ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.11

Implementation of Classification Based Algorithms to Prediction of Students Performance Using Educational System

Jatinderpal Singh[1], Er. Harwinder Kaur[2]
M.Tech Scholar[1], Assistant Professor[2]
Department of Computer Science and Engineering
St. Soldier Institute of Engg. & Technology, Near NIT, Jalandhar, Punjab, India

Abstract:
In this research work, the performance of various feature selection algorithms was evaluated on different classification algorithm using the students performance dataset generated for the study. The proposed study made several comparisons to evaluate the effectiveness of the feature selection techniques using the measures involving error and accuracy parameters. The overall aim of the study was to analyze the effectiveness of various algorithms to predict poor, average and good learners not only on basis of academic performance but also other factors. The dataset for the study was referred from the college that included the details of the students like QOS, HOCC. Other parameters were also considered using dummy data to evaluate the overall performance of the student except the academic performance. Thus, it could be useful to the educational leaders and management of the colleges, if the features in the currently available data can be acting as the indicator for predicting the performance of the students. The major objective of this study is to analyze the student's data available in the college to identify any specific patterns that might be useful in the prediction of their performance. The specific objective of the study is to classify students according to their performance in the academic as well as other factors apart from academic performance.

Keywords: ENBA, DT, EDM, DM

INTRODUCTION

Data Mining is the automatic discovery of relationships typically in large database and, in some instances, the use of the discovery results in decision making. This is an essential process where intelligent methods are applied in order to extract data patterns.

Two main reasons to use data mining:

- Too much data and too little information
- There is a need to extract useful information from the data and to interpret the data

However, Data mining can automate the process of finding relationships and patterns in raw data and the results can be either utilized in an automated decision support system. This is why to use data mining, especially in science and business areas which need to analyze large amounts of data to discover trends which they could not otherwise find.

Educational Data Mining

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods for better growth of faculty. Data mining is extraction of interesting patterns or knowledge from huge amount of data. As we know large amount of data is stored in educational database, so in order to get required data & to find the hidden relationship, different data mining techniques are developed & used. There are varieties of popular data mining task within the educational data mining e.g. classification, clustering, outlier detection, association rule, prediction etc

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large. Here we apply basic techniques for data classification, such as how to build decision tree classifiers, Bayesian classifiers, a test set is used, made up of test tuples and their associated class labels. These tuples are randomly selected from the general data set. They are independent of the training tuples, meaning that they are not used to construct the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. The associated class label of each test tuple is compared with the learned classifier's class prediction for that tuple. describes several methods for estimating classifier accuracy[2]. If the accuracy of the classifier is considered acceptable. The classifier can be used to classify future data tuples for which the class label is not known. Different classification processes are data cleaning, relevance analysis, attribute subset selection, data transformation, and reduction.

BAYESIAN CLASSIFICATION

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Bayesian classification is based on Bayes' theorem, described below. Studies comparing classification algorithms have found a simple Bayesian classifier known as the *naïve Bayesian classifier* to be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called *class conditional independence*. It is made to simplify the computations involved and, in this sense, is considered "naïve." *Bayesian belief networks* are graphical models, which unlike naïve Bayesian classifiers, allow the representation of dependencies among subsets of attributes. Bayesian belief networks can also be used for classification.

Proposed Method

Data mining plays an important role in the business world and it helps to the educational institution to predict and make decisions related to the students' academic status. The existing system is a system which maintains the student information in the form of numerical values and it just stores and retrieve the information what it contains. So the system has no intelligence to analyze the data. The proposed system is a data mining algorithm which makes use of the enhanced Bayesian algorithm technique for the extraction of useful information. Result proves that enhanced decision tree technique provides more accuracy over other methods like Decision Trees and clustering for comparison and prediction.

As competitive environment is prevailing among the academic institutions, challenge is to increase the quality of education through data mining. Student's performance is of great concern to the higher education. In this research, we have applied data mining techniques by evaluating student's data using enhanced Bayesian algorithm which is helpful in predicting the student's results. We have calculated the Entropy of the attributes taken in Educational Data Set and the attribute having highest Information Gain is taken as the root node to split further. The results generated using Data Mining Techniques help faculty members to focus on students who are getting poor class results.

