Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor:8.075(SJIF) Vol-5, Issue- 2,2025 ISSN:2582-5887 www.uijes.com

A REVIEW ON MACHINE LEARNING APPROACH FOR TRACKING AND PREDICTING STUDENT PERFORMANCE

N.S.SAI KARUN [1] B.VIJAYA SRI [2]
1 C.S.E Student, Dept of C.S.E, RK College Of Engineering, Kethanakonda, A.P., India
2 Assistant Professor, Dept of C.S.E, RK College Of Engineering, Kethanakonda, A.P., India

ABSTRACT
Accurately predicting students' future performance based on their ongoing academic records is crucial for effectively carrying out necessary pedagogical interventions to ensure students' on-time and satisfactory graduation. Although there is a rich literature on predicting student performance when solving problems or studying for courses using data-driven approaches, predicting student performance in completing degrees (e.g. college programs) is much less studied and faces new challenges: (1) Students differ tremendously in terms of backgrounds and selected courses; (2) Courses are not equally informative for making accurate predictions; (3) Students' evolving progress needs to be incorporated into the prediction. In this paper, we develop a novel machine learning method for predicting student performance in degree programs that is able to address these key challenges. The proposed method has two major features. First, a bilayered structure comprising of multiple base predictors and a cascade of ensemble predictors is developed for making predictions based on students' evolving performance states. Second, a data-driven approach based on latent factor models and probabilistic matrix factorization is proposed to discover course relevance, which is important for constructing efficient base predictors. Through extensive simulations on an undergraduate student dataset collected over three years at UCLA, we show that the proposed method achieves superior performance to benchmark approaches.

Keywords: Student Performance Prediction,Machine Learning,Degree Completion,Ensemble Predictors,Probabilistic Matrix Factorization

INTRODUCTION

Making higher education affordable has a significant impact on ensuring the nation's economic prosperity and represents a central focus of the government when making education policies [1]. Yet student loan debt in the United States has blown past the trillion-dollar mark, exceeding Americans' combined credit card and auto loan debts [2]. As the cost in college education (tuitions, fees and living expenses) has skyrocketed over the past few decades, prolonged graduation time has become a crucial contributing factor to the evergrowing student loan debt. In fact, recent studies show that only 50 of the more than 580 public four-year institutions in the United States have on-time graduation rates at or above 50 percent for their full-time students [2].

To make college more affordable, it is thus crucial to ensure that many more students graduate on time through early interventions on students whose performance will be unlikely to meet the graduation criteria of the degree program on time. A critical step towards effective intervention is to build a system that can continuously keep track of students' academic performance and accurately predict their future

performance, such as when they are likely to graduate and their estimated final GPAs, given the current progress. Although predicting student performance has been extensively studied in the literature, it was primarily studied in the contexts of solving problems in Intelligent Tutoring Systems (ITSs) [3][4][5][6], or completing courses in classroom settings or in Massive Open Online Courses (MOOC) platforms [7][8]. However, predicting student performance within a degree program (e.g. college program) is significantly different and faces new challenges.

First, students can differ tremendously in terms of backgrounds as well as their chosen areas (majors, specializations), resulting in different selected courses as well as course sequences. On the other hand, the same course can be taken by students in different areas. Since predicting student performance in a particular course relies on the student past performance in other courses, a key challenge for training an effective predictor is how to handle heterogeneous student data due to different areas and interests. In contrast, solving problems in ITSs often follow routine steps which are the same for all students [9]. Similarly, predictions of students' performance in courses are often based on in-course assessments which are designed to be the same for all students [7]

Second, students may take many courses but not all courses are equally informative for predicting students' future performance. Utilizing the student's past performance in all courses that he/she has completed not only increases complexity but also introduces noise in the prediction, thereby degrading the prediction performance. For instance, while it makes sense to consider a student's grade in the course "Linear Algebra" for predicting his/her grade in the course "Linear Optimization", the student's grade in the course "Chemistry Lab" may have much weaker predictive power. However, the course correlation is not always as obvious as in this case. Therefore, discovering the underlying correlation among courses is of great importance for making accurate performance predictions.

Third, predicting student performance in a degree program is not a one-time task; rather, it requires continuous tracking and updating as the student finishes new courses over time. An important consideration in this regard is that the prediction needs to be made based on not only the most recent snapshot of the student accomplishments but also the evolution of the student progress, which may contain valuable information for making more accurate predictions. However, the complexity can easily explode since even mathematically representing the evolution of student progress itself can be a daunting task. However, treating the past progress equally as the current performance when predicting the future may not be a wise choice either since intuition tells us that old information tends to be outdated.

