iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VIII **Month of publication:** August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8077

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

Different Techniques used for Educational Data Mining - A Survey

Tanuja Sharma[1], Rajendra Kumar Gupta[2]
[1, 2]*Department of Computer Science & Information Technology, MITS, Gwalior (M.P.) 474005, India*

Abstract: ***Performance prediction is one of the major discussed topics studied in every educational institution. Prediction in the area of choosing the subjects of students' interest can be better viewed through the associations of the data or process of finding inherent regularities. The variety of data or the attributes and algorithm can be used for the prediction models. It is beneficial to the prediction of students' performance with high accuracy because it is helpful to identify the students who have achieve low grades at the early stage of academics. When we study marketing we generally focus on what the customer is going to buy according to what he wants and we recommend him according to it. Likewise, here in educational mining if a student performs good in particular field of subjects and bad in some other kind of subjects. So, we can make that student prepares for those subjects before his performance degrades. Therefore, educational data mining provides an idea to study the dataset of multiple students under various conditions and influences which a student goes through. We can never predict any kind of person though, their present and past academic records including various activities and habits can imply what they are used to how they perform in various tests that can be labelled accordingly. Educational data mining is a prediction technique which extracts information of students from educational institutes like schools, colleges, universities. The whole idea of focusing on educational sector is the one related to discover new reasons or factors involved which are affecting their education.***
Keywords: EDM, Classification, Prediction, ANN, LR, CART, ID3

I. INTRODUCTION

A lot of research using data mining in education is increasing and the technique implemented in this field can usually be known as educational data mining. The work presented here reflects the usage of educational data in different areas which is a variable approach to mine the datasets of any educational institution to discover interesting patterns and rule generation. Data mining technique used in education sector provide us more customized education, reduce the expenses of education process for universities, improved system efficiency. This support us to increase academic achievements, increase the retention rate of student. Apparently, it raises the question of why we are focusing on mining the data in the educational field. The more we get in to any sector of education the more we get involved in it and hence research is the word for it. The whole idea of focusing on educational sector is the one related to discover new reasons or factors involved which is affecting their education. Well, everybody knows education is the most important to survive. Therefore, we are more concerned and more likely to research in the most required field. It also relevant as everyone in the carrier suffers from almost something that stops them to concentrate on their studies. The motivation factor behind this research is the students' grade degradation problems which are result of multiple factors i.e. parents qualification, past academic records, self-study ability etc. As a result of that, their mentor or teacher can't make valuable decisions for them. So, our research helps to find the flaws at an early stage. The main aim of this study is to identify students' relations using multiple classification techniques like linear regression, naive bayes, decision tree based on the datasets of students to determine predictive models to calculate their performance. To develop a set of rules or associations that can be implemented while choosing the subjects in further semesters. These relations can be discovered through a set of predictive models inculcating the whole ideas that are to be implemented in the dataset. To analyse the data and present a higher accuracy model for performance prediction. The accuracy models or the data represented in further sections shows the past background, abilities and their habits. By predicting, the performance we can show the higher predictive models that can be used for higher predictions and formulations that can enhance their future performances. To formulate a hypothesis to check the factors responsible for their performance issues. The regression model helps in plotting predictive graphs, that can help in switching from educational background to their abilities and habits by which, we check what is a eating more and which one is more responsible.

In this paper, we will discuss the different techniques used for educational data mining. The main aim of this study is to improve the students' performance. It is helpful to both teachers and learners.

©IJRASET: All Rights are Reserved

II. REVIEW METHODOLOGY

Berhanu, F., & Abera, A., et.al[1], From the college of Agriculture, Department of Horticulture–Dilla University, student datasets was accumulated. It contains five years period (2009-2014) of dataset. The pre-processing of attributes and experiment was performed with the help of Rapid miner. Dataset have 199 records and 49 attributes. 27 rules are found in order to improve the performance of student. After pre-processing, the decision tree algorithm was used which has presented 84.95% accuracy.

Buniyamin, N., Mat, U. bin, et al.[2], This paper show the value of student dataset in order to improve the education system. This study also suggested method to get information from educational institution. After that, they presented the tool to faculty of institute for identification, classification and prediction of students' academic performance on the basis of CGPA. After perform prediction, university or institute can arrange the assistance or guide. They help those student who have achieved low grades by providing extra tutorial. This paper proposed Neuro-Fuzzy classification technique to perform prediction of students' performance for Electrical Engineering students in Malaysian public university.

