomy. It is made to improve the calculations included and, in this sense, is considered "gullible." Bayesian conviction systems are graphical models, which dissimilar to innocent Bayesian classifiers, permit the portrayal of conditions among subsets of qualities. Bayesian conviction systems can likewise be utilized for grouping

IV. PROPOSED METHOD

Data mining assumes a critical job in the business world and it serves to the instructive organization to foresee and settle on choices identified with the understudies' scholarly status. [3]The existing framework is a framework which keeps up the understudy data as numerical qualities and it just stores and recover the data what it contains. So the framework has no insight to investigate the information. The proposed framework is an information mining calculation which makes utilization of the upgraded Bayesian calculation procedure for the extraction of helpful data. Result demonstrates that upgraded choice tree system gives more exactness over different strategies like Decision Trees and bunching for examination and forecast.

As aggressive condition is winning among the scholastic establishments, challenge is to expand the nature of training through information mining. Understudy's execution is of extraordinary worry to the advanced education. In this exploration , we have connected information mining systems by assessing understudy's information utilizing improved Bayesian calculation which is useful in foreseeing the understudy's outcomes. we have determined the Entropy of the qualities taken in Educational Data Set and the characteristic having most astounding Information Gain is taken as the root hub to part further. The outcomes created utilizing Data Mining Techniques help employees to concentrate on understudies who are getting poor class results.

V. OBJECTIVE OF STUDY

In this examination work, the execution of different component determination calculations was assessed on various grouping calculation utilizing the understudies execution dataset produced for the investigation. [4]The proposed ponder made a few correlations with assess the viability of the component determination methods utilizing the measures including mistake and precision parameters. The general point of the examination was to break down the adequacy of different calculations to anticipate poor, normal and great students on premise of scholastic execution as well as different components. The dataset for the investigation was alluded from the school that incorporated the subtleties of the understudies like QOS, HOCC. Different parameters were likewise viewed as utilizing sham information to assess the general execution of the understudy with the exception of the scholastic execution. Consequently, it could be helpful to the instructive pioneers and the executives of the universities, if the highlights in the right now accessible information can be going about as the marker for anticipating the execution of the understudies.

The real goal of this investigation is to break down the understudy's information accessible in the school to distinguish a particular examples that may be helpful in the expectation of their execution. [9]The explicit target of the examination is to order understudies as per their execution in the scholarly just as different factors separated from scholastic execution.

Research Objectives are:

1. To execute upgraded Bayesian calculation assembled an explanatory model of understudy accomplishment, utilizing the model in the wake of pruning calculation.
2. To improve the forecasts on understudies execution and information handling rate of Education framework.
3. To examinations of three classifiers and finds the best performing order calculation among all dependent on right arranged example, off base characterized case and blunder rate.

VI. RESEARCH METHODOLOGY

Data mining is the learning revelation process from an immense information volume. The component works in vast dataset where the understudy execution in the semester exercises is assessed.

A. Data preparation

Understudy related information were gathered from the school on the testing technique for software engineering office from the session 2014 to 2017. In this progression, information put away in various tables were joined into a solitary set.

B. Data selection and transformation

In this progression just those fields were chosen which were required for the information mining process. This is appeared in the table 4.1.

TABLE 1: STUDENT RELATED VARIABLES

Sr. No.	DESCRIPTION	
1	QOS	Quality of Syllabus
2	QOIP	Quality of Instruction Plans
3	HOCC	Handling of committees clubs
4	RM	Record Maintenance
5	EVA	Evaluation
6	QP	Question Paper
7	CACP	CA of capstone project
8	RPP	Research Paper published
9	CA	Conferenceattended paper published
10	WA	Workshop attended
11	BP	Book Computers published
12	DM	Discipline Maintenance
13	IWS	Interactivity with students
14	EMTE	Evaluation of MTE and ETE
15	AWA	Award
16	LA	Leadership ability
17	FDP	Faculty Development program
18	QUAL	Qualification

