IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

Prediction of Engineering Students Learning Outcomes using Machine Learning

Dr. H Girisha, Irfan T, M. Ashish Narayana, Mohammed Mushraf Ashrafi, Syeda Sana Zeba Quadri
Department of Computer Science and Engineering
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract: *Data science and machine learning, over the years have proven very well-organized and significant in many sectors including education. Machine learning is an aspect of artificial intelligence in which a computing system can able to learn from data and make conclusions. The recent development in education sector provides assessment tools to predict the student performance by exploring education data using machine learning and data mining techniques. Student performance assessment is an important measurement metrics in education which affects the university accreditation. Student performance improvement plan must be implemented in those universities, by counselling the low performer students. It helps both students and teachers to overcome the problems experienced by the student during studies and teaching techniques of teachers. In this review paper, different student performance prediction literature related to find out low performer student. The survey results indicated that different machine learning techniques are used to overcome the problems related to predicting student at risk and assessment of student performance. Machine learning techniques plays an important role in progress and prediction of student performance, thus improving student performance prediction system.*

Keywords: Student Performance Prediction (SPP), Artificial Intelligence (AI), Machine Learning (ML)

I. INTRODUCTION

Student performance prediction is one of the indispensable challenges in education sec- tor. There are different factors which may affect the student performance and which indirectly affect the university accreditation. Maintaining high learning rate in universities can be challenging due to low performer students. The wide variety of research has open to improve the learning system based on the student needs. Challenging issue in education data mining provide new opportunities of the research. Data mining techniques that applied in education field can be called as educational data mining. It is the process of automatically determining useful information inside from raw data. Machine learning techniques in education sector, examined to control meaningful patterns that improve student knowledge and academic institutions will take the decision from those patterns.

Modern Educational Institute operates in a very competitive and complex environment. Does evaluating student performance, providing high level teaching and learning techniques to predict student performance and identifying the objectives of research are some challenging issues faced by most universities today. Student performance improvement planes are implemented in universities to overcome student problem during their studies. Student performance prediction at entry level and during the academic year helps the organization to develop and assess the involvement plants, where both management such as teacher and student are the beneficiaries of the student performance prediction plan.

This problem statement aims to address the pressing need for a predictive framework that goes beyond traditional assessment methods, enabling educational institutions to proactively support students on their academic journey. By leveraging machine learning, the goal is to empower educators with the tools necessary to foster a more personalized and effective learning environment, ultimately enhancing the success and well-being of students.

II. LITERATURE SURVEY

There are many papers published related to students' performance prediction system, which is beneficial for educational sector. By using educational data and different techniques the students result is evaluated. The summery of those papers

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

on student performance prediction which has been published yet using various machine learningalgorithms are shown in Table 1.

Topic	Attributes	Algorithm	Dataset/Size	Conclusion
Classification and prediction of student performance data using various machine learning algorithms	18 Attributes	SCM, Naïve Bayes, C-4.5, ID3	UCI Machinory Student Performance (649 instanes)	SVM is better
Students' performance prediction using KNN and Naïve Bayesian	DOB, Gender, City, Secondary School Name, Specialization, Fathers Job, Student Status	KNN, Naïve Bayes	Gaza Strip 2015 (500 instances)	Naïve Bayes is better than KNN
Student performance analysis system (SPAS)	Quizzes, Assignments, projects, Final examination Grades, Gender, Programs	J-48, Simple CART, BFTree, Random Tree, J-48 Garft	TMC1013 System Analysis and Design in University of Malaysia Sarawak	BF-Tree is better, TMC-1013 System Analysis assist to predict the student performance accurately
Analysis and Prediction of Student Academic Performance Using Machine Learning	22 Attributes	1. Linear Regression for Supervised Learning 2. Linear Regression for Deep Learning	Kaggle Dataset (648 instances)	Liner Regression for Supervised learning is best.
A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques	MEDU, Attendance, 1st Sem GPA	Naïve Bayes, Decision Tree	Industrial Engineering University, Islam, Indonesia	Naïve Bayes is performing better.
Prediction of Students Performance using Educational Data Mining	19 Attributes	Naïve Bayes	Amrita School of Arts & Science, Mysure	Naïve Bayes is used for knowledge classification

III. PROPOSED SYSTEMS

In the proposed system of "Prediction of Engineering Students Learning Outcomes using Machine Learning," the primary objective is to leverage advanced machine learning techniques to forecast the learning outcomes of engineering students. This innovative approach aims to enhance educational practices by providing insights into students' academic performance, thus enabling educators to intervene promptly and tailor educational strategies to individual needs effectively. Below is a detailed description of the proposed system:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

A. Background and Motivation:

The traditional educational system often relies on retrospective analysis and standardized testing to evaluate students' progress. However, this approach may overlook the unique learning patterns and challenges faced by individual students.

