International Journal of Engineering & Technology, 7 (2.33) (2018) 522-525

International Journal of Engineering & Technology

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

Research paper

An inclusive survey of students performance with various data mining methods

A. S. Arunachalam [1] , K. Rajeswari [2]

[1] *Assistant Professor, Department of Computer Science, School of computing sciences ,Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*

[2] *Research scholar, Department of Computer Science, School of computing sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*

**Corresponding author E-mail: arunachalam1976@gmail.com*

Abstract

Educational institutions are the source of generating quality students in order to make them do better service for the nation. It is a must for all educational institutions to be aware of the competency and academic level of every student so as to study their performance. The key attributes to identify the performance are to have controlled parameter with clear data. So it is important to set the standards and calibration measures in order to make the study of students' performance an efficient one. There are several tools and techniques available to perform this prediction study. Among all, Data mining is the best and the most efficient technique to handle prediction process. Among data mining, EDM (Educational Data Mining) is much more popular in the present century and hence it is beneficial to make a research on the current technique. This survey paper focuses on various Data Mining approaches in order to forecast student's performance and bring clarity in students' results and faculty's contribution as a success.

Keywords: *Educational Data Mining; Students Performance; Patterns; Classification*

1. Introduction

Educational institutions are to facilitate education service to students along with managerial decisions. The most possible and achievable way of attaining this goal is by learning the factors and attributes that has an impact on student's performance. The extracted information from the big data or the raw data can be of extreme help to academic planners in the educational department. It also helps them to develop their decision making process and to have a better idea on students' behavior. It also supports to enhance students performance and thereby reducing the failure rates. Data mining is the concept of analyzing data that uses several software techniques in order to find the unpredicted and unknown patterns hidden in the data. It also helps to identify the relationships in the data set. Data mining techniques are broadly divided into: supervised and unsupervised learning [1] .Data mining techniques are pertained to forecast students academic performance based on some of the attributes like socioeconomic condition, earlier academic performances and so on. Classification is one type of data mining technique of the predictive types that classifies data based on the training set. It then uses the generated pattern to classify a new data [2]. Educational data mining is a rising task that uses several new techniques in order to extract the new data. These techniques also help to have a greater prediction rate in understanding academic performance, students' behavior, subject interest etc [3] Educational data mining is a derivate of data mining that is oriented to educational field. Educational data mining is specially designed to be applied on data derived from educational background. EDM is similar to data mining process where raw data aggregated by education systems is converted into some useful information. There are some unique algorithms and tools specially designed for educational data mining to analyze data patterns. There are numerous data mining techniques like classification, clustering, rule mining etc to extract the useful information from the data.

2. Related works

Pal and Yadav [4] used the VBS University student's data like Discipline, Category and Student grade in high school and family size. It also uses the previous year database of the students so as to predict students who are likely to fail in the examination using CART, C4.5 and ID3 Algorithm. The experiment concluded that C4.5 is the best algorithm among the rest. Hijazi et al [5] used a sample of 300 students (75 females, 225 miles) from Punjab university of Pakistan. The results of the study state that family income, student's attitude, hours spent in study after college, mother's age and education are important attributes to decide student performance.

A. S. Arunachalam and T. Velmurugan [6] are given the detailed thought of educational data mining and core paths of EDM. The techniques and tools discussed in this survey will provide a clear-cut idea to the young educational data-mining researchers to carry out their work in this field. R.Shanmuga Priya [7] proposed a paper work on improving student's performance using Educational Data Mining. It considered 50 students from Hindustan College of Arts and Science, Coimbatore, India. The decision tree classifier was implemented with 8 attributes and it was observed that seminar, class assessment, lab practical and attendance is used in order to predict the student functioning. This prediction tool helps teachers to pay specific concentration on students who need help and confidence in their studies.

C. Marquez, et al [8] performed a research to analyze the variables that expresses attention to reduce low performance of students. 670 students data from Zacatecas, Mexico was used for this experiment. Classification algorithm was implemented on the various chosen components and the outcome stated that economic or educational characteristics and sociological characteristics are the best variables to identify the academic performance of the students. Pal and Pandey [9] implemented a research on 600 student from various colleges of Dr. R. M. L. Awadh University, Faizabad, India.