PROPOSED ALGORITHM

Information: the preparation test S; competitor property set trait
list; grouping trait C.
Yield: an upgraded Naïve Bayesian calculation.
a) Create a root hub N;
b) If the preparation set is unfilled, at that point come back to hub N and
set apart as Failure;
c) IF T have a place with a similar class C, at that point return a leaf
hub N, marked as class C;
d) IF property list is vacant OR T in the remaining
tests is not exactly a given esteem, at that point return a leaf hub N,
marked N is the most continuous class in T;
e) FOR each property list {
In the event that competitor quality is nonstop, at that point scattering
the trait;
Compute data gain proportion ;}
f) Choosing the hopeful quality with the most noteworthy
data gain proportion, and stamping properties as hubs N;
g) For each another leaf hub created by N {
In the event that the leaf hub compares to a subset of the
test * T is vacant
At that point
This leaf hub is part to create another leaf
hub, mark it as the most incessant class in T
ELSE
Performed on the leaf hubs in adjusted structure tree
(T, * T. Quality rundown) keeps on isolating it

VII. RESULTS AND DISCUSSION

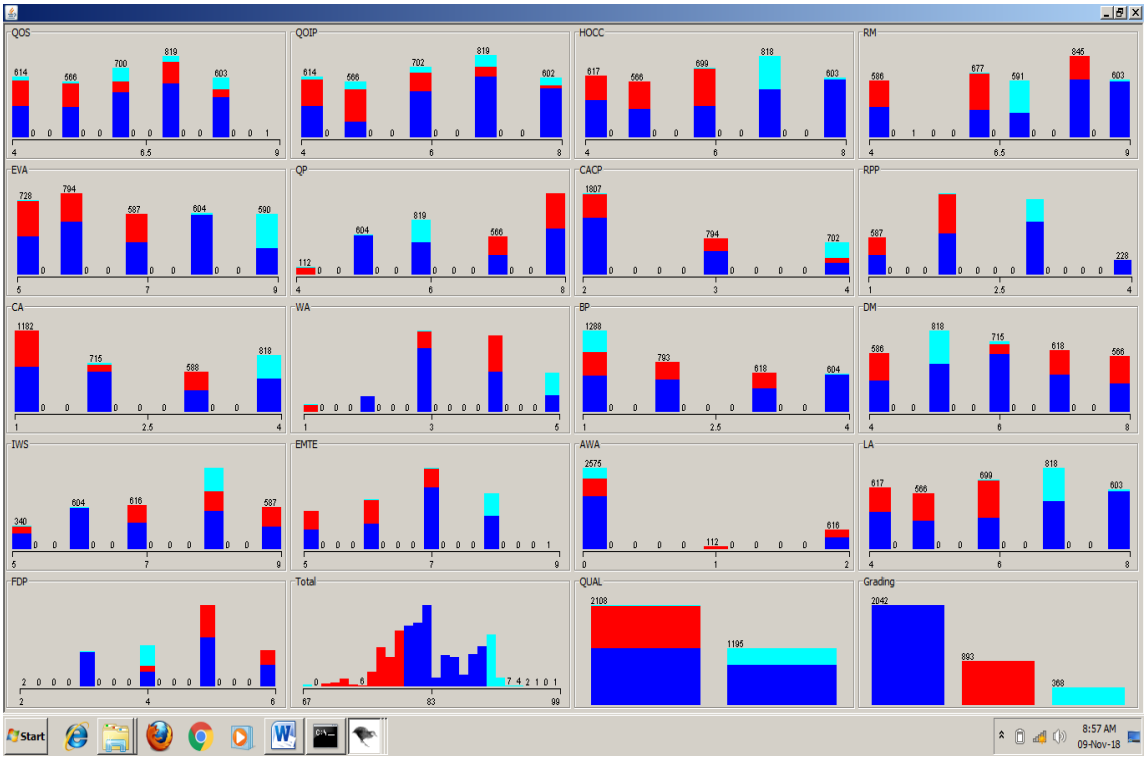


Figure 1: Visualisation of all attributes

Table 2: Result Analysis of different algorithms with ENBA

	Naïve Bayes	Decision Stump	User Classifier	ENBA