With the advent of machine learning algorithms and the availability of vast educational datasets, there is an opportunity to develop predictive models that can anticipate students' learning outcomes based on various factors such as demographic information, past academic performance, and engagement metrics.

By accurately predicting students' learning outcomes, educators can personalize the learning experience, identify at-risk students early, and implement targeted interventions to improve overall academic success rates.

B. Objectives:

Develop a predictive model capable of forecasting engineering students' learning outcomes with high accuracy.

Identify relevant predictors and features that significantly influence students' academic performance.

Evaluate the performance of the predictive model using real-world educational datasets to ensure its reliability and effectiveness.

Provide actionable insights to educators, administrators, and policymakers to optimize educational practices and enhance student success rates.

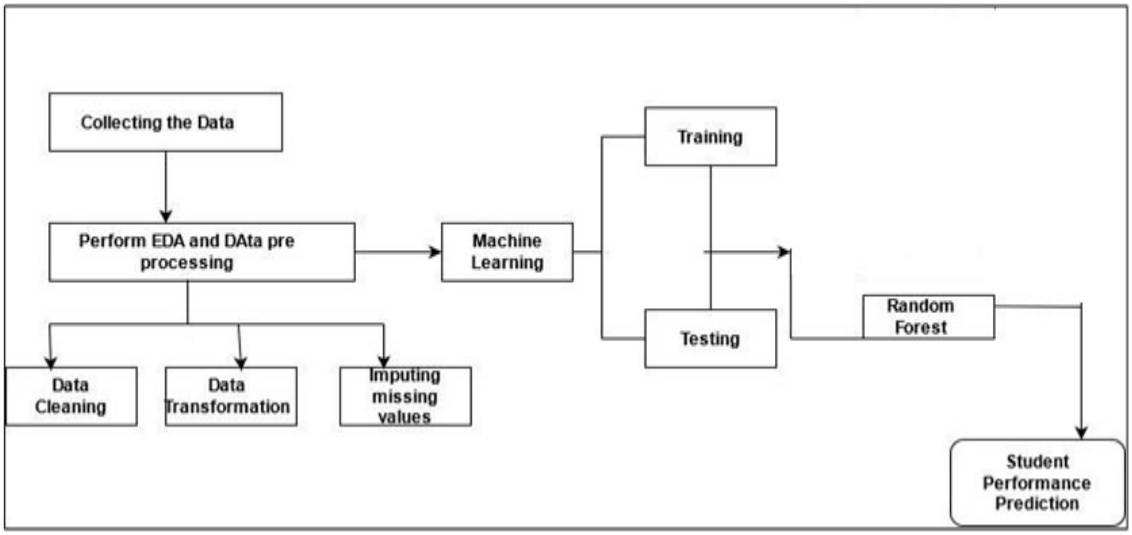
C. Proposed Methodology:

The proposed methodology involves the application of machine learning algorithms, specifically the Random Forest model, to predict engineering students' learning outcomes.

Random Forest is a powerful ensemble learning technique that combines multiple decision trees to generate robust and accurate predictions. Unlike traditional linear regression models, Random Forest can capture complex nonlinear relationships and interactions among predictor variables, making it well-suited for educational prediction tasks.

By leveraging Random Forest, the proposed system can handle high-dimensional datasets with heterogeneous features, handle missing data effectively, and mitigate overfitting issues commonly encountered in linear regression models.

Additionally, Random Forest inherently provides feature importance scores, allowing educators to identify the most influential factors driving students' academic performance, thus facilitating data-driven decision-making and targeted interventions.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

IV. METHODOLOGY SECTION

In this project, the chosen method for predicting engineering students' learning outcomes is the Random Forest model, a robust and versatile machine learning algorithm. Here's a detailed description of why Random Forest is selected and its advantages over traditional linear regression models:

A. Power of Ensemble Learning:

Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. Each tree in the forest independently learns from a subset of the training data and contributes to the final prediction.

By aggregating the predictions of multiple trees, Random Forest reduces the risk of overfitting and improves generalization performance, making it particularly suitable for complex prediction tasks with high-dimensional datasets.

B. Nonlinear Relationships:

Unlike linear regression models, which assume linear relationships between predictors and the target variable, Random Forest can capture complex nonlinear relationships and interactions among features.

In educational prediction tasks, where the relationship between student attributes and learning outcomes may be nonlinear and nontrivial, Random Forest offers a more flexible modeling approach, potentially leading to more accurate predictions.

C. Robustness to Outliers and Noise:

Random Forest is inherently robust to outliers and noisy data due to its ensemble nature. Outliers or erroneous data points are less likely to significantly impact the overall prediction, enhancing the model's resilience to data imperfections.

This robustness is particularly beneficial in educational datasets, which may contain outliers or measurement errors, ensuring reliable predictions even in the presence of noisy data.

D. Handling of Heterogeneous Features:

Educational datasets often comprise diverse types of features, including categorical variables, numerical attributes, and text data. Random Forest can effectively handle heterogeneous feature types without requiring extensive preprocessing.

