International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Students Semester Exams Performance Prediction and Recommendation Machine

Prakruthi S T[1], Gowri Prasad[2]
[1, 2]*Department of Information Science and Engineering, New Horizon College of Engineering*

Abstract: ***Predicting students semester exams performance becomes more challenging in recent days. The existing system 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. So the proposed students performance prediction and recommendation machine is carries out analysis of student performance throughout the semester using different attributes. It has a unique set of attributes and values which is mainly designed for the engineering institutions. It learns through the data set (training data) which is given to the machine and machine splits the test data, compares the data among the training data and gives the prediction. The naïve Bayes classification algorithm is used in improve predictions. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data. The set of data divided into the five attributes start from attendance, internal score, project attributes, extra-curriculum and the class attribute performance which has the two prediction values pass or fail. Every student's semester end examination result is predicted and the system also recommends the student for further improvement.***

Keywords: Machine Learning, **Naïve** ***Bayes Classification Algorithm, Recommendation Machine, Students performance***

I. INTRODUCTION

Students performance is an essential part in higher learning institutions. This is because one of the criteria for a high quality university is based on its excellent record of academic achievements [1]. There are a lot of definitions on students' performance based on the previous literature. Usamah et al. (2013) stated that students performance can be obtained by measuring the learning assessment and co-curriculum [2]. However, most of the studies mentioned about graduation being the measure of students success
Machine learning has many algorithms such as clustering, classification, association etc for prediction of students performance. Performance analysis is the process of learning and finding the results. Several efforts provided prediction of students results based on applying classification algorithms. The major objectives of such classification may help the students and educators to improve the performance before they finish their study The machine usually predicts the performance of the student which is future result. Various attributes relevant to engineering institutions can be considered for prediction such as attendance, internal marks score, project submission status , extra-curriculur participation. After learning these data the machine is able to classify the class attribute whether the student will be passed or failed..

II. RELATED WORK

A. Yadav and Pal [7] carried out a comparative study to predict the final exam performance for engineering students. They applied ID3, C4.5 and CART decision trees algorithms on student's datasets that include personal, social, psychological and environmental factors for the prediction task. The obtained results reveal that C4.5 decision tree prediction model gives better result than ID3 and CART with accuracy of 67.77% for identifying the weaker students before the examination and that help them to improve their study for better exams results.

B. Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah[10] also carried out performance prediction. A data set of students from different countries was used. In addition to using individual machine learning methods, the researchers also applied ensemble methods, and compared the results between them. Decision trees, A classification technique provided the best results

C. Paulo Cortez and Alice Maria Gonçalves reviewed about performance prediction. It was done at the University of Minho, Portugal[11]. The data set contained information about whether the student had passed the exam in the subjects of different languages. Decision trees, random forest, neural networks, and support vector machines were used[11].

D. CH.M.H.Sai Baba1, AkhilaGovindu2 , Mani Krishna Sai Raavi3, VenkataPraneethSomisetty4[9] considered academic details like students 10th Class percentage, 12th Class Percentage, B E Previous years percentage are considered and result analysis has been done. In future, more attributes like student's social networking interest is student academic details are considered. Also based on student's social networking interests, parent's economic status, parent's educational details and many other attributes can be considered. Number of attributes should be increased because the result of the student can't be predicted just by just by his previous year's marks.

III.OBJECTIVE

1) To apply efficient prediction algorithm to find out the success rate whether the student pass or fail in semester end exams based on several attributes like internal score, attendance, project, extracurricular activities participation.

2) Recommend the student for further improvement based on the prediction made.

IV.BACKGROUND

A. CLASSIFIER USED

The Naïve Bayes classifier technique is based on Bayesian theorem, whereas it performs better when data dimensionality is high [6]. The Bayesian classifier is capable of calculating the most possible output based on the input.

In this algorithm, the presence of a particular feature in a class is unrelated to the presence of any other feature. Bayesian theorem provides an equation for calculating posterior probability P(c|x) from P(c), P(x) and P(x|c):

$$p(c \mid x) = \frac{p(x \mid c)p(c)}{P(c)}$$

- $P(c|x)$: the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).
- $P(c)$: the prior probability of *class*.
- $P(x|c)$: the likelihood, which is the probability of *predictor* given *class*.
- $P(x)$: the prior probability of *predictor*.

B. Evaluation Method

In order to evaluate the effectiveness of a prediction model, predicted values must be compared with actual values. There are multiple criteria for prediction effectiveness. Table 1 shows the possible results of prediction for binary values.