Correctly Classified Instances	65.18	88.85	61.82	89.76
Incorrect classified Instances	34.81	11.14	38.17	10.23
Error Rate	64.55	35.79	99.97	24.7

```
Time taken to build model: 0.08 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        2965               89.7669 %
Incorrectly Classified Instances       338               10.2331 %
Kappa statistic                          0.8213
Mean absolute error                      0.0877
Root mean squared error                  0.26
Relative absolute error                 24.7083 %
Root relative squared error             61.7153 %
Total Number of Instances             3303

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.852     0.016      0.989     0.852     0.915    Average
  0.999     0.026      0.935     0.999     0.966    Poor
  0.908     0.087      0.566     0.908     0.697    good

=== Confusion Matrix ===

    a    b    c   <-- classified as
 1739   47  256 |    a = Average
    1  892    0 |    b = Poor
   19   15  334 |    c = good
```

Figure 2: shows the results depict the Correctly Classified Instances, Incorrect Classified Instances and Error Rate of ENBA algorithm.

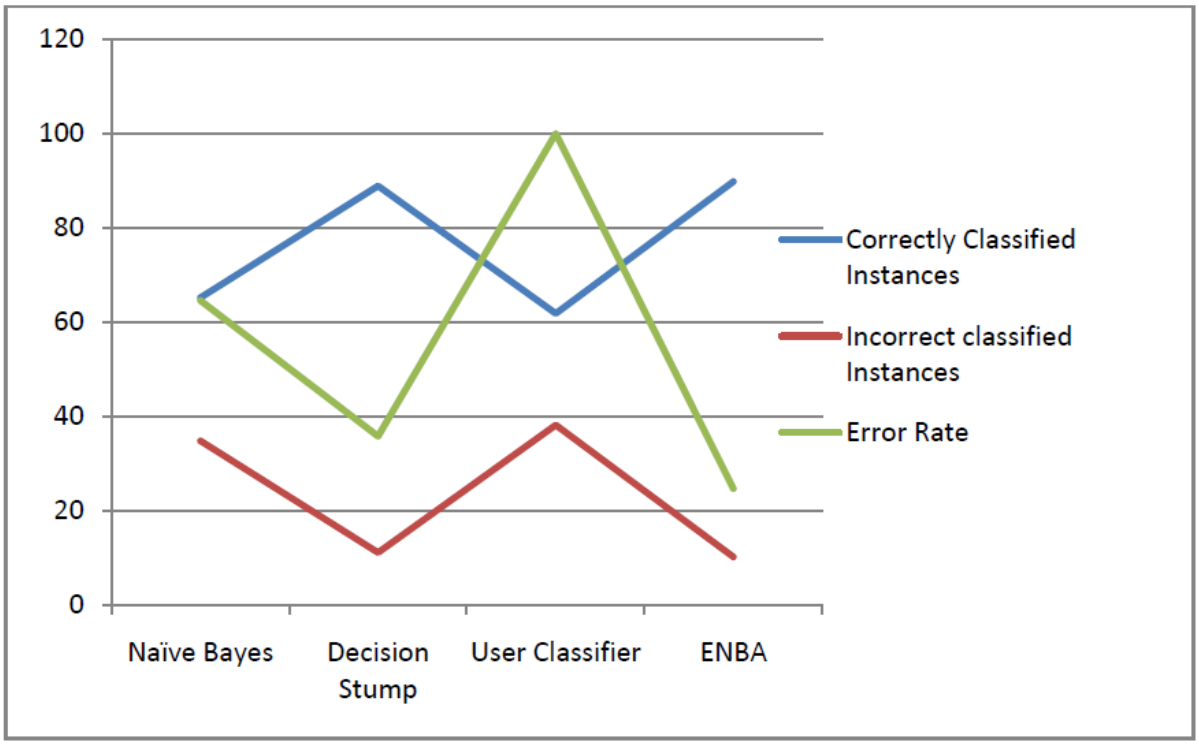


Figure 3: shows the Result Analysis of different algorithms with ENBA

Table 3: the comparative results of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms on Correctly Classified Instances

