

A Review on EDM Techniques with Special Focus on Student Performance Enhancement

Simmi John^{1*}, Anuj Mohamed²

^{1,2}School of Computer Sciences, Mahatma Gandhi University, Kerala, India

^{*}Corresponding Author: *simmijohn77@gmail.com*

DOI: <https://doi.org/10.26438/ijcse/v7i2.384388> | Available online at: www.ijcseonline.org

Accepted: 20/Feb/2019, Published: 28/Feb/2019

Abstract - Today many of the institutions use data mining techniques, especially in the field of education. The main purpose of Educational Data Mining (EDM) is to increase the quality of education. Use of data mining methods in educational scenarios will help us to learn student behaviour, their performance, enhance the present student models and efficiently design the course curriculum. Teachers will get an overview into the academic performance and administrators can make policies, execute programmes, and adapt the policies and programmes to enhance the teaching–learning process. Using EDM we can improve student's achievements and success more efficiently and effectively. Machine Learning methods are very efficient for predicting student performance. The student data depends on the various educational environments. Selection of the correct dataset plays a vital role in these predictions. EDM uses computational approaches to analyze educational problems and data. By applying data mining techniques we can extract valuable information from huge amounts of data. For extracting knowledge from huge volume of data we require sophisticated set of algorithms and data pre-processing techniques. This paper surveys the most relevant studies carried out in the field of student performance enhancement. It also discusses EDM, areas of student performance enhancements and enhancement methods based on classification.

Keywords— Educational Data Mining, Classification, Knowledge Discovery, Machine Learning, Prediction

I. INTRODUCTION

Educational Data Mining includes various applications of data mining techniques on educational data. Student performance enhancement has been an important research area in EDM which uses both data mining and machine learning methods to explore data from educational environment [1]. EDM is defined by the journal of educational data mining and Baker [2] as an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. The major objective of EDM is the proper understanding of educational phenomena and enhancement of higher education quality. Quality education plays a vital role in the development of a nation. EDM is an interdisciplinary subject which includes features and applications from various fields such as statistics, machine learning, psychometrics, scientific computing, information retrieval, and artificial Intelligence [3]. The educational researchers mainly depend on data mining techniques for analysis and evaluation of their data set. The major goal of EDM is to increase the higher education quality in the field of improvement in decision making, organizational productivity and academic performance of students [4]. The

EDM is a wide area because the data produced from data mining can come from various sources. The application of EDM is different according to the different educational environment. Educational data can come from different areas like administrative data, actual classroom data, intelligent tutoring system (ITS), computer-supported collaborative learning activities, learning management systems (LMS) [5], e-learning, discussion forum, social networks etc [6]. Evaluation of student - teacher performance plays a vital role in higher education quality. Different factors that affect student performance include their prior achievements, academic performance, family background, learning capability, and social factors.

This paper is organized into 5 different sections. Section I gives an introduction about the importance of student performance enhancement. Section II illustrates a brief description about different areas of performance enhancement. Section III presents some classification methods related to this study. Section IV illustrates related researches and conclusion is included in section V.

II. AREAS OF STUDENT PERFORMANCE ENHANCEMENTS

Students' performance can be increased by Predicting and evaluating their performance. It allows academic institutions to provide appropriate support for problems faced by students. Performance assessment can improve their learning capabilities and strategic knowledge. Teachers can improve or take new strategies based on the evaluation result. Some areas for enhancing the student performance are discussed below.

A. Student modeling and grouping students

Modeling student performance in higher education level is an important task of EDM. Various machine learning algorithms have been used for this purpose. Grouping students can be based on their personal behaviour, customized features, learning capabilities etc. It is very helpful for teachers to promote effective group learning. Proper decisions can be made by effectively managing and analyzing the growing volume of data.

