International Journal of Innovative Research in Computerand Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 7, July 2020||

Student Performance Analysis For Academic Ranking Using Gradient Boosting Machine

K.Nithya[1], Dr.V.Narayani [2]

Department of Computer Science, St. Xavier's College, Palayamkottai, Tirunelveli, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science, St. Xavier's College, Palayamkottai, Tirunelveli, Tamil Nadu, India[2]

ABSTRACT: Educational organizations are unique and play utmost significant role for the development of any country. As Education transforms the lives of individuals, families, communities, societies, countries and ultimately the world! This is why we live comfortable lives today. Now a day's education is not limited to only the classroom teaching but it goes beyond that like Online Education System, Web-based Education System, Seminars, Workshops, MOOC course. becomes It's more challenging to Predict student's performance because of the huge bulks of data stored in the environments of Educational databases, Learning Management databases. Students' performance can be evaluated with the help of various available techniques. Data Mining is the most prevalent techniques to evaluate students' performance and is extensively used in Educational sector known as Educational data mining. This paper proposed an automated solution for the performance prediction of the students using machine learning. Gradient Boosting Prediction Model is proposed in this paper to improve students achievements. The main objective of this paper is to use Gradient Boosting Prediction to predict students performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data.

KEYWORDS: Student performance, Educational Data Mining; Learning Analytics model; FPSO; GA; PSO; Gradient Boosting Prediction Model

I. INTRODUCTION

The search for knowledge from large databases is known as data extraction. It detects hidden information from various data sources related to different regions. Many techniques can be used in various fields of data extraction, including weather forecasting, oil exploration, pharmaceutical business, marketing and EDM etc. [2]. A sub domain for data retrieval, called Educational data mining (EDM) was also created to extract and analyze the knowledge contained in educational data sources. Statistical data retrieval and machine learning are applied to EDM data to extract knowledge from the educational environment.

It is now in demand and receiving more attention due to the rise of educational data on the education system and even the evolution of traditional education. Starting with evolving techniques for discovering different types of data available in the learning environment, he sought to extract meaningful information to stimulate and evaluate the learning process from large amounts of raw data [6]. A study of traditional database records can provide answers to problems such as "finding students who pass the exam", while EDM provides answers to additional problems such as "predicting students who are likely to pass the exam". The arrival of an educational institution, improving the student model so that student characteristics or outcomes can be predicted is an important part of an EDM application.

Therefore, many researchers have begun to explore different data extraction techniques to help teachers or mentors evaluate and improve their course design. [7] Predicting student performance is the worst in our current education system. If student performance is anticipated, it can maintain or improve the quality of education by anticipating student interests, student-level activities, and helping to improve their performance on campus learning and educational institutes. The drop-out classification classification can also be made from this [4]. Through some institutions today, machine learning methods are implemented in conjunction with EDM, a system of continuous assessment. These schemes are useful for improving student performance. The benefit of full-time students is the main motto of the continuous assessment system. Pipeline hypocrisy and data exchange are the result of strategic outreach efforts by machine learning methods. They contribute to the presentation of data to provide active machine learning and focus on the lack of existing learning algorithms [1]. To predict student outcomes, knowledge discovery is proposed here to dig into the rules from the dataset of the study management system.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 7, July 2020||

The paper remains are prepared as follows. Section 2 reviews your work. The comparative work of student prediction are proposed in Part III. Section IV describes the experimental results. The conclusion of this study are presented in Section V.

II. LITERATURE SURVEY

Many researchers have used statistics and machine learning methods to predict student achievement in educational institutions. Edin Osmanbegovic et al. [1]uses three supervised learning algorithms, Bayesian, Decision trees and Neural Networks for evaluating data to predict the number of successful students in the course, and the effectiveness of teaching methods are evaluated based on their predictions. Precision and ease of training and easy-to-use features. It has been shown that this approach can be used to benefit students and teachers to achieve student success and reduce failure rates by taking appropriate action at the right time to improve the quality of education.