	Naïve Bayes	Decision Stump	User Classifier	ENBA
Correctly Classified Instances	65.18	88.85	61.82	89.76

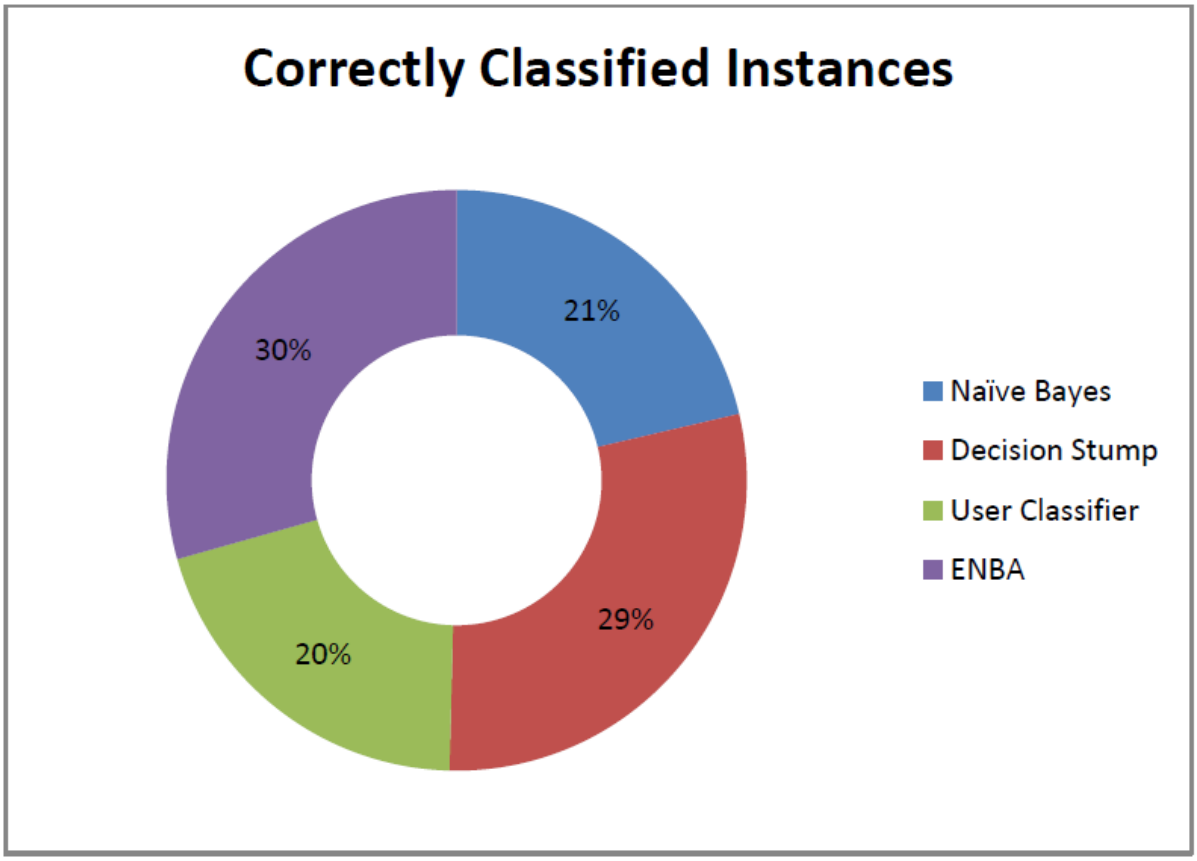


Figure4:Shows the percentage of Correctly Classified Instances of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

Table 4: the comparative results of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms on Incorrect Classified Instances

	Naïve Bayes	Decision Stump	User Classifier	ENBA
Incorrect classified Instances	34.81	11.14	38.17	10.23