By accommodating various types of features, Random Forest simplifies the modeling process and reduces the need for feature engineering, allowing researchers to focus more on model development and evaluation.

E. Feature Importance Analysis:

One of the key advantages of Random Forest is its ability to provide feature importance scores, indicating the relative contribution of each predictor variable to the model's predictions.

By analyzing feature importance, educators and researchers can gain insights into the factors driving students' learning outcomes, identify influential variables, and prioritize intervention strategies accordingly.

V. RESULTS

```
"C:\Users\ashis\PycharmProjects\Student Perfomance prediction\pythonProject\venv\Scripts\python.exe" "C:\Users\ashis\PycharmProjects\Student Pe
Enter the USN of the candidate: 3VC20CS602

Candidate Details:
 Name       DOB Department  Attendance
Irfan 2301-08-25       CSE         65

Rank of this student: 9
Area of improvement: IA 1

Overall Predicted Performance (0-100): 40.0
Enter the actual performance: 90

Difference between Actual and Predicted Performance: 50.0
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

VI. CONCLUSION

The Random Forest model offers several advantages over traditional linear regression models, including enhanced predictive performance, robustness to outliers and noise, flexibility in modeling nonlinear relationships, and the ability to handle heterogeneous features effectively. By leveraging these strengths, the proposed system aims to deliver accurate and actionable predictions of engineering students' learning outcomes, ultimately enhancing educational practices and student success rates.

REFERENCES

[1]. Harikumar Pallathadka, Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz, KhongdetPhasinam, "Classification And Prediction Of Student Perfor- mance Data Using Various Machine Learning Algorithms", Materials Today: Proceedings Elsevier, 2021.

[2]. Ihsan A. Abu Amra, Ashraf Y. A. Maghari, "Students Performance Prediction Using KNN And Naïve Bayesian", 8th International Conference on Information Technology (ICIT), 2017.

[3]. Dr. R Senthil Kumar, Jithin Kumar.K.P, "Analysis Of Student Performance Based On Clas- sification And Mapreduce Approach In Bigdata", International Journal of Pure and Applied Mathematics, Volume 118 No. 14, 141–148, 2018.

[4]. Emmy Hossain, Mohammad Hossin, "Student Performance Analysis System (SPAS)", https:// www.researchgate.net/publication/282956807, 2015.

[5]. Shanmugarajeshwari, R. Lawrance, "Analysis of Students' Performance Evaluation Using Classification Techniques", IEEE, 2016.

[6]. Leila Ismail, HunedMaterwala, Alain Hennebelle, "Comparative Analysis Of Machine Learning Models For Students' Performance Prediction", https://www.researchgate.net/pub lication/350057919, 2021.

[7]. Ajibola Oyedeji, Olaolu Folorunsho, Olatilewa Raphael Abolade, "Analysis And Prediction Of Student Academic Performance Using Machine Learning", https://www.researchgate.net/ publication/340310208, 2020.

[8]. Annisa UswatunKhasanah, Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques", IOP Conference Series: Materials Science and Engineering, 2017.

[9]. Leena H. Alamri, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, Nida Aslam, "Predicting Student Academic Performance Using Support Vector Machine and Random Forest", https://www.researchgate.net/publication/351653053, 2020.

[10]. Ms.TismyDevasia,Ms.VinushreeTP,Mr.VinayakHegde,"PredictionOfStudentsPerfor- mance Using Educational Data Mining", ResearchGate, 2020,

[11]. Snje˘zanaKri˘zanic, "Educational Data Mining Using Cluster Analysis And Decision Tree Technique:ACaseStudy",InternationalJournalofEngineeringBusinessManagement,2020.

[12]. C.A.Palacios,J.A.Reyes-Suárez,L.A.Bearzotti,V.Leiva,andC.Marchant,"Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile", Entropy, vol. 23, no. 4, pp. 1–23, 2021.

[13]. S. Rovira, E. Puertas, and L. Igual, "Data-driven system to predict academic grades and dropout", PLoS One, vol. 12, no. 2, Feb. 2017.

[14]. T.Muncie,"UsingMachineLearningModelstoPredictStudentRetention:BuildingAState- Wide Early Warning System", Morehead State University, 2020.

[15]. G.M.RaoandP.K.K.Kumar,"StudentsPerformancePredictioninOnlineCoursesUsing Machine Learning Algorithms," United Int. J. Res. Technol., vol. 02, no. 11, 2021.

[16]. Hassan Zeineddine and Udo C. Braendle, "Enhancing Prediction of Student Success: Automated Machine Learning Approch", Article in Computers & Electrical Engineering, 2021.

[17]. Fiseha Berhanu, "Students Performance Prediction Based on Their Academic Record", International Journal of Computer Application, 2015.

[18]. M. Tight, "Student Retention and Engagement in Higher Education", Journal of Further & Higher Education, 2019.