Prediction of Student Performance Using Machine Learning Techniques: A Review

Chapter · May 2023

DOI: 10.2991/978-94-6463-136-4_63

2 authors, including:

Sonal Deshmukh
Jawaharlal Nehru Engineering College
14 PUBLICATIONS **13** CITATIONS

Prediction of Student Performance Using Machine Learning Techniques: A Review

Nitin Ramrao Yadav(✉) and Sonal Sachin Deshmukh

Jawaharlal Nehru Engineering College, MGM University, Aurangabad, Maharashtra, India
nitin.r.yadav17@gmail.com, hodmcajnec@mgmu.ac.in

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)

1 Introduction

Student performance prediction is one of the indispensable challenges in education sector. 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.

S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 735–741, 2023.
https://doi.org/10.2991/978-94-6463-136-4_63

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.

E-learning is speedily growing and advance form of education, where students are enrolled in online classes. E-learning platforms such as Massive Open Online Course (MOOC) take maximum advantages of ADM in deploying and building automatic grading system. This platform utilizes intelligent tools that collect valuable user information such as frequency of student access to the e-learning system, accuracy of the student answer to the question, number of hours spent for reading and watching videos tutorials.

Machine learning is a subfield of AI, where machine learning system study from the data, generate patterns and predict the outcomes. Machine learning techniques can automatically and quickly analyze large data and very complex data with truthful results. Machine learning algorithm are useful tool for early predicting low performance student based on the derived log data. This technique is more advanced than the traditional on campus, where student records such as: Attendance, Quizzes, Exam, Marks are used to evaluate and predict the student academic performance. This study found that machine learning algorithm was most used for prediction of student performance. Most used machine learning algorithms are Artificial Neural Network (ANN), Naïve Bayes, Decision Tree and SVM.

2 Discussion

In this study, we have reviewed many papers aimed at predicting student performance in education sector. We can draw some conclusions from the analysis of this papers.

Student Performance Prediction (SPP) aims to evaluate the great that a student will reach before enrolling in a course or taking an exam [12]. This will do using five data mining steps including Data Collection, Problem Formulation, Used Methods, Prediction Target, and Practical Applications. Specially three ways of education i.e., Online, Offline and Blended are available. Data mining is the technique of data analysis which aims to extract hidden knowledge from the raw data. Data mining techniques such as cluster analysis and decision tree on educational data of higher educational institution in Croatia [11]. Cluster analysis results in group of students while decision tree used for used for deeper analysis of grouping results. One more data mining technique i.e., classification used to predict the students division on the premise of previous information in Amrita Vishwa Vidyapeetham Mysore. Naïve Bayesion mining technique for data extraction for useful information provides more accuracy [10]. It takes students' academic history as input.

Machine learning is a set of techniques that gives computes ability to learn without any intervention of human programming data from University of Minho, Portugal [1], as input then preprocessed and various machine learning algorithm are applied i.e., Naïve Bayes, SVM, ID3, C4.5, etc. to get more accuracy as a result. In this experiment SVM proven best. Educational data set of secondary schools collected from ministry of education in Gaza Strip for 2015 applied two classification algorithm that is KNN and Naïve Byes, results showed that highest accuracy value of Naïve Byes proven it best than KNN [2]. Learning analytics and predictive analytics method to identify students with lower marks, MapReduce Mongo DB framework model used to implement student performance prediction, which will help teacher to analyze the student performance in Big Data approach [3]. In university Malaysia Sarawak, student result in the form of grid for system analysis and design used classification technique to classify students [4]. Open software tool WEKA is also used. Ayya Nadar Janki Ammal college Sivakasi, Tamil Nadu from computer application department used classification techniques based on C5.0 algorithm, to help learner and teacher to improve the performance as pass or reappear [5]. Comparative analysis of different machine learning algorithm that is DT, NB, ANN, SVM and RF are done for student performance prediction, on student academic data set, Portuguese [6]. Accuracy rate and F measure of data set is measured with two forms that is with attribute selection and without attribute selection. With attribute selection DT performs better and without attribute selection RF perform better.

Machine learning approach to enhance the student performance prediction for the source data set of academic institution in United Arab Emirates [16]. They select 1491 students record as input and use Synthetic Minority Oversampling Technique (SMOTE) to balance the data classes. They select 13 features of student data set such as program, ethnicity, gender, age group, scholarship, transfer status, in drom, course load, admitted on prohibition, result, math level, English level, School system, etc. Then perform 10 folds cross validation method to get the accuracy of models and perform different classification techniques such as ANN, NB, SVM, DT, K-means Cluster, K-Nearest Neighbor and Linear Regression to find out the highest accuracy rates.

As Berhanu and Abera [17] uses 5 years of undergraduate student data of department of horticulture, college of agriculture, Dilla University. They select 199 records and 49 attributes based on the 27-rule set generated by them. After preprocessing of data, data classified by loading it to Rapid Miner (RM) software, which is a standalone application for data analysis. Decision tree techniques results better to provide high accuracy and handle high dimensional data.

