International Journal of Engineering & Technology, 7 (2.33) (2018) 634-636

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A survey on educational data mining techniques in predicting student's academic performance

P. Amutha [1] *, Dr. R. Priya [2]

[1] Research Scholar, Department of Computer Applications, VISTAS, India
[2] Associate Professor, Department of Computer Applications, VISTAS, India
**Corresponding author E-mail: priyaa.research@gmail.com*

Abstract

The massive growth in the educational sector needs to create awareness about handling the huge volume of student data. The educational data mining is a technique to extract information from these volumes of data. Nowadays educational data mining technique plays a vital role in predicting academic performance. The objective of this study is to explore the extended knowledge of different educational data mining techniques, which have been used to predict the academic performance.

Keywords: *Academic Information; Educational Data Mining; Student Performance; Prediction Techniques*

1. Introduction

Educational data mining is a current trend in the field of data mining and knowledge discovery in databases which have been extracted useful patterns and discovered knowledge from educational information systems. Data mining techniques are very significant tools to analyzing the large datasets and extracting previous unknown, credible, potentially used and hidden patterns from these large datasets. Researchers and professionals in education focus on discovering useful knowledge either to help educators to work according to the learning standards of the students or to help student enhance their future performance.
This paper is focused on various educational data mining techniques such as classification, clustering, neural network, association rule that can be used to identify the student attributes and data mining techniques which are mostly used for predicting the academic performance in education [14].
Advantages of educational data mining are

- Improving the student performance
- Recognition of slow learners
- Modeling indicators to predict the student dropout
- Predicting student performance in UG level
- Improving student attendance
- Predicting learning behavior and to warn student at risk
- Select student groups with characteristics towards learning strategies

2. Educational data mining phases

Educational Data Mining is related to the translation of new hidden information from the raw data collected from educational systems. EDM generally contains the following phases.

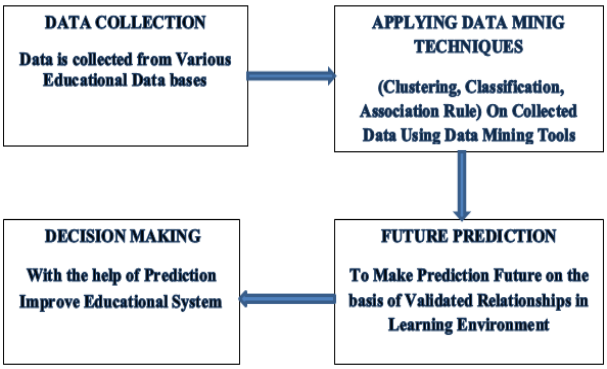


Fig. 1: Educational Data Mining Phases.

The first phase of educational data mining is to collect data which is to be mined from different educational system resources i.e. from course management system (different institutes), E-learning environment, web-based data (i.e. YouTube, Twitter) which is relevant to students activities during learning process(i.e. their academic grades, students posts on social networking sites etc.)
The second phase of educational data mining is a validation of discovered relationships between data so that uncertainty can be avoided using data mining techniques i.e. classification, clustering, regression etc. The third phase is to make predictions for future on the basis of validated relationships in the learning environment.
The fourth phase is supporting decision-making process with the help of predictions [15].

3. Educational data mining methods

Educational data mining not only incorporates data mining techniques classification, clustering, and association analysis but also includes statistics, machine learning, text mining, web log analysis, etc., which is derived from different areas related to EDM. There are so many methods of educational data mining but all kind of methods spread over in one of following categories [15], [16].

Copyright © 2018 P. Amutha, Dr. R. Priya. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3.1. Prediction

Using prediction methods namely classification, regression (when the predicted variable is a continuous value), or density estimation (when the predicted value is a probability density function) infer single trait of the data (predicted variable) from some combination of other traits of the data (predictor variables).

3.2. Regression

Regression is a fundamentally statistical technique used regularly in data mining. Regression analysis showed the relationship between a dependent or outcome variable and a set of predictors. Regression is supervised learning data mining technique. Supervised learning partitions the database into training and validation data. There are two types of regression technique.

- Linear regressions.
- Non-Linear regressions

3.3. Classification

Arranges a data item into some of several predefined categorical classes. The algorithm exploited for classification are

- Decision tree
- Naive biased classification
- Generalized Linear Models (GLM)
- Super vector machine etc.

3.4. Clustering

In clustering technique, the dataset is divided into various groups, known as clusters. The data point of one cluster and should be more similar to other data points of the same cluster and more dissimilar to data points of another cluster. Two ways to implement clustering algorithm.

