Machine Learning-Based Academic Result Prediction System

Megha Bhushan, School of Computing, DIT University, Dehradun, India*

https://orcid.org/0000-0003-4309-875X

Utkarsh Verma, School of Computing, DIT University, Dehradun, India

Chetna Garg, School of Computing, DIT University, Dehradun, India

Arun Negi, Deloitte USI, Gurgaon, India

ABSTRACT

Students' academic performance is a critical issue as it decides his/her career. It is pivotal for the educational institutes to track the performance record because it can help to enhance the standard of their quality education. Thus, the role of the academic result prediction system comes into existence which uses semester grade point average (SGPA) as a metric. The proposed work aims to create a model that can forecast the SGPA of students based on certain traits. It predicts the result in the form of SGPA of computer science students considering their past academic performance, study, and personal habits during their academic semester using different machine learning models, and to compare them based on different accuracy parameters. Some models that are widely used and are found effective in this field are regression algorithms, classification algorithms, and deep learning techniques. The results conclude that deep learning techniques are the most effective in the proposed work because of their high accuracy and performance, depending upon the attributes used in the prediction.

KEYWORDS

Academic performance, Deep Learning, Educational Data Mining, Machine Learning, Semester Grade Point Average (SGPA)

INTRODUCTION

Education is a crucial aspect in terms of the economy, due to which researchers are developing several methods to enhance the performance of the students. One way to do so is to track the student's performance. Through research and development in this field, students can be benefitted in numerous ways, like faculty can give special attention to the students whose predicted semester grade point average (SGPA) is low or not up to the mark and this will be very helpful for the student and for the university as well as their entire result percentage will improve, on the other hand students can also track their performance and hence, can improve their study pattern and accordingly perform well.

Educational data mining is an emerging discipline, used to explore the distinctive and increasingly sizable data gathered from various educational institutes, and using data mining techniques to

*Corresponding Author

comprehend the students and the methods in which they learn. By exploring these huge datasets and using different aforementioned techniques, unique patterns can be identified which will help to study, predict, and improve academic performance of students.

COVID-19 outbreak has brought unique challenges that were not expected earlier, especially in the education system. Due to the pandemic, schools, colleges and universities were closed. Therefore, few universities decided to take the exams in online mode, while some went for promoting the students to next semester by providing them grades based on the assignment marks and the marks obtained by the students in the previous semester only. This cannot be the only factor for the evaluation since there are other personal factors that contribute to a student's result.

Secondly, it was difficult for faculty to identify the students who are at a risk of not performing well due to following variation in students' performance, (i) who perform well in the beginning but degrade their performance by the end., (ii) who perform worse in the beginning but improve their performance by the end, and (iii) who consistently perform either better or worse. Predicting their SGPA can be very helpful to identify the students who need more attention and work hard.

A model is proposed in this work in which the results of students will not be dependent on a single feature, rather the features will include day-to-day personal habits along with academic habits and past academic performance, to predict the future result as fair as possible. Various machine learning (ML) (Kedia & Bhushan, 2022; Kholiya et al., 2021; Singh & Bhushan, 2022; Verma et al., 2019) techniques have been used for predicting the result of the student. It is crucial to assess the data quality (Bhushan & Goel, 2016; Bhushan et al., 2018; Bhushan et al., 2021) used by ML algorithms. Predicting the SGPA will also help the student in planning his/her academic goals and accordingly put in efforts to improve the results. It will also help the faculty as well as the university to maintain a record of responses submitted by the student and hence, to better understand and supervise their students accordingly. The objective of this work is to ensure that each student is being monitored and being given the guidance he or she needs in order to improve their academics.

The remaining paper incorporates related literature review followed by the experiment details along with the results and discussions. Later, the conclusion along with the future scope is discussed.

RELATED WORK

There are many existing works related to the prediction of students' performance using ML algorithms (Arcinas, 2022; Albreiki et al., 2021; Chakrapani & D, 2022; Gajwani & Chakraborty, 2021; Verma et al., 2022; Yağcı, 2022).

According to Sharma and Aggarwal (2021), a dataset of around 400 students was selected and the analysis was conducted to check the level of parental influence on academic performance. The attributes taken into account for prediction were family size, parents' education, educational support from the family, internet access at home, paid classes and semester wise marks. During the analysis, correlations were found between the attributes, and linear regression (LR) was used as the predictive model. Further, training was done on 90% of the data and accuracy was calculated in terms of mean absolute error (MAE) as well as root mean square error (RMSE) which were 3.155 and 3.76, respectively. Gradient boost and support vector machine (SVM) were also applied but LR proved to be the best with a value of adjusted R square as 0.4771.

The analysis of students' performance based on a subset of behavioral and academic parameters was done using the techniques of feature selection and supervised ML algorithms (logistic regression (LOGR), decision tree (DT), naïve bayes' (NB) classifier and ensemble ML algorithms like bagging and boosting) (Gajwani & Chakraborty, 2021). The attributes such as demographic, behavioral and academic were taken into consideration. These included nationality, gender, place of birth, student's participation in discussion groups, raising hand in classes, using external resources, grade as well as semester marks. The dataset of 500 records was taken from Kaggle which was in turn obtained from learning management system (LMS). Further, 70% records were used in training and 30% for

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/machine-learning-based-academic-result-prediction-system/334715

Related Content

Software Principles of 5G Coverage: Simulator Analysis of Various Parameters

Himanshu Kumar Sinha, Devasis Pradhan, Abhishek Saurabhand Anand Kumar (2023). *The Software Principles of Design for Data Modeling (pp. 276-285).*

www.irma-international.org/chapter/software-principles-of-5g-coverage/330502

Content-Based Image Retrieval Using Hybrid Densenet121-Bilstm and Harris Hawks Optimization Algorithm

Sanjeevaiah K., Tatireddy Subba Reddy, Sajja Karthik, Mahesh Kumarand Vivek D. (2023). *International Journal of Software Innovation (pp. 1-15).*

www.irma-international.org/article/content-based-image-retrieval-using-hybrid-densenet121-bilstm-and-harris-hawks-optimization-algorithm/315661

Determination of Melting Point of Chemical Substances Using Image Differencing Method

Anurag Shrivastavaand Rama Sushil (2022). *International Journal of Software Innovation (pp. 1-10).*

www.irma-international.org/article/determination-of-melting-point-of-chemical-substances-using-image-differencing-method/297985

Project Quality of Off-Shore Virtual Teams Engaged in Software Requirements Analysis: An Exploratory Comparative Study

Dhruv Nath, Varadharajan Sridhar, Monica Adyaand Amit Malik (2009). *Software Applications: Concepts, Methodologies, Tools, and Applications (pp. 2115-2136).*

www.irma-international.org/chapter/project-quality-off-shore-virtual/29498

Rapid Agile Transformation at a Large IT Organization

Pan-Wei Ng (2016). *Emerging Innovations in Agile Software Development (pp. 85-102).*

www.irma-international.org/chapter/rapid-agile-transformation-at-a-large-it-organization/145035