B. Analysis of student data

It helps teachers and administrators to analyze learner's activities and learning styles. Statistics and visualization of data are the mainly used techniques for this task. To identify learner's behaviour, learning skill, time taken by each learner, student's interaction with online learning environments are different important tasks. For analyzing student performance we can use different performance metric like accuracy, precision, recall, which can be obtained from confusion matrix.

C. Recommendations for students

EDM plays a vital role in this particular task. Different data mining techniques have been used for this purpose. For recommendations of online learning activities *association rule mining* plays an important role. For grouping students based on special criteria clustering is very useful. The recommendation system helps to improve current curriculum practice.

D. Feedback for instructors and students and for enhancement of their performance

Data mining is playing an important role in identifying and solving educational problems and assessing the performance of instructors in higher education. According to the analysis they can improve their teaching methods [7]. Data mining use various techniques like classification, clustering, association rule mining, and statistic correlation analysis to provide feedback mechanism. This feedback mechanism plays a vital role in improving instructor - student performance. It allows them to evaluating their performance and make necessary corrections [8].

E. Predicting student performance

For enhancing higher education quality predicting student performance is very important [9]. It is one of the most popular applications of EDM. Different factors like students

family background, prior academic performance, psychological, and social factors affect student performance. Using Educational Data Mining techniques we can improve students' achievements and success more efficiently and effectively. Machine learning techniques are very useful for predicting the student performance. The prediction can be done in different environments like traditional classroom education, e-learning environments, intelligent tutoring systems, web based education etc. While measuring the performances of online classes correlation analysis is very helpful. Mainly the performance measuring task includes data collection, pre-processing of the data collected, and applying data mining techniques like classification, interpretation and evaluation of results.

III. STUDENT PERFORMANCE ENHANCEMENT METHODS BASED ON CLASSIFICATION

Classification technique has wide applications in both traditional and modern educational settings. Classification is a machine-learning method for constructing prediction models from data [10]. It is a process in which objects are classified based on different criteria. In educational paradigm, teachers classify their students based on their performance, behaviour and knowledge. We use different algorithms and techniques based on the dataset used. Classification is a supervised learning technique. It assigns the data into a predefined classes or groups. It is also called predictive data mining technique because it is used for prediction purposes. The predicted variable can be either categorical or binary variable. There are different types of classification methods. Some of the popular methods are described below [11].

A. Decision Trees

Most of the researchers use decision tree classifier for analyzing their data set. Decision tree builds classification models in the form of a tree like structure [12]. It will help the researchers to reach a final decision. The topmost decision node in a decision tree is called the root node and the leaf node represents a decision. Commonly used algorithms for building decision trees are ID3 (Iterative Dichotomiser 3), ID4.5, ID5 [13], and CART (Classification and Regression Trees). The split criterion like Information Gain, Gain Ratio, and Gini Index are different for these algorithms. There are many studies that apply this algorithm for analyzing undergraduate students' performance [24].

B. Neural network

Artificial neural network is one of the best prediction models used in educational data mining. Neural network algorithms are mainly designed to identify patterns. It is composed of a large number of interconnected processing elements called neurons. The learning algorithm is an essential part of this model. It has the ability to detect all possible interactions between predictor variables [14].

Neural network can be used for solving many complex classification problems. Multi Layer Perception is commonly used in learning network model for academic needs.

C. Naïve Bayes Classifier

It is a supervised learning classifier based on Bayes theorem. This is highly scalable in nature. One of the advantages of this classifier is its requirement of less number of training data. This technique uses probabilistic relation between the classes and their attribute. In a supervised learning environment Naïve Bayes classifiers can be trained very efficiently and effectively. Using this algorithm many complex real world situations can be handled. There are different studies on enhancing student performance using Naïve Bayes Classification algorithm [21] [23].