OBJECTIVE OF STUDY

In this research work, the performance of various feature selection algorithms was evaluated on different classification algorithm using the students performance dataset generated for the study. The proposed study made several comparisons to evaluate the effectiveness of the feature selection techniques using the measures involving error and accuracy parameters. The overall aim of the study was to analyze the effectiveness of various algorithms to predict poor, average and good learners not only on basis of academic performance but also other factors. The dataset for the study was referred from the college that included the details of the students like QOS, HOCC. Other parameters were also considered using dummy data to evaluate the overall performance of the student except the academic performance. Thus, it could be useful to the educational leaders and management of the colleges, if the features in the currently available data can be acting as the indicator for predicting the performance of the students.

The major objective of this study is to analyze the student's data available in the college to identify any specific patterns that might be useful in the prediction of their performance. The specific objective of the study is to classify students according to their performance in the academic as well as other factors apart from academic performance.

My Research Objectives are:

1. To implement enhanced Bayesian algorithm built an analytical model of student achievement, using the model after pruning algorithm.
2. To improve the predictions on students performance and data processing speed of Education system.
3. To analyses of three classifiers and finds the best performing classification algorithm among all based on correct classified instance, incorrect classified instance and error rate.

RESEARCH METHODOLOGY

Data mining is the knowledge discovery process from a huge data volume. The mechanism works in large dataset where the student performance in the semester activities is evaluated.

A. Data preparation

Student related data were collected from the college on the sampling method of computer science department from the session 2014 to 2017. In this step, data stored in different tables were joined into a single set.

B. Data selection and transformation

In this step only those fields were selected which were required for the data mining process. This is shown 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	Conference attended 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

Input: the training sample S; candidate attribute set attribute list; classification attribute C.

Output: a enhanced Naïve Bayesian algorithm.

```
a) Create a root node N;

b) If the training set is empty, then return to node N and
marked as Failure;

c) IF T belong to the same class C, then return a leaf node N,
labeled as class C;

d) IF attribute list is empty OR T in the remaining samples is
less than a given value, then return a leaf node N, labeled N is
the most frequent class in T;

e) FOR each attribute list {
        IF candidate attribute is continuous then dispersing
        the attribute;
        Calculate information gain ratio ;}

f) Choosing the candidate attribute with the highest
information gain ratio, and marking properties as nodes N;

g) For each a new leaf node produced by N {
    IF the leaf node corresponds to a subset of the sample * T
    is empty

    THEN
    This leaf node is split to generate a new leaf node, mark it
    as the most frequent class in T
```

ELSE
Performed on the leaf nodes in modified form tree (T, * T.
Attribute list) continues to divide it