Mladen Dragicevic et al. [2] Use the decision tree classification to estimate student outcomes based on GPA criteria. Two parameters are considered as degree analysis and GPA, and this GPA provides better results in predicting student outcomes. CH.M.H.Sai Baba, et al. [3] used the decision tree and multilateral regression analysis to predict the number of students employed. Behrouz MinaeiBidgoli Iet al. [4] Compare four different rankers and combine the results into a higher ranking group. Their research divides data into three different classes: high, medium, and low. Genetic algorithms were used to improve prediction accuracy in all classes. For less functional data sets, the feature weighing mechanism is better than the function selection. With the help of L-CAPA, the results of the e-learning platform were validated.

In addition, in research work Rahel Bekele et al. [5] Implement a biennial educational network to predict student outcomes. The model was also tested on real-world data in which students were assigned to complete data, and real-world data were analyzed to predict their effectiveness. This shows that the results are very important for teachers to help students improve their learning outcomes. Paulo Cortes et al. [6] Focused on predicting high school students' results in two subjects, mathematics and Portuguese, using previous scores in the previous section and other demographic factors. Business analysis (BI) and data extraction techniques such as decision trees, random forest, vector maintenance methods, and real-world data of neural networks were used. Actual data may contain information related to student assessment, social function, and school. This model was tested with and without the previous semester.

This model demonstrates the accuracy of good predictions by Surjeet Kumar Yadav et al. [7] Conclude that decision trees are very popular because they make classification rules easier to understand than other classification methods. Commonly used decision tree classifiers are studied and experiments are conducted to find the best classifier for student data that predicts student results in exams at the end of the semester. Experimental results show that the tree divides and regresses (DTT). Kathy is the best algorithm for classifying data. It provides an accuracy of 56.25% compared to other algorithms such as ID3 and C4.5. This study is useful for failing students and is used to identify poor students who need more attention.

From a study conducted by Zill Koko Jie. Kovacic [8] proposed a case study to extract educational data to determine the amount of data that could be used to predict student success. Two CHAID and CART algorithms were applied to student enrollment data in New Zealand's newly opened multimedia technology system to achieve two decisions: classify successful and unsuccessful students. The results show that the accuracy obtained with CHAID and CART is 59.4 and 60.5, respectively. Ahmed et al. Ahmed et al. [9] Use classification techniques to predict students' final score. It was created using the Decision Tree method. To determine the optimal attribute for a specific node in the tree, measures to increase the information above, the collected sample S is used. The Midtrem attribute gets the highest, so it is selected as the root node for the decision tree and follows the same process for the entire attribute classification.

Approximately S et al. [10] showed that Data Processing Capacity (DMT) provides an effective tool to improve student performance. In addition, the study showed how data retrieval is useful in higher education, especially for predicting student outcomes. Researchers collected data from students using questionnaires to find out the relationship between a student's behaviour and his or her study results. Data extraction techniques were applied. They get a model of prediction using the decision tree as well as apply the rules in the Support Vector Machine (SVM) algorithm to predict the final score of the students. Also, students were grouped using A-median core clusters. The study showed a strong correlation between students' mental state and the results of their final study.

Ogunde A. O, et al. [11] used Iterative Dichotomiser (ID3) decision tree algorithm to predict student score in Nigeria. 79.556% of the forecast accuracy was obtained from the sample. They recommended different model trees for the solution to perform similar analysis with the expanded data set for better results. DorinaKabakchieva et al. [12] Proposed a classification method for predicting student outcomes. This document compares different data extraction algorithms using WEKA tools and the results are likely to vary between 52% -67%. For analysis, sample data related to students were collected with an assessment of enrollment, number of failures in the first year etc.

Hasmaiham, et al. [13] proposed a model of conducting the study using decision trees and ambiguous genetic algorithms. In this research, various parameters such as internal class, middle school enrollment, etc. are taken into account to identify the results of undergraduate and postgraduate students. Student performance in the degree is assessed by a tree algorithm that puts students at greater risk. Unclear genetic algorithms give students more traversal by looking at students between risk and safety.