LITERATURE SURVEY

[1] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification," Cybernetics and Information Technologies, vol. 13, Mar. 2013.

Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making. This paper presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management. Universities today are operating in a very complex and highly competitive environment. The main challenge for modern universities is to deeply analyze their performance, to identify their uniqueness and to build a strategy

for further development and future actions. University management should focus more on the profile of admitted students, getting aware of the different types and specific students' characteristics based on the received data. They should also consider if they have all the data needed to analyze the students at the entry point of the university or they need other data to help the managers support their decisions as how to organize the marketing campaign and approach the promising potential students. This paper is focused on the implementation of data mining techniques and methods for acquiring new knowledge from data collected by universities.

[2] S. K. Mohamad and Z. Tasir, ''Educational data mining: A review,'' Procedia Social Behav. Sci., vol. 97, pp. 320–324, Nov. 2013.

Data mining is a technique for extraction of valuable patterns from multiple sources. Data mining plays an important role in marketing, electronic-commerce, business intelligent, healthcare and social network analysis. Advancement in these applications, many researchers show their interest in development of data mining applications in educational context. Educational data mining is a technique defined as a scientific area making inventions within rear types of data that derived from educational surroundings. This Paper reviews different case studies based on data mining educational systems. These systems and mining methods are considered for gathering and analysis of information. Due to huge amount of data in Educational databases, it becomes very challenging to evaluate student performance. Currently in Pakistan, there is dire need to monitor and examine student's academic progress. There are two main causes of why existing systems were not able to analyze performance of students. First, the study on present evaluation methods is still not satisfactory to analyze the appropriate methods for evaluating the progress and performance of students in institutions of Pakistan. Second is because of absence of investigations on parameters; that effects student's success in specific courses. Thus, a comprehensive review is proposed on evaluation of student's performance by using techniques of Data Mining methods to progress student's achievements. The aim of paper is to improve students' academic performance by identifying most suitable attributes by using techniques of EDM.

EXISTING SYSTEM

Most higher education institutions face challenges when they analyze their large educational databases to predict students' performance. This is because they use only conventional statistical methods rather than new and efficient predictive techniques.

Disadvantages of Existing system

- All existing algorithms were concentrating on past data to predict future performance

PROPOSED SYSTEM

In this paper using Machine Learning algorithms author is suggesting concept to predict future courses performances of students by using students previous terms result data as feature vectors. Every year in all universities only 50 % students completing graduation courses successfully and remaining students are failed to complete course so by using this paper machine learning algorithms college peoples can predict future performance of students by giving his past performance GPA as input to the machine learning algorithms.

All existing algorithms were concentrating on past data to predict future performance but in this paper we will used past data as well as on going course performance data to predict future course GPA. Students' evolving progressneeds to be incorporated into the prediction.

Advantages of Proposed System

- a novel machine learning method for predicting student performance.
- High accuracy.

MODULE DESCRIPTION

'Run Random Forest Algorithm' to generate training model using Random Forest and to get it accuracy and MSE

'Run SVM Algorithm' to train SVM classifier and to get it accuracy and MSE value

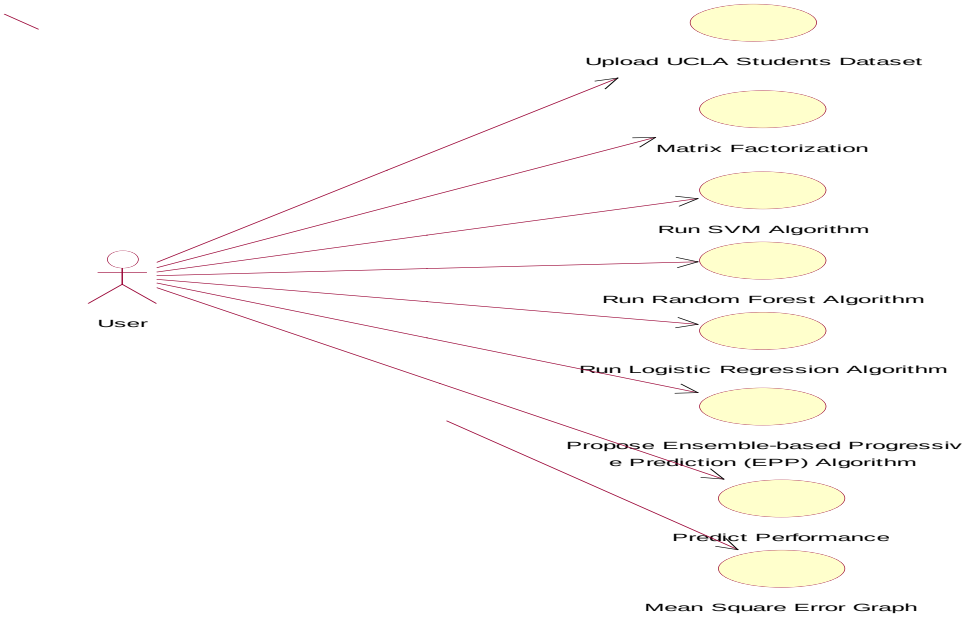
'Run Logistic Regression Algorithm' button to get it accuracy and MSE