RESULTS AND DISCUSSION

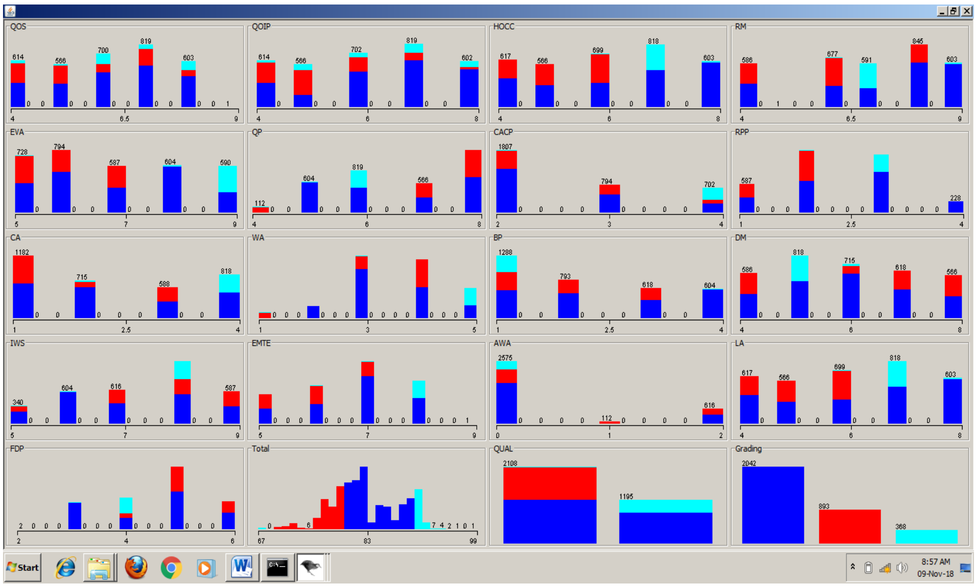
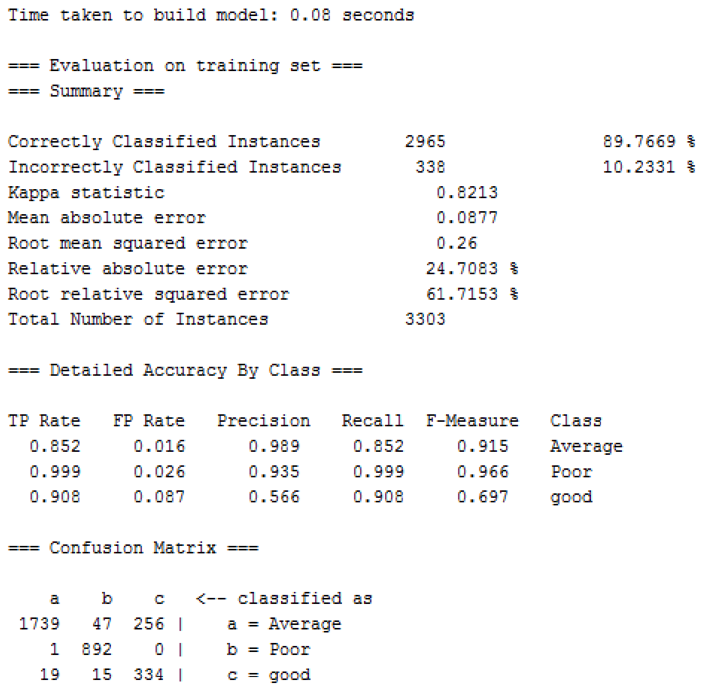