III. PROPOSED WORK

In this paper, we propose an effective system for predicting student outcomes. To do this, the work introduces effective forecasting algorithms based on FPSO, as well as various methods of machine learning. The outline of the work we offer is shown in Figure 1. The proposed work consists of four modules. They are

1. Data Preprocessing
2. Attribute Selection
3. Feature Extraction
4. Prediction Model Generation

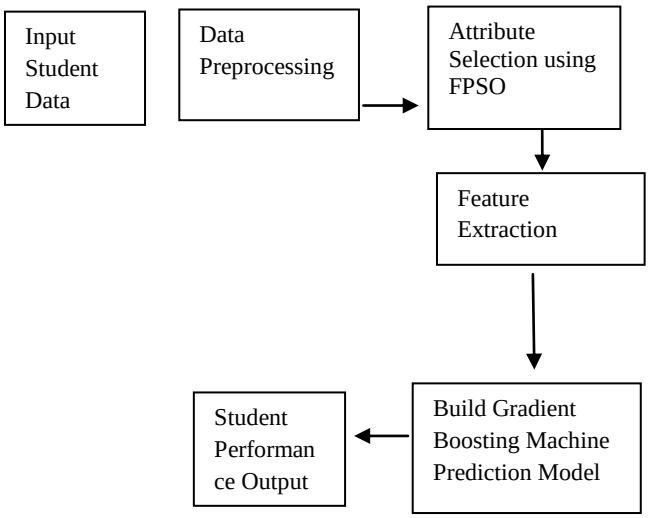


Fig.1 Outline of the Proposed Work

3.1 Data Preprocessing

Initially, data on student records were collected from the University Enterprise Database. The data is then reformatted at the data conversion stage to prepare for the following algorithm. In the data cleaning process, the parameters used in the data analysis are defined and the missing data is removed or filled in with a value of zero..

3.2 Attribute Selection

When selecting attributes, the most important attributes in the student database are selected only using the new FPSO attribute selection method.

3.2.1 Fuzzy Particle Swarm Optimization

To get good predictive results, several types of feature are performed simultaneously. Because different types of feature may contain additional information, this can lead to better predictive efficiency by selecting characteristics that discriminate between different feature spaces. The advantage of feature selection is to determine the definition of the initial feature set.

3.3 Feature Extraction

After selecting the key attributes, the next step is to create a feature matrix. In this step, three matrix characteristics are developed: grade matrix, performance matrix, interest matrix based on the students mark, performance and interests.

There are three main conditions in a referral system: user, item, and grade. The task of the recommendation is to predict which rating the user will give for all the unskilled positions, then recommend the user with the highest predicted results. Similarly, in a in the grading management system (GMS) that has three main components: students, courses, and assessment / evaluation. In this setting, the task is to anticipate marking subjects that students have not

learned. There is a similar matching between the student model in GMS and recommender systems where student, course, and mark/grade would become user, item, and rating, respectively {Student → User; Course → Item; Grading → Ratings}. As well as, similar mapping {Student → User; Course → Item; Performance → Ratings}, {Student → User; Course → Item ; Contextual Information → Ratings} and {Student → User; Course → Item; Interests → Ratings} are generated.

3.4 Prediction Model Generation

In this module, prediction models are generated using different machine learning methods. Among some machine learning methods, this work uses K Nearest Neighbour, Navie Bayes and Support Vector Machine as prediction model generation.

3.4.1 Gradient Boosting Prediction Model

Gradient boosting is a machine learning technique for regression and classification problems that creates predictive models in the form of weak model predictions, usually the decision tree. It creates step-by-step models of other reinforcement methods and summarizes them, allowing the functional optimization of any of the various boosting methods. Increasing the slope size is usually used with fixed size decision trees as a basic learner. In this particular case, Friedman suggested proposing a change in approach to increasing gradient requirements, which improved the quality of individual study adjustments at the base learner.