Table 1: Prediction Results

	Predicted as True	Predicted as False
Actually True	True Positive	False Negative
Actually False	False Positive	True Negative

The matrix that shows the possible prediction results is called a confusion matrix[8]. There are different evaluation criteria that can be obtained from these values. one is accuracy, defined as accuracy is basically the ratio of correct predictions. However, accuracy has limitations in evaluating the prediction performance. Especially, accuracy does not show how the cases of minority class are classified, when the class distribution is imbalanced. As an example, a data set that contains 100 students, 90 of which has passed the exam, might be considered. a crude prediction (known as the majority rule) that does not use any machine learning method, but instead predicts that every student will pass the exam, has 90% accuracy. The model should perform better than just guessing that each case belongs to the majority class.

C. Method Comparision

The first step in evaluating the results is to compare the machine learning methods in terms of their prediction performance. Table 2 show the prediction results of the three machine learning methods for the first data set, with the raw data and the modified data, respectively.

Table 2 :Method comparision of different classification algorithms

Method	Accuracy (%)	Precision (%)	Recall (%)	F-measure
Linear regression	93.3	93.5	97.7	0.952
Decision trees	93.3	91.7	100	0.956
Naïve Bayes classification	95.8	95.6	98.9	0.972

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

V. METHODOLOGY

The various modules implemented are listed. Module functionalities and implementation details are included in the following subsections. As a prototype the below modules are implemented in a simpler and cost effective way. These can be replaced in real use. Architecture of the proposed system is shown[Fig 2.1].

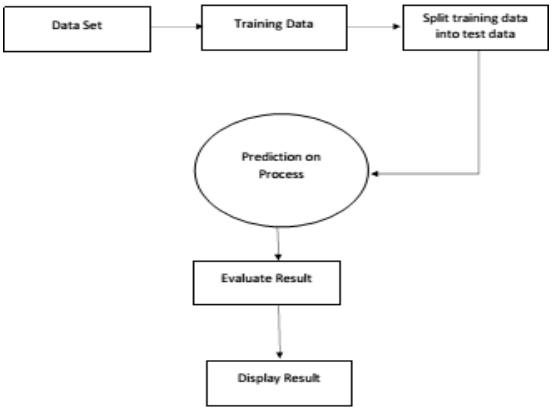

Fig 2.1: Architecture of the proposed system.

A. Data Extraction

Import all required data set and libraries. In order to work with the available datasets, we must first import the datasets into our program. So, we make use of 'pandas' package which is already available in python for data manipulation and analysis. Pandas is an open-source, BSD-licensed Python library providing high-performance easy to use data structures and data analysis tools for the Python programming language. It is built on the Numpy package and its key data structure is called the Data Frame. Data Frames allow you to store and manipulate tabular data in rows of observations and columns of variables.

```
def loadCsv(filename):
lines = csv.reader(open(filename))
dataset = list(lines)
for i in range(len(dataset)):
dataset[i] = [float(x) for x in dataset[i]]
return dataset
filename= 'data/Student.csv'
splitRatio = 0.9
dataset = loadCsv(filename)
trainingSet, testSet = splitDataset(dataset, splitRatio)
```

B. Data Preprocessing And Clustering Of Data

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Clustering is one of the main tasks in exploratory data mining and is also a technique used in statistical data analysis. Data preprocessing and clustering also involves split the input data into training and test data after extraction.[Fig: 2.3]

```
def separateByClass(dataset):
separated = {}
for i in range(len(dataset)):
vector = dataset[i]
if (vector[-1] not in separated):
separated[vector[-1]] = []
separated[vector[-1]].append(vector)
return separated
```

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

C. Running Classification Algorithm For Prediction Of Results

The important attributes used in predicting student's performance are Attendance, Internal Score, Project Submission Status, extra Curricular Activities Participation status . The set of data divided into the five attributes start from attendance has values low, medium, high respectively, same like the internal score has poor, ok, good, very good and excellent values, project attributes has two values submitted and not submitted, extra-curriculum has two values participated and not participated and the class attribute performance which has the two prediction values pass and fail

The machine can generate the prediction based on the class attributes which is giving as a performance attribute. The classification algorithm classify and compare the test data with the training data attributes and its values randomly and provides the probability which is called as a prediction. The UI part in the recommendation has given. And finally if the prediction percentage displayed into the UI

The user can interact with the UI and can predict the student performance using predict button The system provide the recommendation after the prediction if the predicted data is fail then the system displays message.

D. Prediction and Recommendation IN UI.

Naïve Bayes algorithm is implemented for classification task that have been applied to predict students performance. Naïve Bayes performs better than C4.5 decision tree algorithm in predicting the students at high risk of failing the Module with an accuracy result of 88.48%

VI.EXPERIMENTAL RESULTS

Fig 2.2: Data extraction from csv file.

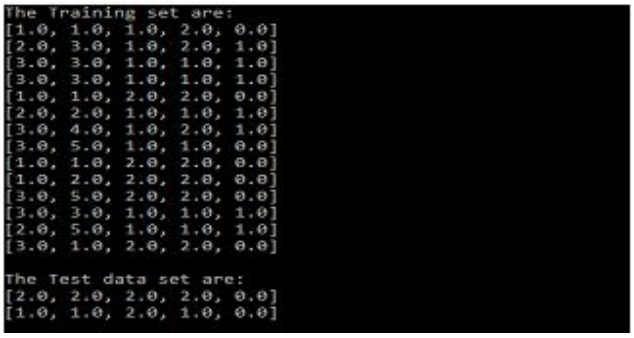

Fig 2.3:Separation of training data and test data.