- With no prior assumption.
- With a prior postulate

3.5. Relationship mining

It is used for discovering relationships between variables in a dataset and encoding them as rules for later use. There are different types of relationship in mining techniques such as association rule mining (any relationships between variables), sequential pattern mining (temporal associations between variables), correlation mining (linear correlations between variables), and causal data mining (causal relationships between variables). In EDM, relationship mining is used to identify relationships between the student's online activities and the final marks and to model learner's problem-solving activity sequences.

3.6. Discovery with models

Its goal is to use a validated model of a phenomenon (using prediction, clustering, or knowledge engineering) as a component in further analysis such as prediction or relationship mining. It is used for example to identify the relationships between the student's behavior and characteristics .

3.7. Outlier detection

An outlier is a different observation (or measurement) that is usually larger or smaller than the other values in data. In educational data mining, it can be used to detect deviations in the learner's or educator's actions or characteristics, irregular learning processes, and learning difficulties.

4. Literature review

4.1. Machine learning based student performance prediction

Garima Sharma et al. [1] were implemented ID3 decision tree to help the students, instructors, and guardians to take necessary action to improve the performance of the students in future. In this study, 90% accuracy had achieved. They had considered student semester data as attributes to measure the performance.

Raheela Asif et al. provided a good implementation such feature selection algorithm to predict the final year graduation of the students with greater accuracy. They used attributes as university pre-examination marks and first year and second-year marks of the civil engineering department at EDUET [2].

K. Karthikeyan and P. Kavipriya show that K- means to reduce the number of support vector for classification by using Academic and classroom activities of UG students in the year 2015 as attributes to predict performance [3]. They are examined on two dataset namely data set 395 and data set 690.

Raheela Asif et al. were implemented decision tree to provide better indicators to predict good and poor performance in the degree programmes of the students. University pre- examination marks and first and second-year marks are considered as attributes [4].

4.2. Association rule mining in predicting student performance

Sushil Kumar Verma and R.S. Thakur had shown that Fuzzy association is used to decrease the failure rate and taking action for the next semester examination. They used two Dataset, data set 50 and data set 154 from students of Samrat Ashok Technological Institute [13].Different educational data mining techniques are applied to predict student's performance at various circumstances shown in table1, table2, and table3 respectively

Table 1: Classification Techniques in Predicting Student Performance

Reference paper	Algorithm/ technology	Test data	Finding
Garima Sharma, Santosh K. Vishwakarma., Feb 2017	ID3 Decision Tree, Rapid Miner Platform	Semester data of the student at Computer Science Dept, G.G.I.T.S Jabalpur	Help the students, instructors, and guardians to take necessary action to improve the performance of the students in future.
Raheela Asif et al., May-2017	Decision tree , Feature selection –Rapid Miner	Pre university and Examination (1^{st}, 2^{nd} year) marks. 2005-2006 and 2006-2007 batches	Predict the final year graduation of Students with greater accuracy
K. Karthikeyan and P. Kavi Priya, May- 2017	K-Means with SVM	Academic and Classroom activities of UG student in the year 2015. Data set1 395 and Data set2 650	Reduce the number of support vector for classification
Raheela Asif et al., May 2017	Decision tree	Pre university and Examination (1^{st}, 2^{nd} year)marks	Provides Indicators of good and poor performance in the degree programme.

Table 2: Comparative Study in Predicting Student Performance

Reference paper	Algorithm/ technology	Test data	Finding
Vinaya Patil et al., April-2017	Classification-Naïve Bayes, LibSVM, C4.5, Random Forest and ID3	CSE data set of SSBT COET	Predict the grades and drop out of a student.
Saurabh Pal and Vikas Chaurasia, April 2017	Decision Tree (BFTree, J48, RepTree and Sample cart	450 students from MCA Dept. VBS Purvanchal University	Among the techiques BFTree is having, more accuracy ad took less time to build the model.
E. Venkatesan and S. Selvaragini, 2017	Classification – C4.5 and ID3	Behavior and results in various type of examinations	The performance of C4.5 well in processing the student's data.
Bhrigu Kapur et al., March –April 2017.	Decision tree, Random forests, Naïve Bayes, Rule induction, K- Star, IBK algorithm	480 student record with 16 attributes	Random Forest is more suitable for the large data set.
Surbhi Agrawal et al., August 2017	Decision tree C4.5 algorithm, random forests, naïve bayes, rule induction – Rapid Miner	Two Portugal school Students' living habits and background details (corpus) in the 2005-2006.	Prediction accuracy percentages of each classifier analyzed. Among these, decision tree produces the highest accuracy to predict final semester grade.