'Propose Ensemble-based Progressive Prediction (EPP) Algorithm' to generate model using propose EPP algorithm and to get it accuracy and MSE

'Predict Performance' to upload student on going test marks and to predict GPA for future course

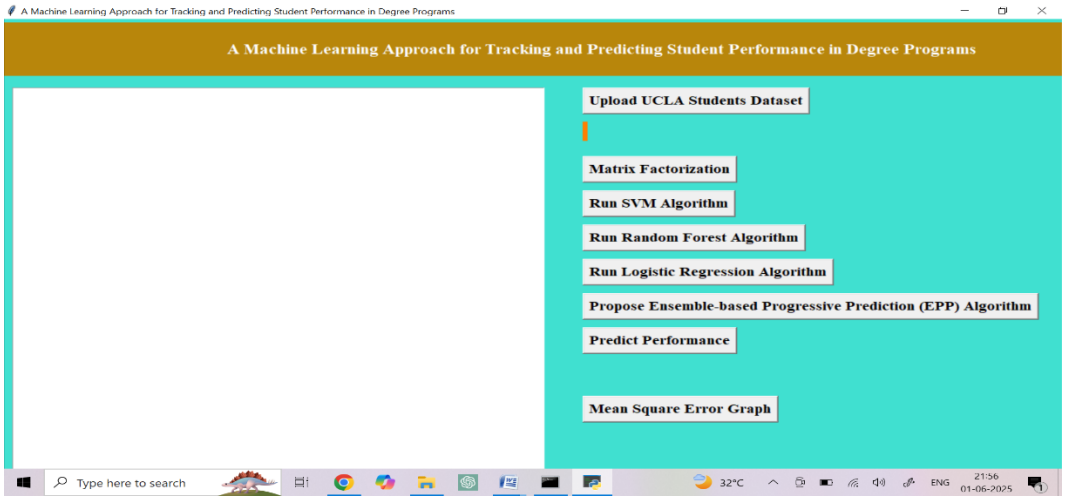
USE CASE DIAGRAM

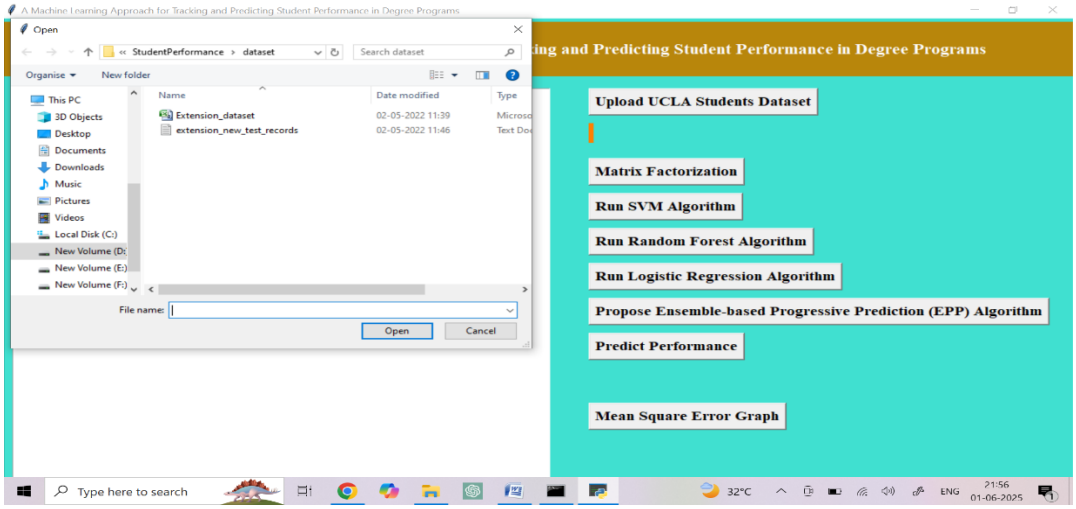
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