D. Support vector machine (SVM)

SVM is a machine learning algorithm used for classification purpose. SVM are supervised learning models used for classification and regression analysis. Using SVM we can avoid over fitting and under fitting. We can use SVM to predict something belongs to a particular group. In SVM a hyper plane is used to separate the classes to minimize the classification error and to maximize the margins. A hyper plane is a line that split the input variable space. In SVM, when the data is mapped to a higher dimension, the classes become linearly separable [15]. Various comparative studies are being conducted for evaluating the performance of students [18].

E. KNN

K-Nearest Neighbour algorithm (KNN) is mainly used for classification and regression. It is a non parametric and simplest of all machine learning algorithms. Non parametric means it does not make any assumptions on the underlying data. Major drawback of this method is that it needs high memory requirement. In KNN Euclidean distance is the most commonly used distance metric.

IV. RELATED RESEARCHES

Ashraf Abazeed and Moaiad Khde [16] proposed a model for predicting and evaluating student performance. Their study suggested two techniques for this purpose. The first one is based on classification techniques and the second one is based on association rule mining. According to this study student can improve their performance and reduce the failure rate. They used random tree classifier for their studies. Random tree is a supervised machine learning classifier. It is a collection of individual decision trees in which each tree is generated from different samples and subset of the training data. One of the advantages of random tree classifier is that it resists over fitting and can work with segmented fields. The data set contain 242 students from the

computer science department of the Applied Science University Bahrain.

Zollanvari, A et al. [17] proposed a machine learning techniques to construct a predictive model of GPA using self-regulatory learning behaviors. They used maximum-weight first-order dependence tree technique. According to their studies cognitive and metacognitive self regulatory behaviors play a vital role in predicting students' academic performance. Based on this study teachers can interact with students according to their personality and behavior.

Nieto, Y et al. [18] used two machine learning algorithms like artificial neural network and support vector machine to compare student performance. According to their studies, students' academic performance has a great influence rather than their socio economic features. They analyzed the results based on confusion matrix and the receiver operating characteristic curve.

Jie Xu, Kyeong Ho Moon, Michaela van der Schaar [19] use a machine learning algorithm to predict student performance in degree level. It is based on two features. The first approach is a bilayered structure based on multiple base predictors and a cascade of ensemble predictors. The second approach is a data-driven clustering method; it is based on latent factor models and matrix factorization techniques.

Altujjar, Yasmeen [20] suggests predicting student performance is one of the important applications of EDM. In academic institutions this type of research is very helpful for decision making. The institutions can easily evaluate and detect the problems of students by using EDM techniques. Here they use ID3 Decision Tree Induction algorithm for predicting student performance. A decision tree is constructed in top-down recursive approach and it is based on greedy algorithm. Maximization of information gain is depending on the splitting attribute selection. The performance measures used by them are accuracy, precision, and recall. They used Rapid Miner tool for their analysis.

Leena Khanna, Dr. Shailendra Narayan Singh, Dr. Mansaf Alam [4] focused on improving the quality and effectiveness of education. These are the main objectives of higher educational institutions. EDM techniques can predict the future of the students and make necessary improvements in teaching learning process. In higher educational institutions the quality of education mainly depends on student's academic performance. The main EDM users are classified into students, teachers, administrators, educational authorities, research persons etc. The educational environment includes traditional class rooms, e-learning systems, adaptive intelligent and web based educational system.

Mueen, Ahmed, Bassam Zafar, and Umar Manzoor [21] used data from two under graduate courses for predicting the academic performance. The classification algorithms used are Naïve Bayes, Neural Network, and Decision Tree. According to their analysis Naïve Bayes gave better result. Various factors which affect student performances are previous academic performance, family background, psychological factors etc. So it is a challenging task. Different factors or combination of different factors affect the performance of student and it varies from student to student, institution to institution and nation to nation.

Huseyin Guruler and Ayhan Istanbul [22] proposed modelling student performance as successful or unsuccessful based on their GPA. According to their studies the demographic data has a great influence in students GPA. Universities mainly depend on EDM techniques for decision making purpose. They introduced a data driven educational approach called Student knowledge discovery software

(SKDS) based on data mining techniques. They used Decision tree classification techniques for evaluating their data. The dataset includes the demographic data of students in the faculty of Economics and Social science of Mugla Sitki Kocman University.