Using Bayes Classification of group, it was concluded that accent and background adequacy, are the reasons for a newcomer to perform or not. Ryan S.J.D. Baker and Kalinayacef [10] studied the history and development in educational data mining (EDM), 2009. The study reviewed the increased importance on prediction and was proposed to be applied on the existing models to make scientific discoveries. Elaraby and Abeer [11] carried out a similar study that targets on generating classification rules and also to predict students' performance based on previously filed students' activities and behavior. They used previously enrolled students' data (2005–2010) and analyzed it and processed it. Multiple attributes were taken into consideration. The study was able to predict students' final grades and also help the student's to progress on their performance and identify students who require special attention and reduce the failing ratio [11]. Sumam Mary Idikkula, Joseph Alexander and Sudheep Elayidom[14] proved that data mining is an effective tool to predict employment and aid students to select a good branch that can assure them placement in future. [12]

3. Data mining approaches in prediction of students performance

Educational data mining is a recent field in data mining and there are many researches being proposed to understand the advantage and uses of data mining in the field of education. Various techniques like decision tree, rule induction, neural networks, k-nearest neighbor, naïve Bayesian are being implemented in educational data mining. Using the above approaches huge knowledge can be derived from data by setting up classifications, association rules and clustering.

3.1. Classification

Classification follows a mining strategy with the help of pre classified attributes so as to build a model that can collect huge data. This technique uses neural network based classification or decision tree algorithms in general. Data classification involves learning of data and then followed by classification. In the process of learning training data they are analyzed using algorithm. The accuracy of the proposed rule is experimented on the test data. Classification is so far the best and the most natural data mining technique we ever had. Educational Data Mining (EDM) is also the best method so far to analyze collected students' information. EDM analyses the data collected through a survey and retrieves information through classification technique so as to classify students' academic performance in future.

3.2. Clustering

Clustering is the concept of aligning data points into clusters that possess similar character. Clustering works best when the user is aware of the common categories that is present in the data set.[18] Clusters are formed in various possible grain-sizes: like in case of schools the group are clustered to identify the similarities and differences among schools, students and departments and data can be clustered accordingly to study the similarity between students. Clustering does not require a pre hypotheses or a specific hypothesis that was generated earlier research with a distinct data set (Expectation Maximization algorithm).The quality of the clusters is based on how efficiently the clusters fit the data. This can be identified using the Bayesian Information Criterion.

3.3. Prediction

Prediction is a way of gaining that single aspect (predicted variable) from the different blend of aspects of the information (predictor factors). This is the most commonly used method in educational data since it aids to predict student educational outcomes (Romero et al, 2008) without the need to predict intermediary or mediating components. It is also used to predict the expected output value in context where it does not need to directly obtain a label for that construct.

Baker et al (2008) implemented a prediction model with observational methods so as to label a small data set. The model builds up an expectation show that utilizes the gathered information from cooperation amongst understudies and also the product for indicator factors; later proving the model's exactness by summing up to the extra understudies and settings that are available. There are three types of prediction: classification, regression and density estimation. In classification the predicted variable belongs to the binary or categorical variable type. In regression, the variable is of consistent type. The regression procedures mostly uses support vector machine regression, linear regression and the neural networks. [19]

4. Data mining techniques

Educational data mining is an upcoming technique in the field of data mining. Many researches are being proposed to understand the advantage and uses of data mining in the field of education. Several techniques like rule induction, decision tree, neural networks, naïve Bayesian, k-nearest neighbour are being implemented in educational data mining.

4.1. Decision tree

Decision tree techniques are simple to be understood and implemented. It supports addition of new possible scenarios. It also aids to find worst, best and expected values for different scenarios. It can also be merged with other decision tree techniques in order to generate rules with no trouble [20]. There are many advantages and disadvantages of using this technique. Since the number of training data increases like over fitting it becomes complicated to construct the decision tree. It does not support numeric data and therefore pruning is difficult. Decision trees help to understand a lead in instructive situation, like his/her enthusiasm towards a particular subject or to predict the result in an examination.