Figure 1: Visualisation of all attributes



```
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.

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

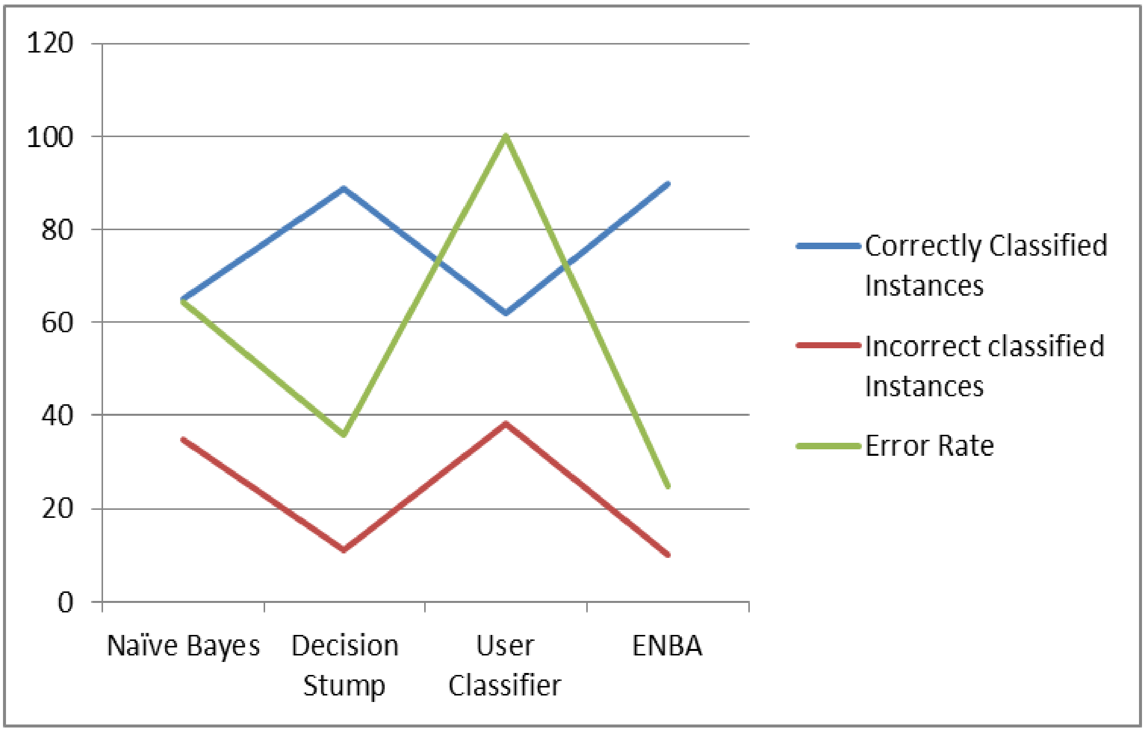


Figure 3: shows the Result Analysis of different algorithms with ENBA

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

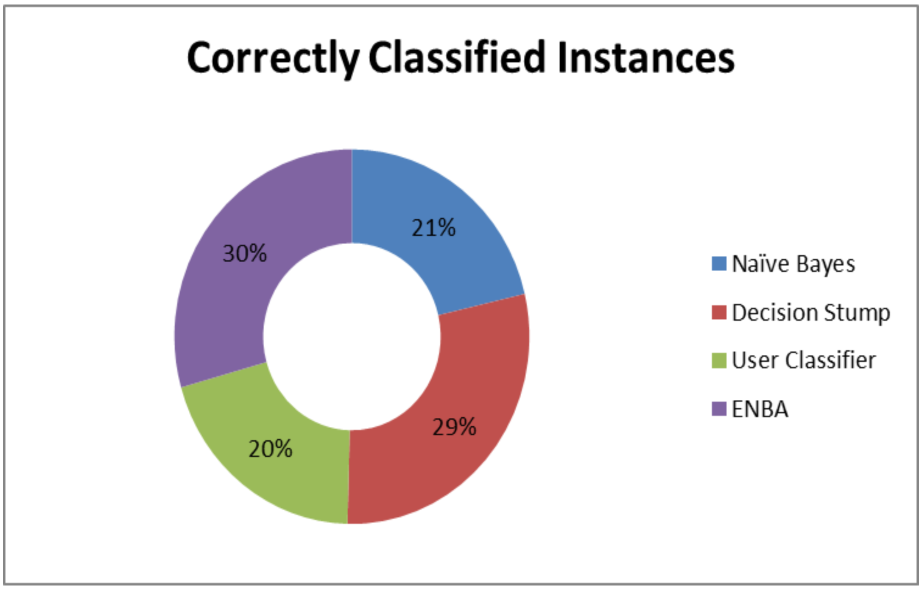


Figure 4: Shows the percentage of Correctly Classified Instances of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

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

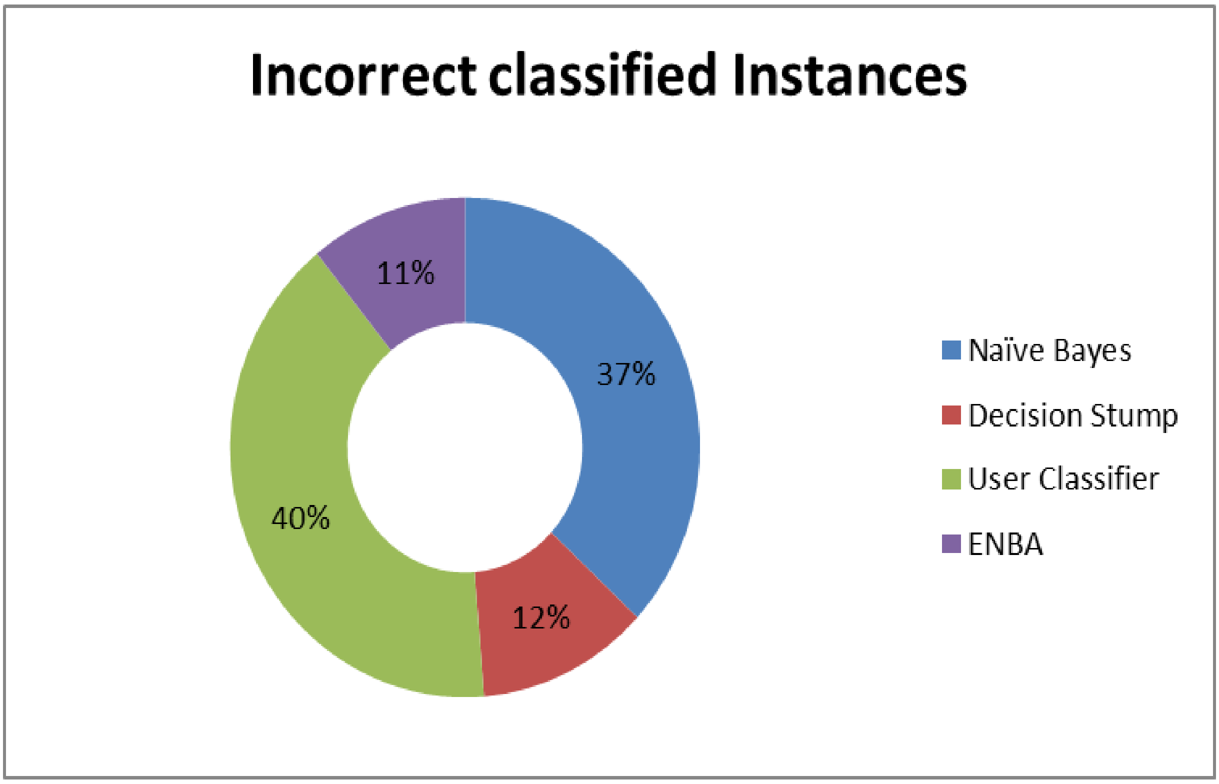


Figure 5: Shows the percentage of Incorrect Classified Instances of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

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

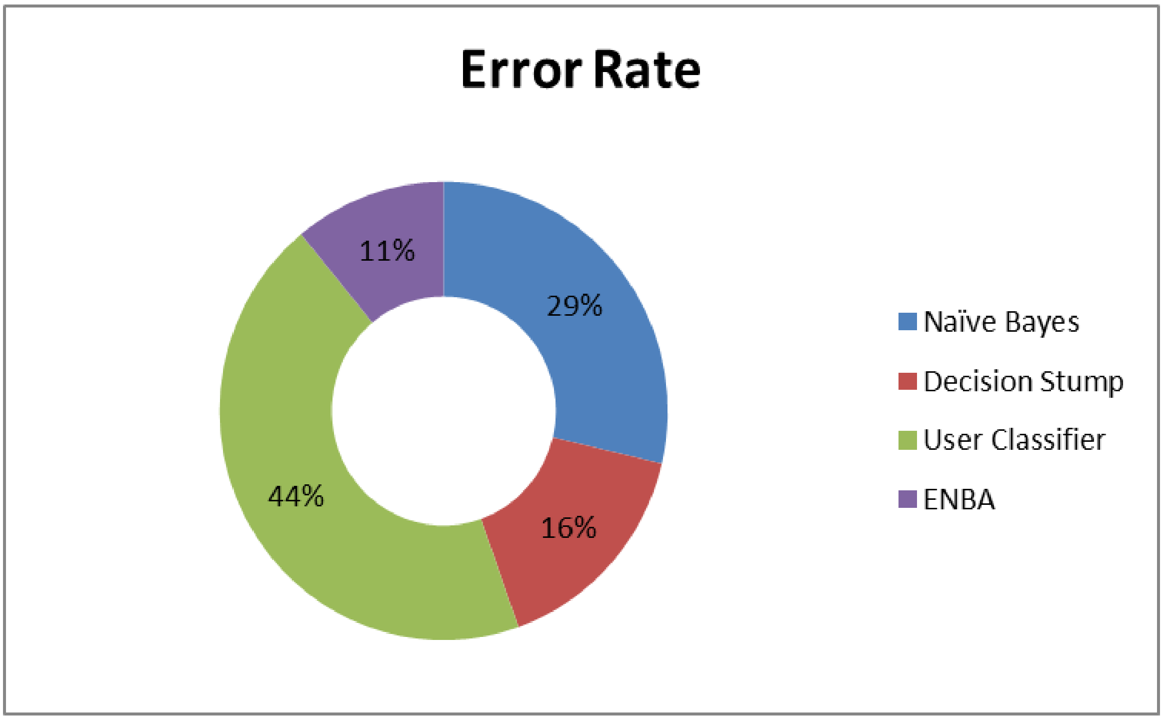


Figure 6: Shows the percentage of Accuracy of Naïve Bayes , Decision Stump , User Classifier and ENBA Algorithms

CONCLUSION AND FUTURE SCOPE

In this research, Using Enhanced Naïve Bayesian Algorithm and Decision Tree Classifiers we have addressed the prediction of Class Grading of students based on the attributes taken. Enhanced Naïve Bayesian Algorithm and Decision tree classifiers are used on student's academic data to predict the student's performance in class grading. This study will help in identifying those students who are below academic activities and shown poor, average and good performance in education system. The result shows that poor performance in grading is main factor of student's failure in final exams. The