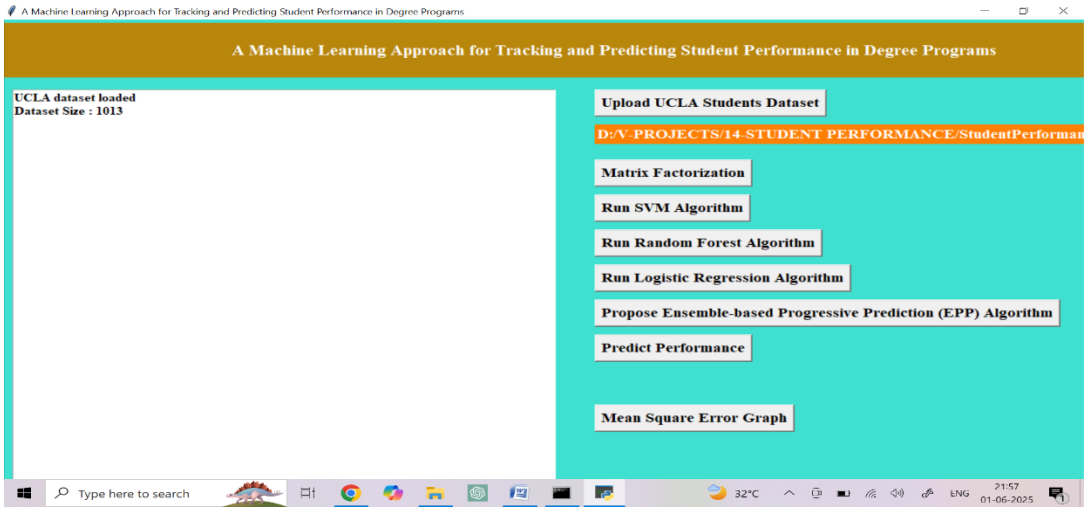


Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor:8.075(SJIF) Vol-5, Issue- 1,2025 ISSN:2582-5887 www.uijes.com

SCREEN SHOTS







CONCLUSION

In this paper, we proposed a novel method for predicting students' future performance in degree programs given their current and past performance. A latent factor model-based course clustering method was developed to discover relevant courses for constructing base predictors. An ensemble-based progressive prediction architecture was developed to incorporate students' evolving performance into the prediction. These data-driven methods can be used in conjunction with other pedagogical methods for evaluating students' performance and provide valuable information for academic advisors to recommend subsequent courses to students and carry out pedagogical intervention measures if necessary. Additionally, this work will also impact curriculum design in degree programs and education policy design in general. Future work includes extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

References:

[1]The White House, "Making college affordable," https:// www.whitehouse.gov /issues/ education/higher-education/ making-college-affordable, 2016.

[2] Complete College America, "Four-year myth: Making college more affordable," http://completecollege.org/wp-content/uploads/2014/ 11/4-Year-Myth.pdf, 2014.

[3] H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis–a general method for cognitive model evaluation and improvement," in International Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164–175.

[4] M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," User Modeling and User-Adapted Interaction, vol. 19, no. 3, pp. 243–266, 2009.

[5] H.-F. Yu, H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T. G. McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei et al., "Feature engineering and classifier ensemble for kdd cup 2010," in Proceedings of the KDD Cup 2010 Workshop, 2010, pp. 1–16.

[6] Z. A. Pardos and N. T. Heffernan, "Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset," Journal of Machine Learning Research W & CP, 2010.

[7] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Personalized grade prediction: A data mining approach," in Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015, pp. 907–912.

[8] C. G. Brinton and M. Chiang, "Mooc performance prediction via clickstream data and social learning networks," in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2299– 2307.

[9] KDD Cup, "Educational data minding challenge," https://pslcdatashop. web.cmu.edu/KDDCup/, 2010.

[10] Y. Jiang, R. S. Baker, L. Paquette, M. San Pedro, and N. T. Heffernan, "Learning, moment-by-moment and over the long term," in International Conference on Artificial Intelligence in Education. Springer, 2015, pp. 654–657.