Table 3: Machine Learning Based Student Performance Prediction

Reference paper	Algorithm/ technology	Test data	Finding
Jie Xu et al., 2017	Ensemble learning	Real world student dataset from UCLA Mechanical and Aerospace engineering	Provide valuable information for academic advisors to recommend courses to student
S. Kalaivani et al., January -2017.	Reweight enhanced Boosting Algorithm	Grade sheet consists of sub code, CCET1, CCET2, CCET3, Semester grade	Final semester marks are predicted from internal marks
Zafar Iqbal et al., August 2017	Collabrative Filtering Matrix Factorization Restricted Boltmann Machine	Real world data from 255 undergraduate students of EE Dept. at ITU	Restricted Boltzmann Machine is better than others are.

5. Conclusion

This paper has reviewed lots of research papers on analyzing educational related data to develop models for improving the performance of the students in future, predicting the grades and dropout, predicting the higher education using pre-examination marks of the students, predicting student career and their capacity take up the course. This research work investigated which data mining technique is best suited for educational datasets; in classification technique, the decision tree is best algorithm to predict the performance of the student and also more accurate than other algorithms. The clustering and association rule mining will be beneficial for the researchers to do future research in our educational environment to make grouping among students based on their performance and predicting the regular performance of the student.

References

[1] Garima Sharma and Santosh K. Vishwakarma, Analysis and prediction of Student's Academic Performance in University Courses, International Journal of Computer Application, Vol.160-No 4, February 2017.

[2] Raheela Asif, Saman Hina, Saba Izhar Haque, "Predicting Student Performance Using Data Minig Methods, International Journal of Computer Science and Network Security, Vol. 17, No5, May 2017.

[3] K. Karthikeyan and P. Kavi Priya, On Improving Student Performance Prediction Using Enhanced Data Mining Techniques, International Journal of Advanced Research in Computer Science and software Engineering, Vol. 7, May 2017.

[4] Raheela Asif, Agathe Merceron, et al., Analyzing undergraduate Student's performance using Educational Data Mining, Journal of Computers and Education 113(2017) 177-194.

[5] Vinaya Patil, Shiwani Suryawanshi, et al., Student Performance Prediction Using Classification Data Mining Techniques, International Journal for Research in Emerging Science and Technology, Vol.4 April 2017.

[6] Saurabh Pal and Vikas Chaurasia, Is Alcohol Affect Higher Education Student Performance: Searching and predicting pattern using Data Mining Algorithms, International Journal of Innovations and Advancement in Computer Science, Vol. 6, and April 2017.

[7] E. Venkatesan and S. Selvaragini, A Study on the Result Based Analysis of Student Performance Using Data Mining Techniques, International Journal of Pure and Applied Mathematics, Vol. 116,No. 16 , 2017.

[8] Bhrigu Kapur, Nakin Ahluwalia and Sathyaraj, Comparative Study of Marks Prediction Using Data and Classification Algorithms, International journal of Advanced Research in Computer Science, Vol. 8, No. 3, March –April 2017.

[9] Surbhi Agrawal, Santosh K. Vishwakarma and Akhilesh K. Sharma, Using Data Mining Classifier for Predicting Student's International Journal of Computer Applications, August 2017.

[10] Jie Xu, Yuli Han,et al., Progressive Prediction of Student Performance in College Programs, Association for the Advancement of Artificial Intelligence, 2017.

[11] S. Kalaivani, B.Priyadharshini, et al., Analyzing Student's Academic Performance Based on Data Mining Approach, International journal of innovative Research in Computer Science and Technology, Vol. 5, January 2017.

[12] Zafar Iqbal , Junaid Qadir, et al. Machine Learning Based Student Grade Prediction: A Case Study, August 2017.

[13] Sushil Kumar Verma and R.S. Thakur, Fuzzy Association Rule Mining based Model to Predict Students' Performance, International journal of Electrical and Computer Engineering, Vol.7, No. 4. August 2017.

[14] Mukesh Kumar, et al. Literature Survey on Student performance in Education using Data Mining Techniques, International Journal Education and Management Engineering, 6, 2017.

[15] N. Upadhyay, K. Katiyar, A Survey on the Classification Techniques in Educational Data Mining, International Journal of Database Management System, 3(11), 2014.

[16] Baradwaj, B. Kumar, Mining Educational Data to Analyze Students Performance, International Journal Advanced Computer Science and Applications, 2(6), 2011.