main finding of this research is the gathering of knowledge from student's academic performance and to identify those students who need special attention. This study also verifies that entropy calculated by us shows the same attribute having highest information gain which is being taken by WEKA simulation tool. Further, data mining techniques can be helpful in analysing trends in education, medical, business sector etc.

In this research, using Enhanced Naïve Bayesian Algorithm built a analytical model of student performance, using the model after pruning algorithm the model was optimized by using the model after pruning algorithm and to evaluate Correctly classified instances, Incorrect classified instances and accuracy ratio of the classification method, it achieved good the excavation results.

In future course, we will review the various classification algorithms and significance of Prediction approach in designing of efficient classification algorithms for data mining. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain.

REFERENCES

[1] Khokhoni Innocentia Mpho Ramaphosa, Tranos Zuva, 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 Hua Phung, "Towards a robust incomplete data handling approach to effective educational data classification in an academic credit system", 2014 IEEE.

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

[4] Sneha Chandra, Maneet Kaur, "Creation of an Adaptive Classifier to Enhance the Classification Accuracy of Existing Classification Algorithms 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] Nikolaos Dimokas, Nikolaos Mittas, Alexandros Nanopoulos, Lefteris Angelis, "A Prototype System for Educational Data Warehousing and Mining", Panhellenic Conference on Informatics, IEEE 2008.

[7] Ping Dong, Vice-processor,Junjun Dong, Engineer, Tiansheng Huang, Graduate student, "Application of Data Warehouse Technique in Educational Decision Support System", IEEE 2006.

[8] Ning Fang and Jingui Lu, "Work in Progress - A Decision Tree Approach to Predicting Student Performance in A High-Enrollment, High-Impact, and Core Engineering", 39th ASEE/IEEE Frontiers in Education Conference, October 18 - 21, 2009, San Antonio.

[9] Pratiyush Guleria, Niveditta Thakur, Manu Sood, "Predicting Student Performance Using Decision Tree Classifiers and Information Gain", 2014 International Conference on Parallel, Distributed and Grid Computing, IEEE

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

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

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

[13] Sajan Mathew and Dr. John T. Abraham, "APPLICATION OF DATA MINING IN HIGHER SECONDARY DIRECTORATE OF KERALA", IEEE 2016.

[14] Ms.Tismy Devasia ,Ms.Vinushree T P, Mr.Vinayak Hegde, "Prediction of Students Performance using Educational Data Mining", International Journal of Innovative Research in Science, Engineering and Technology, vol 3, Special iss 3, March 2016.

[15] Fezile Matsebula, "A BIG DATA ARCHITECTURE FOR LEARNING ANALYTICS IN HIGHER EDUCATION", IEEE Africon 2017 Proceedings.