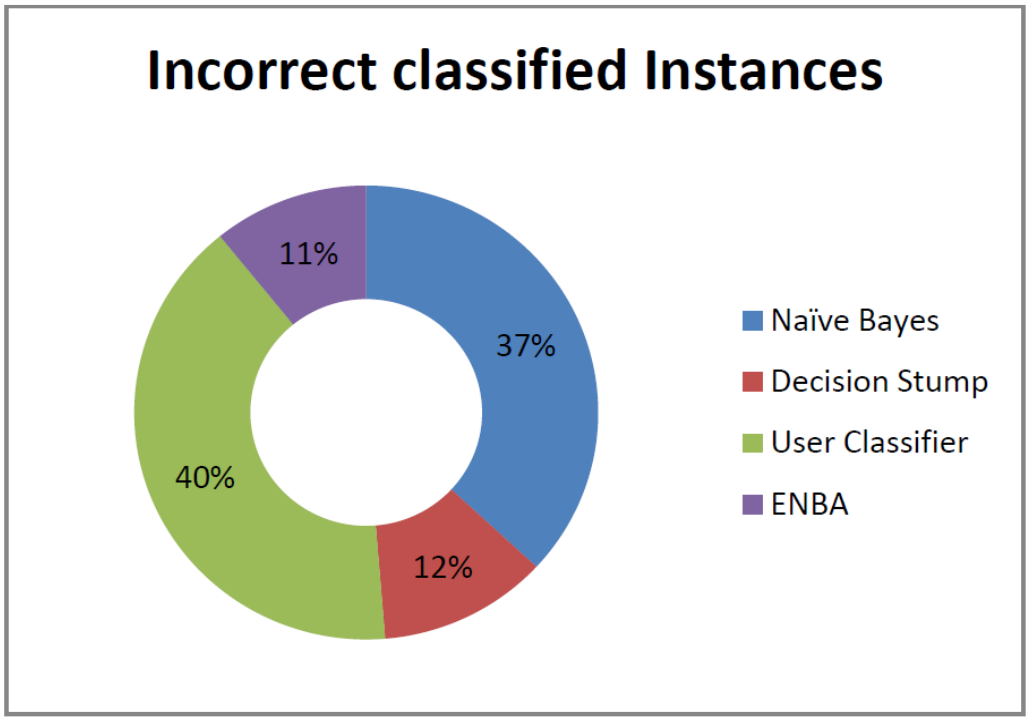


Figure 5: Shows the percentage of Incorrect Classified Instances of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

Table 5: the comparative results of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms on Error Rate

	Naïve Bayes	Decision Stump	User Classifier	ENBA
Error Rate	64.55	35.79	99.97	24.7