Algorithm: Gradient Boost Prediction Model

Inputs:

• input matrix derived from feature extraction step (k, l)
• Total iterations T
• Assume the loss-function as (l, g)
• Assume the base-learner model f(k, θ)

Algorithm: set f0 with a constant value
2: **for** i = 1 to T do
3: Calculate the negative gradient gr(k)
4: fit a new base-learner function f(k, θi)
5: Calculate the best gradient descent step-size ρi:

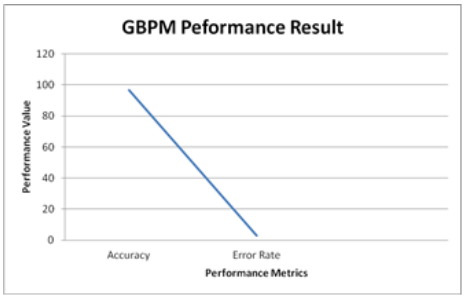
$$\rho i = \min(p)*\text{sum}(l-f(xi)+ \rho f(ki, \theta i))$$

6: update the prediction estimate value:

$$fi= fi-1+ \rho f(ki, \theta i))$$

7: **end for**

The above the algorithm take values of the training dataset in order to generate the prediction model. The goal of this prediction model is to compare the prediction capability of student performance considering two different sets of variables from training and testing dataset. This prediction model was created for testing dataset, enabling the evaluation and comparison of its performances with training dataset and accurately predicting whether a student will pass/fail at the end of the year. This pass/fail is considered as the final predicted result of the particular student. This algorithm is implemented using R.3.3.2 programming language. The result of the GBPM with 2018 year dataset is shown in below figure.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 7, July 2020||

IV. RESULT AND ANALYSIS

4.1 Dataset Used

Datasets of student can be collected from Xavier's College, Thirunelveli. Several years of datasets are collected. But in this work 2018 year dataset are used for analysing the performance of the GBPM with FPSO.

This dataset consists of 2861 student records; each record consists of 24 attributes with their domain values. The dataset was divided two parts, training dataset (75%) and testing dataset (25%). This work is implemented using R.3.3.2 programming language.

4.2 Efficiency Parameters

To assess the efficiency of the proposed sentiment constructing process, several efficiency metrics are available. This paper employs the Detection Accuracy and Error Rate to analyses the efficiency.

4.2.1. Detection Accuracy

Detection Accuracy is the measurement system, which measure the degree of closeness of measurement between the original results and the correctly prediction results.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

4.2.2. Error Rate

Error Rate is the measurement system, which measure no of falsely predicted result from given input data.

$$\text{Error Rate} = (FP + FN)/(TP+FP+TN+FN) \quad (2)$$

4.3 Experimental Results

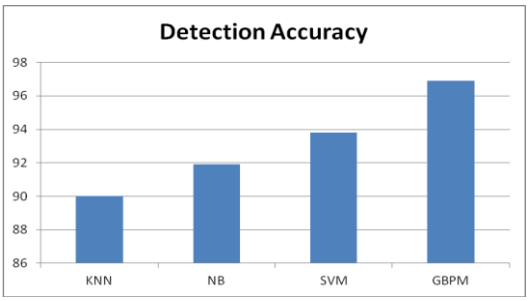
4.3.1 Experiment No #1 : Accuracy Analysis of Proposed Prediction Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy and Error rate measures are employed Table 1 lists the accuracy analysis of FPSO with GBPM.

Table 1: Detection Accuracy Analysis of Proposed Prediction Model

Prediction Model	Detection Accuracy
KNN	90
NB	91.9
SVM	93.8
GBPM	96.92

As observed from Table 1, the Accuracy of the FPSO with GBPM in range 93-97, which is superior than other methods. So the FPSO with GBPM classifier is considered to be the best for sentiment analysis.



As observed from above figure, the Accuracy of the FPSO with GBPM in range 93-97, which is superior than other method. So the FPSO with GBPM classifier are best for sentiment analysis.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 7, July 2020||

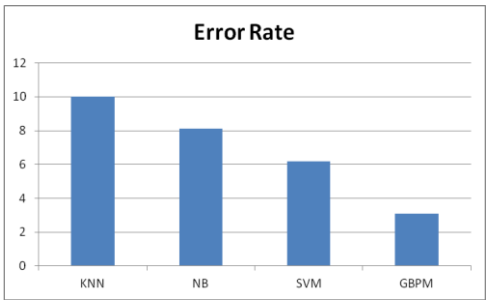
4.3.2 Experiment No #2 : Error Rate Analysis of Proposed Prediction Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy and Error rate measures are employed. Table 2 lists the Error Rate analysis of FPSO with GBPM.