Brijesh Kumar Bhardwaj and Saurabh Pal [23] proposed a predictive DM model for predicting student performance and differentiate students as high learners and low learners. They used datasets of five degree colleges conducting BCA courses affiliated with Dr. R. M. L. Awadh University, Faizabad, UP, India. A total of 300 students were included in their study. Their study considered various attributes including student's academic, demographic and socio-economic variables. The data set is directly collected through questionnaire and University database.

Some of the recent findings in EDM based on student performance are summarized in Table 1.

Table 1: Summarisation of papers based on student performance enhancement

ID	AUTHORS	METHODS USED	OUTCOMES
1	Abazeed et al. (2017) [16]	Classification techniques (random tree classifier) and association rule mining	This study help the students to determine future direction in their studies and who needs any special attention can be highlighted.
2	Zollanvari et al.(2017)[17]	Constructed a predictive model using maximum weight dependence trees.	A set of cognitive and meta cognitive self-regulatory behaviours are considered. Probabilistic predictive model of GPA (Grade Point Average) is helpful to improve their academic performance.
3	Nieto, Y et al. (2018)[18]	Artificial neural network and Support vector machine	Based on this study, graduation rate predictions mainly depend on Students' academic performances rather than their socio - economic factors.
4	Jie Xu et al. (2017) [19]	Used a bilayered structure based on multiple base predictors and a cascade of ensemble predictors.	Performance in degree level can be improved by using this proposed machine learning algorithm.
5	Altujjar et al. (2016) [20]	Used ID3 Decision Tree Induction algorithm.	The goal of this study is to predict the performance of students and identify critical courses in IT program. Performance prediction at early stages can eliminate student's problems and enables allocation of proper support to students.
6	Leena Khanna et al. (2016) [4]	Classification, Association rule mining	This study focus on various factors which will improve the quality of higher education and also compare different data mining methods and tools.
7	Mueen et al. (2016) [21]	Naïve Bayes, Neural Network, and Decision Tree	Based on this study, teachers can early detect the weak students and provide special attention to them.
8	Jacob John et al. (2015) [9]	Decision tree and correlation	This method will greatly help the university teachers to know what changes need to be made, identify weak students at risk and provide remedial courses to weak students.
9	Guruler et al. (2014) [22]	Decision tree classification techniques for evaluating the data.	This research aimed to evaluate and develop a data driven approach to improve the performance of university students using a new developed educational technology system called 'student knowledge discovery software'.
10	Bhardwaj et al.(2012) [23]	Bayes classification prediction model construction	This study will help the teachers and students to improve their performance and to identify the differences between high learners and low learners.

V. CONCLUSION

Education is the most important element in the evolution of the nation. So education plays a vital role in the economic prosperity of a country. Data mining techniques in educational institutions help us to learn students' academic performance, learning capability, their behavior etc. which in turn can be used in designing course and curriculum and also to motivate students on various activities. This paper mainly discusses educational data mining, and its application in the field of student performance enhancement. The main objectives of EDM is to eliminate current research issues in the field of education, development of flexible and user friendly EDM tools, provision of consistent feedback mechanism and development of different strategies to meet their goals.

ACKNOWLEDGEMENT

The authors acknowledge the support extended by DST-PURSE (Phase II), Government of India.