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

Table 1. Summary of related work on student performance prediction system.

Topic	Attributes	Algorithm	Dataset/Size	Conclusion
Classification and prediction of student performance data using various machine learning algorithms [1]	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 [2]	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) [4]	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 [5]	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 [6]	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 [7]	19 Attributes	Naïve Bayes	Amrita School of Arts & Science, Mysure	Naïve Bayes is used for knowledge classification

(*continued*)

Table 1. (*continued*)

Topic	Attributes	Algorithm	Dataset/Size	Conclusion
Using Machine Learning models to predict student retention: Building A State-wide early warning system [8]	15 Attributes	LR, NB, DT	Administrative student data (179517 records)	Predict at risk students with 80% probability.

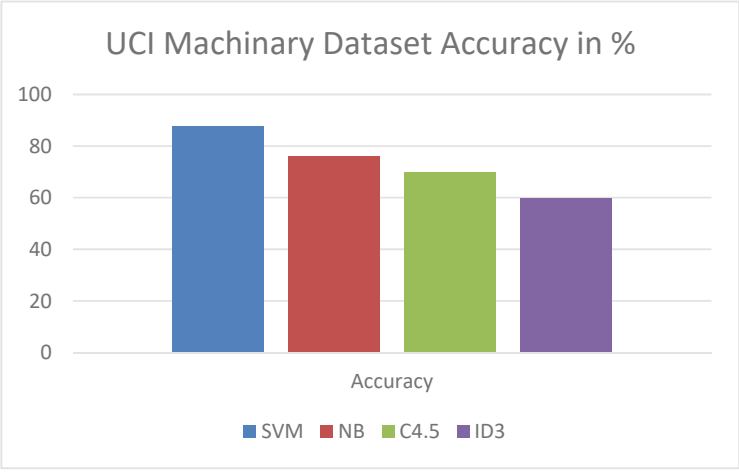

Fig. 1. Student Data Classification Result

3 Experimentation

In experimental analysis of literature, different Machine Learning algorithms are mostly used. The result of algorithms measured in terms of accuracy. In this literature, the base for this review paper shows the performance of different machine learning algorithms in terms of accuracy. They used UCI machinery student dataset as input. Then this data is preprocessed to select the attributes. In this paper they select set of 18 attributes. The machine learning algorithms Naïve Bayes, ID3, C4.5 and SVM are applied on those input dataset. The result obtained by output of algorithms are evaluated in terms of accuracy. Based on the accuracy measure for this dataset with selected 33 attributes, SVM (Support Vector Machine) is the most accurate technique providing highest accuracy that is 88% than other machine learning algorithms (Fig. 1).

If we want to increase the accuracy of the algorithm, then we have to select the relevant set of attributes which affect the accuracy measure. For the student performance prediction, in the review it was found that machine learning algorithms are frequently

used. For different educational datasets different attribute sets are selected which directly affects the accuracy measure of algorithm. For future work we can select the best attribute sets like Stud_Name, Gender, Previous Exam Marks, Address, Parents Education, etc. with machine learning algorithm to provide the high accuracy in result of prediction of student performance system.

4 Conclusion

It is challenging issue to evaluate student performance prediction in education sector or universities. This review shows the intended results of Learning Analytics, Prediction Analytics and different Machine Learning algorithms and Data Mining techniques to improve the student outcome. This study results that machine learning technique is most used approach to enhance student performance and helps teachers to predict student achievements.

The most widely used Machine Learning algorithms to enhance student performance at entry level and during academic year are Artificial Neural Network (ANN), Support Vector Machine (SVM), Naïve Bayes (NB), Linear Regression (LR) and Decision Tree (DT). The data used in this study verity of datasets. Prediction of student performance is calculated in the form of Grade Range, Pass/Fail or Pass/Reappear. The accuracy rate of algorithm depends on the selected attributes set, so it is important to select most relevant attributes to enhance the output of algorithm used.

References

1. Harikumar Pallathadka, Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz, Khongdet Phasinam, "Classification And Prediction Of Student Performance 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 Classification 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, Huned Materwala, Alain Hennebelle, "Comparative Analysis Of Machine Learning Models For Students' Performance Prediction", https://www.researchgate.net/publication/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 Uswatun Khasanah, 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.Tismy Devasia, Ms.Vinushree T P, Mr.Vinayak Hegde, "Prediction Of Students Performance Using Educational Data Mining", ResearchGate, 2020,
11. Snjeˇzana Kriˇzanic, "Educational Data Mining Using Cluster Analysis And Decision Tree Technique: A Case Study", International Journal of Engineering Business Management, 2020.
12. C. A. Palacios, J. A. Reyes-Suárez, L. A. Bearzotti, V. Leiva, and C. 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, "Using Machine Learning Models to Predict Student Retention: Building A State-Wide Early Warning System", Morehead State University, 2020.
15. G. M. Rao and P. K. K. Kumar, "Students Performance Prediction in Online Courses Using 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.