Table 2: Error Rate Analysis of Proposed Prediction Model

Prediction Model	Error Rate
KNN	10
NB	8.1
SVM	6.2
GBPM	3.08

As observed from Table 2, the error rate of the FPSO with GBPM in range 3-10, which is lower than other method. So the FPSO with GBPM classifier is considered to be the best for sentiment analysis.



As observed from above figure, the error rate of the FPSO with GBPM in range 3-10, which is lower than other method. So the FPSO with GBPM classifier are best for sentiment creation.

V. CONCLUSION

Performance of student's using EDM is carried out in this research work. Classification is done in order to predict students in different class categories like High, medium and low. This paper has compared various machine learning approaches and feature selection approaches on predicting students performance with various analytical methods. Classification is done in order to predict students in different class categories like High, medium and low. The results of both feature selection approaches and machine learning were compared on the basis of accuracy and precision. It was found and detected that classification implemented by GBPM with FPSO is more efficient compare to other classifiers as seen in the accuracy and precision. Based on the results, GBPM with FPSO technique is more efficient compared to other technique in prediction of students' performance.

REFERENCES

1. Osmanbegovic E. and Suljic M., "Data mining approach for predicting student performance",Journal of Economics and Business,Vol. X, Issue 1, 2012.
2. MladenDragicevic, Mirjana Pejic Bach, and VanjaSimicevic, "Improving University Operations with Data Mining: Predicting Student Performance", International Journal of Social, Behavioral, Educational, Economic and Management Engineering Vol. 8, Issue 4, 2014.
3. CH.M.H.Sai Baba, AkhilaGovindu,Mani Krishna Sai Raavi, and VenkataPraneethSomisetty, "Student Performance Analysis Using Classification Techniques", International Journal of Pure and Applied Mathematics, Vol. 115, No. 6, pp. 1-7, 2017.
4. Behrouz M, Karshy D, Korlemeyer G and Punch W., "Predicting student performance: an application of data Mining methods with the educational web-based system", IEEE Frontiers in Education Conference,2003.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 7, July 2020||

5. BekeleR. andMenzelW. “A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students”, Journal of Information Science 2016.
6. Cortez P and Silva A., “Using data mining to predict Secondary school student performance”, Journal of information science, Vol. 2, issue 6, 2013.
7. Surjeet K, Yadav, Bharadwaj B and Pal B., “Data Mining Applications: A comparative Study for Predicting Student’s performance”, International journal of innovative technology & creative engineering, Vol. 1, issue 12,2012.
8. Kovacic Z., “Early prediction of student success: Mining student enrollment data”, Informing Science & IT Education Conference,pp. 647-665, 2010.
9. Ahmed A. B. E and Ibrahim S. E.,“Data Mining: A prediction for Student's Performance Using Classification Method”, World Journal of Computer Application and Technology, Vol. 2, issue 2, pp. 43- 47,2014.
10. Sembiring S, Zarlis M, Hartama D., Ramliana S andElvi W., “Prediction of student academic performance by an application of data mining techniques.”, International Conference on Management and Artificial Intelligence, Vol.6, pp. 110-114,2011.
11. Ogunde A.O. andAjibade D.A., “A data Mining System for Predicting University Students Graduation Grade Using ID3 Decision Tree approach”, Journal of Computer Science and Information Technology, Vol. 2, No 1, pp. 01-26,2014.
12. DorinaKabakchieva, “Predicting Student Performance by Using Data Mining Methods for Classification”, Cybernetics and Information Technologies, Vol. 13, No 1, pp. 61-72, 2013.
13. HashmiaHamsa, SimiIndiradevi and Jubilant J. K., “Academic Performance Model Using Decision Tree and Fuzzy Genetic Algorithm”,International Conference on Recent Advancement and Effectual Researches in Engineering, Science and Technology, pp. 326-332, 2016.