REFERENCES

- [1] Guo, Bo, et al. "Predicting students' performance in educational Data Mining", Educational Technology (ISET 2015), International Symposium on IEEE, 2015.
- [2] Baker, R.S.J.d, "Data mining for education", In: McGaw, B., Peterson, P., Baker, E. (eds.) International Encyclopedia of Education, vol. 7, 3rd Edn. pp. 112-118. Elsevier Amsterdam 2010.
- [3] Romero, Cristobal, Sebastian Ventura, "Data mining in education", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3.1, 12–27, 2013.
- [4] Leena Khanna, Dr. Shailendra Narayan Singh, Dr. Mansaf Alam, "Educational Data Mining and its Role in Determining Factors Affecting Students Academic Performance: A Systematic Review", Information Processing (IICIP), I st India International Conference On IEEE, 2016.
- [5] Alom, B.M., Matthew Courtney, "Educational Data Mining: Case Study Perspectives from Primary to University Education in Australia", 2018.
- [6] Romero, Cristobal, Sebastian Ventura, "Educational data mining: a Review of the state of the art", IEEE Transactions on Systems, Man, And Cybernetics, Part C (Applications and Reviews) 40.6: 601-618, 2010.
- [7] Agaoglu, Mustafa, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education", IEEE Access 4:2379-2387, 2016.
- [8] Kaur, Amanpreet, Amanpreet kaur Bath, "Predicting Instructors Performance using Data Mining Techniques", International Journal of Engineering Science 12835, 2017.
- [9] Jacob John, et al. "Educational Data Mining techniques and their Application", Green Computing and Internet of Things (ICGCloT), International Conference on IEEE, 2015.
- [10] Han, J., Kamber, M., Pei, J., "Data Mining: Concepts And Techniques", 3rd edn. Elsevier, Oxford, 2011.
- [11] Shahiri, Amirah Mohamed, and Wahidah Husain, "A review on Predicting Students Performance using data mining techniques", Procedia Computer Science 72: 414- 422, 2015.
- [12] Walte, Dipesh, et al. "Overview of algorithm in Educational Data Mining for Higher Education: An Application Perspective", International Journal of Engineering Research & Technology 3.2, 2014.
- [13] Kawchale, Neha, "Prediction of Student Performance in Higher Education System using R Programming", 2017.
- [14] R. Hecht-Nielsen, "Theory of the back propagation neural Network", In Proceedings of the International Joint Conference on Neural Networks (IJCNN), volume 1, pp. 593–605. IEEE, Washington, DC, 1989.
- [15] V.N. Vapnik, "Statistical Learning Theory", John Wiley & Sons, New York, 1998.
- [16] Abazeed, Ashraf, and Moaiad Khder, "A Classification and Prediction Model for Student's Performance in University Level", 2017.
- [17] Zollanvari, A., Kizilirmak, R.C., Kho, Y.H. and Hernández - Torrano, D., "Predicting students' GPA and developing intervention Strategies based on self-regulatory learning behaviors", IEEE Access, 5, pp.23792-23802, 2017.
- [18] Nieto, Y., García-Díaz, V., Montenegro, C. and Crespo, R.G., "Supporting academic decision making at higher educational institutions using machine learning-based algorithms", Soft Computing, pp.1-9, 2018.
- [19] JieXu, Kyeong Ho Moon, Mihaela Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs", IEEE Journal of Selected Topics in Signal Processing 11.5: 742-753, 2017.
- [20] Altujjar, Yasmeen, et al. "Predicting Critical Courses Affecting Students Performance: A Case Study", Procedia Computer Science 82: 657, 2016.
- [21] Mueen, Ahmed, Bassam Zafar, and Umar Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques, International Journal of Modern Education and Computer Science 8.11: 36, 2016.
- [22] Guruler, Huseyin, Ayhan Istanbullu, "Modeling Student Performance in Higher Education using Data Mining", Educational Data Mining. Springer, Cham, 105-124, 2014.
- [23] Bhardwaj, Brijesh Kumar, Saurabh Pal, "Data Mining: A Prediction for Performance Improvement using Classification", ArXiv Preprint arXiv: 1201.3418, 2012.
- [24] Asif, R., Merceron, A., Ali, S.A., Haider, N.G, "Analyzing Undergraduate students' performance using educational data Mining", Computers & Education, 113, pp.177-194, 2017.