4.2. Bayesian classifier

This involves a Naive Bayes algorithm and also implements its variations. This is an easy and simple technique to understand and needs small amount of training data in order to estimate the parameters. It is space efficient and insensitive to irrelevant features. It is capable to handle both real and discrete data [21]. Patterns that are derived using Bayesian Classifier for educational data helps to enhance the decision making in terms of identifying students at risk, to increase students' success, decrease student dropout rate and increase students learning outcome.

4.3. Neural networks

Neural network is another most preferred and commonly used technique in educational data mining. The benefit of using neural network is that it has the ability to identify all possible interactions between the predictor variables. This involves Multilayer Perception algorithm in it [22]. This is a generalized technique and works well with noise. But it gets complicated and does not support scaling from small research system to a large real-time system. It is achievable to design a model with artificial neural network that can predict a candidate's performance based on some given pre admission data for that particular student.

4.4. Support vector machine

Support Vector Machine (SVM) was invented by Vapnik and it is the best tool for machine learning research community. Literature on SVM states that it is proficient in delivering high accuracy in classification than the other data classification algorithms. There are numerous advantages of using SVM since it has greatest marginal hyperactive plane for classifying linearly separable data. In relation with Educational Data Mining, SVM Classifier is valuable technique to provide information to departmental faculty members in making decisions.

4.5. C4.5 tree

The most recent and commonly used decision tree algorithm is the C4.5. Professor Ross Quinlan invented a decision tree algorithm called as C4.5 in the year 1993. It relates to the results of the research in the ID3 algorithm (also proposed by Ross Quinlan in the year 1986). C4.5 came with additional features like categorization of continuous attributes, handling missing values, rule derivation, pruning of decision trees and others. The basic design of C4.5 algorithm uses divide and conquer concept to create a suitable tree from a training set. It is used in educational data mining in order to predict academic performance of students.

5. Comparison of data mining techniques in predicting academic performance of students

This work is a comparison study proposed by Mamta Sharma and Monali Mavani. The study includes three algorithms with respect to prediction of student's result [23]. In 2009 there was another comparison study proposed by Fadzilah Siraj and Mansour Ali Abdoulha. This study includes three techniques in order to understand the undergraduate's student enrolment data. The results of the study were published in their contribution in IEEE [24].

The paper proposed by Pathom Pumpuang, Anongnart Srivihok and Prasong Praneetpolgrang states that Nbtree is the best classifiers to predict student sequences for course registration planning [25]. Pedro G. Espejo, Sebastián Ventura, César Hervás and Cristóbal Romero stated that classifier model is the most appropriate technique in order to predict results in educational field [26]. Decision tree is 3-12% more accurate than the Bayesian network to predict academic performance of students in “A Comparative Analysis of Techniques for Predicting Academic Performance” in the year 2007 [27], IEEE.

6. Conclusion

The study was carried out based on research papers that were published by various authors. The past works focuses on various approaches that are used for the prediction process. Different parameters were considered by the researchers in order to classify student's quality assessment based to their IQ (Internal Quality) and capability factors. Their research also involves various educational data for prediction. The study concludes that there are several efficient data mining methods available for performance analysis.

References

[1] Amirah Mohamed Shahiria, Wahidah Husaina,Nur'aini Abdul Rashida, “A Review on Predicting Student's Performance using Data Mining Techniques”, The Third Information Systems International Conference, Procedia Computer Science 72 pp 414 – 422, 2015.

[2] Umadevi, D.Sundar, Dr.P.Alli,“A Study on Stock Market Analysis for Stock Selection – Naïve Investors' Perspective using Data Mining Technique”, International Journal of Computer Applications (0975 – 8887),Vol 34– No.3, 2011.

[3] NityaUpadhyay, VinodiniKatiyar, “A Survey of the Classification Techniques In Educational Data Mining”, International Journal of Computer Applications,Technology and Research, Vol.3,Issue 11, pp 725 – 728, 2014.

[4] S. K. Yadav and S. Pal, “Data Mining: A prediction for performance improvement of Engineering students using classification”, World of Computer Science and Information Technology Journal (WCSIT), Vol. 2, No. 2, pp51-56, 2012.