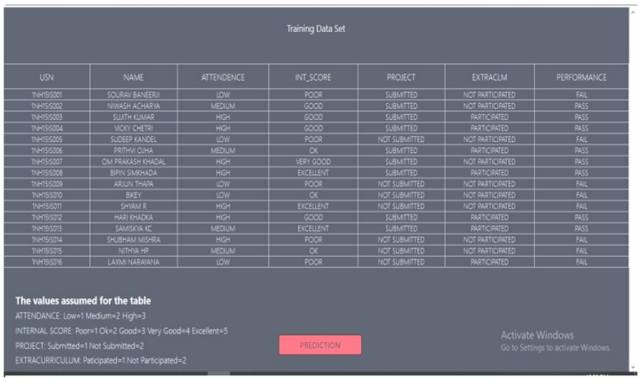

Fig 2.4:Training data with all attributes.

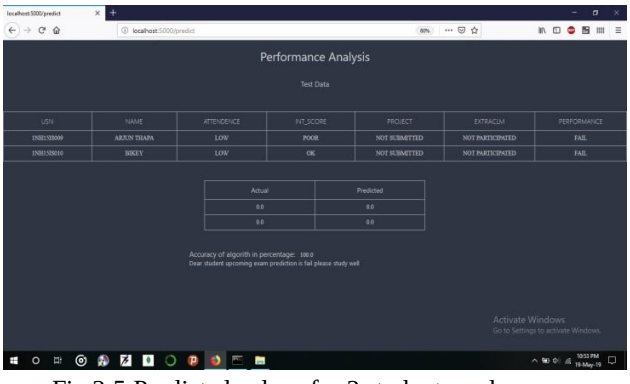

Fig 2.5:Predicted values for 2 students and accuracy.

VII. CONCLUSION

Predicting students performance is mostly helpful to help the tutors and learners improving their learning and teaching process. Result proves that Naive Bayesian algorithm provides more accuracy over other methods like Regression, Decision Tree, Neural networks etc., for comparison and prediction of semester end examination results. Also a recommendation is made for a student who has high risk of failure based on attributes like attendance, internal marks score, project submission status, extra-curriculum participation. After learning of these data, the machine could classify the class attribute of each student whether he passes or fails in semester end examinations.

VIII. ACKNOWLEDGMENT

We deeply express our sincere gratitude to our family, friends and management of New Horizon College of Engineering, Bengaluru for able guidance, providing the necessary state of art infrastructure, assistance and encouragement throughout the preparation of paper.

REFERENCES

[1] Amirah MohamedShahiriWahidahHusainNur'aini AbdulRashid, "A Review on Predicting Student's Performance Using Data Mining Techniques"(2015),URL:HTTPS://WWW.SCIENCEDIRECT.COM/SCIENCE/ARTICLE/PII/S1877050915036182
[2] Ministery of Education Malaysia, National higher education strategic plan(2015).URLhttp://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025
[3] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
[4] R. Patil, S. Salunke, M. Kalbhor and R. Lomte, "Prediction System for Student Performance Using Data Mining Classification," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4.
[5] A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naïve Bayesian," 2017 8th International Conference on Information Technology (ICIT), Amman, 2017, pp. 909-913.doi: 10.1109/ICITECH.2017.8079967
[6] S. S. Nikam, "Acomparative Study of Classification Techniques in Data Mining Algorithms," Orient. J. Comput. Sci. Technol., vol. 8, no. 1, pp. 13–19, 2015.
[7] Yadav, S., Pal, S., (2012) Data Mining: A Prediction for Performance Improvement of EngineeringStudents using Classification. World of Computer Science and Information Technology Journal(WCSIT) 2, 51–56.
[8] Tom Fawcett. 2005. An introduction to ROC analysis. Pattern Recognition Letters 27(8), 861-87
[9] CH.M.H.Sai Baba1, AkhilaGovindu2 , Mani Krishna Sai Raavi3 , VenkataPraneethSomisetty "Student Performance Analysis Using Classification Techniques" International Journal of Pure and Applied Mathematics Volume 115 No. 6 2017, 1-7ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)url: http://www.ijpam.eu\
[10] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. 2016. Mining educational data to predict student's academic performance using ensemble methods. InternationalJournal of Database Theory and Application 9(8), 119-136.
[11] Paulo Cortez and Alice Maria Gonçalves Silva. 2008. Using data mining to predict secondary school student performance. In: Proceedings of 5th Annual Future Business Technology Conference, Porto, 5-12.