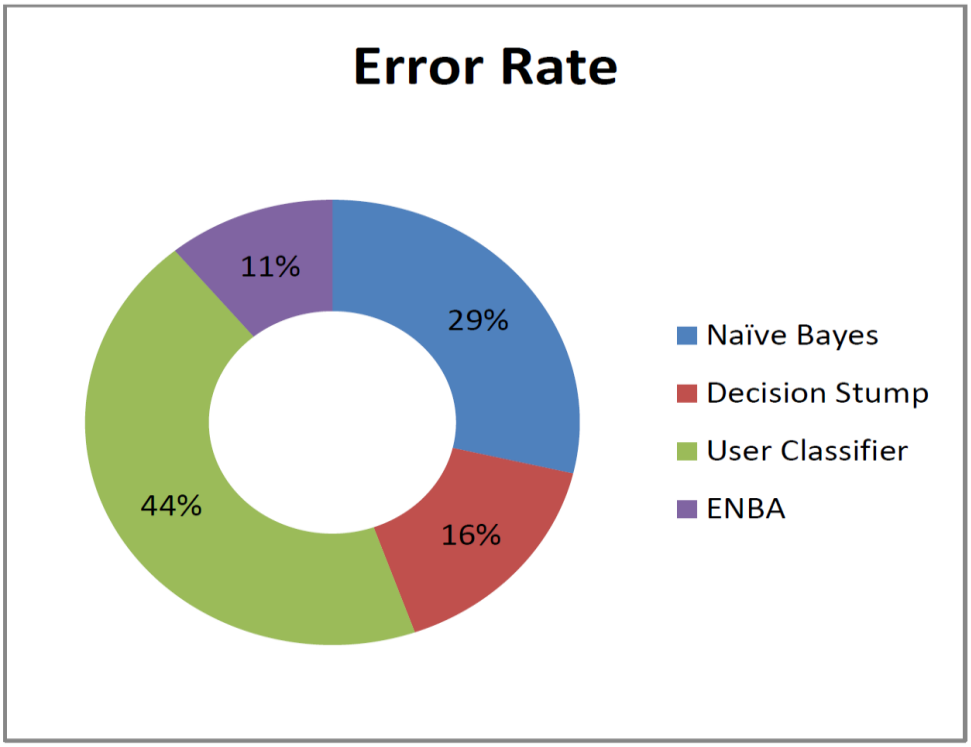


Figure 6: Shows the percentage of Accuracy of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

CONCLUSIONS AND FUTURE SCOPES

In this examination, Using Enhanced Naïve Bayesian Algorithm and Decision Tree Classifiers we have tended to the expectation of Class Grading of understudies dependent on the characteristics taken. Upgraded Naïve Bayesian Algorithm and Decision tree classifiers are utilized on understudy's scholarly information to anticipate the understudy's execution in class reviewing. This examination will help in distinguishing those understudies who are beneath scholarly exercises and appeared, normal and great execution in instruction framework.

The outcome demonstrates that poor execution in evaluating is principle factor of understudy's disappointment in last tests of the year. The fundamental finding of this examination is the social occasion of information from understudy's scholastic execution and to distinguish those understudies who need extraordinary consideration. [11]This think about likewise confirms that entropy determined by us demonstrates a similar property having most astounding data gain which is being taken by WEKA recreation instrument.

Further, information mining strategies can be useful in breaking down patterns in instruction, medicinal, business division and so forth.

In this exploration, utilizing Enhanced Naïve Bayesian Algorithm fabricated an investigative model of understudy execution, utilizing the model subsequent to pruning calculation the model was improved by utilizing the model in the wake of pruning calculation and to assess Correctly ordered cases, [12]Incorrect arranged examples and precision proportion of the grouping technique, it accomplished great the exhuming results.In future course, we will survey the different characterization calculations and essentialness of Prediction approach in structuring of proficient characterization calculations for information mining. Determination of information and strategies for information mining is an imperative undertaking in this procedure and requirements the learning of the area.

REFERENCES

[1] KhokhoniInnocentiaMphoRamaphosa, TranosZuva, Raoul Kwuimi, "Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools", 2018 IEEE.

[2] Nguyen Truc Mai Anh, Vo Thi Ngoc Chau, Nguyen HuaPhung, "Towards a robust incomplete data handling approachto effective educational data classification in anacademic credit system", 2014 IEEE.

[3] Ricardo Timarán Pereira, Javier CaicedoZambrano, "Aplication of Decision Trees for Detection of Student Dropout Profiles", 16th IEEE International Conference on Machine Learning and Applications 2017.

[4] Sneha Chandra, ManeetKaur, "Creation of an Adaptive Classifier to Enhance theClassification Accuracy of Existing ClassificationAlgorithms in the Field of Medical Data Mining", 2015 IEEE.

[5] Boris Delibašić, Member, IEEE," White-Box or Black-Box Decision Tree Algorithms:Which to Use in Education?", IEEE TRANSACTIONS ON EDUCATION 2012.

[6] NikolaosDimokas, NikolaosMittas, AlexandrosNanopoulos, Lefteris Angelis, "A Prototype System for Educational Data Warehousing and Mining", Panhellenic Conference on Informatics, IEEE 2008.

[7] Tomas Hasbun, Alexandra Araya and Jorge Villalon," Extracurricular activities as dropout prediction factors in higher education usingdecision trees", 2016 IEEE 16th International Conference on Advanced Learning Technologies.

[8] Hanjun Jin, Tianzhen Wu, "Application of Visual Data Mining in Higher-education Evaluation System", 2009 First International Workshop on Education Technology and Computer Science.

[9] xiaojian.long, "Application of decision tree in student achievementEvaluation", 2012 International Conference on Computer Science and Electronics Engineering.

★★★