[5] Hijazi,S.T., and Naqvi, R.S.M.M., ” Factors Affecting Student's Performance: A Case of Private Colleges”, Bangladesh e-Journal of Sociology, 2006.

[6] A. S. Arunachalam and T. Velmurugan , “A Survey on Educational Data Mining Tools and Techniques,” International Journal of Data Mining Techniques and Applications, vol. 5, no. 2, pp. 167–171, Dec. 2016.

[7] Shanmuga Priya, “Improving the student's performance using Educational data mining”, International Journal of Advanced Networking and Application, Vol.4, pp1680- 1685, 2013

[8] C. Marquez-Vera, C.Romero and S.Ventura, “Predicting School Failure Using Data Mining”,2011

[9] U.K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690.

[10] BakerRSJd, Yacef K, “The state of educational datamining in 2009: A review and future visions”, 2009.

[11] Ahmed, A.B.E.D. and Elaraby, I.S., “Data Mining: A prediction for Student's Performance Using Classification Method”, World Journal of Computer Application and Technology, Vol 2, pp.43-47, 2014.

[12] Sudheep Elayidom, Sumam Mary Idikkula& Joseph Alexander,“A Generalized Data mining Framework for Placement Chance Prediction Problems” , International Journal of Computer Applications, Volume 31– No.3, 2011.

[13] Tripti, Dwivedi Diwakar Singh, “Analyzing Educational Data through EDM Process: A Survey”, International Journal of Computer Applications, Vol 136 ,No.5, 2016

[14] Baker RS,Yacef K, “ The state of educational data mining in 2009: A review and Future visions”. JEDM-Journal of Educational Data Mining, 2009.

[15] Umadevi, D. Sundar, Dr. P. Alli, “An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50”,International Journal of Data Mining & Knowledge Management Process (IJDKP),Vol.3, No.1, 2013.

[16] B. UmadeviD.Sundar, Dr.P.Alli,”An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors”, Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on (IEEE Xplore Digital Library),pg1 -5.,2013.

[17] K.R.Kavyashree, LakshmiDurga,“ A Review on Mining Students' Data for Performance Prediction”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, 2016.

[18] Ryan S.J.d. Baker, “Data Mining for Education”, International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier, 2009.

[19] B. Umadevi,D.Sundar, Dr.P.Alli, “An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50”,International Journal of Data Mining & Knowledge Management Process (IJDKP),Vol.3, No.1, 2013.

[20] B. UmadeviD.Sundar, Dr.P.Alli,”An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors”, Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on (IEEE Xplore Digital Library), p.1 -5.,2013.

[21] K.R.Kavyashree, Lakshmi Durga, “A Review on Mining Students' Data for Performance Prediction”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, 2016.

[22] Sharma, Mamta, and Monali Mavani. "Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector." Advances in Computing, Communication and Control. Springer Berlin Heidelberg, 2011. 189-194.

[23] Siraj, Fadzilah, and Mansour Ali Abdoulha. "Uncovering hidden information within university's student enrollment data using data mining." Modelling & Simulation, 2009. AMS'09. Third Asia International Conference on. IEEE, 2009.

[24] Pumpuang, Pathom, Anongnart Srivihok, and Prasong Praneetpolgrang. "Comparisons of classifier algorithms: Bayesian network, C4. 5, decision forest and NBTree for Course Registration Planning model of undergraduate students." Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on. IEEE, 2008.

[25] Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás. "Data Mining Algorithms to Classify Students." In EDM, pp. 8-17. 2008.

[26] Nghe, Nguyen Thai, Paul Janecek, and Peter Haddawy. "A comparative analysis of techniques for predicting academic performance." Frontiers in Education Conference-Global Engineering: Knowledge without Borders, Opportunities without Passports, 2007. FIE'07.37th Annual. IEEE, 2007.

[27] Dharmarajan, K., and M. A. Dorairangaswamy. “Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm. & quot”, International Journal of Emerging Technologies in Learning (iJET) Vol.12 (10), PP. 202-211 2017