Costa, E. B., Fonseca,et.al.[3] This the study was performed a comparision of four techniques of EDM (Support vector machine, decision tree, neural network and naïve bayes). In this data sources are taken from on-campus and distance education from a Brazilian university: one comes from distance education and the other from on-campus. When the experiment is conducted, They also used fine-tuning tasks and pre-processing.

T.Devasia et al.[4], conducted a research by using 19 attributes on 700 undergraduate students in Amrita Vishwa Vidhyapeetham, Mysore using Naïve Bayes classifier. This proposed system makes use of Naïve Bayesian technique to extract useful information and this system was a web based application. The main motive behind this research is to increase the success graph of students by using system that maintains the course details, attendance details, subject details, marks details of all student.

Amjad Abu Saa [5], This research has explored various factors that has great impact on students' performance and presents a qualitative model to predict students' performance on the basis of social and personal factors. In this research, firstly a survey has conducted in order to collect students' academic and personal data from university. After that, Pre-processing has performed to make dataset suitable for performing prediction. After completion of pre-processing, implementation of various technique has been performed then performed testing. There are four decision tree algorithm C4.5, ID_3, CART and CHAID used in this study. CART gives the best accuracy. After that, naïve bayes classifier is applied which gives accuracy of 36.40%. It has been analysed that performance of student is not only affected by academic achievement but also various factors.

Amirah Mohamed Shahiri et al. [6], conducted a review to show an overview of prediction techniques of data mining on a large amount of data in Malaysia for predicting the students' performance. This study focused on the way in which various algorithm can be implemented to find the most essential attributes in a dataset of student. The whole dataset compared the various data mining techniques. This study also suggested a table with attributes for common methods. These methods are used together to give 100% results. This study presents a scenario for neural network, naïve bayes, K-nn, decision tree and support vector machine. It showed the prediction accuracy which used the classification method grouped by prediction algorithm since 2002 to 2015. Neural network performed to be the best of all. The attributes included the students' past and present background of academic career, psychometric factors, Internal and external assessments.

Thi-oanh Tran et al. [7], conducted a research on students of information technology in Vietnam National University, Hanoi based on multiple strategies used in a recommender system based approach using regression. They implemented on contributing the certain steps in building a dataset of students consisting of their academic background and investigating on the methods to enhance the performance of student, designing course related skills that was used in regression based models and proposed a hybrid method to give the best result by combining the best outputs. It used linear regression, ANN, decision tree and support vector machine. The results of this experiment have proved that unlike the students' performance prediction in E-learning system, the performance of regression based approach is good in comparison of the recommender system-based approach. To integrate the proposed features is also helpful for enhancing the performance of regression-based system. Finally, we can say that this proposed hybrid method got the excellent RMSE score of 1.871 for targeted elective courses and 1.668 for hybrid approach.

Surjeet Kumar Yadav et al. [8], conducted a research on engineering students using multiple decision tree algorithms which resulted in comparative analysis of prediction algorithm to show the better results in their upcoming semester. The multiple decision tree algorithms like CART, C4.5 and ID3 are applied for predicting the performance of student in the end semester or final exam. Their objective clearly stated that the steps included the generation of dataset with predictive variables, identification of the students learning behaviour, construction of predictive model using classification and lastly, validation of the model with universities which can be applied to all the institutions. The outcomes of decision tree found the probability of total students to fail and pass (promoted

©IJRASET: All Rights are Reserved

to next year). The result provides the way to enhance the students' performance that were predicted to promoted or fail. The outcomes prove that C4.5 likely, proved to be the best with 67% correctly classified instances compared to CART and ID3.

Pauziah Mohd Arsad,et al.[9], conducted a research on matriculation and diploma students to predict their academic performance at semester eight by using comparison between ANN (artificial neural network and LR (linear regression). Academic achievement of 8th semester was measured in the form of CGPA. This research was conducted in Universiti Teknology MARA (UiTM), Malaysia. The first semester's result of student was used as input predictor variable or independent variable instead of this the final semester's result (8th sem.) was used as dependent variable or output. The coefficient of correlation R was used for measuring the performances of the models. The outcomes of ANN and LR show a strong correlation between the first semester CGPA with the final CGPA. It compared the residuals of the methods executed in SPSS that represented actual and predicted values on a graph. It also Performed the comparison of the correlation coefficient for both values. Henceforth, it resulted in different cases for different prediction models.

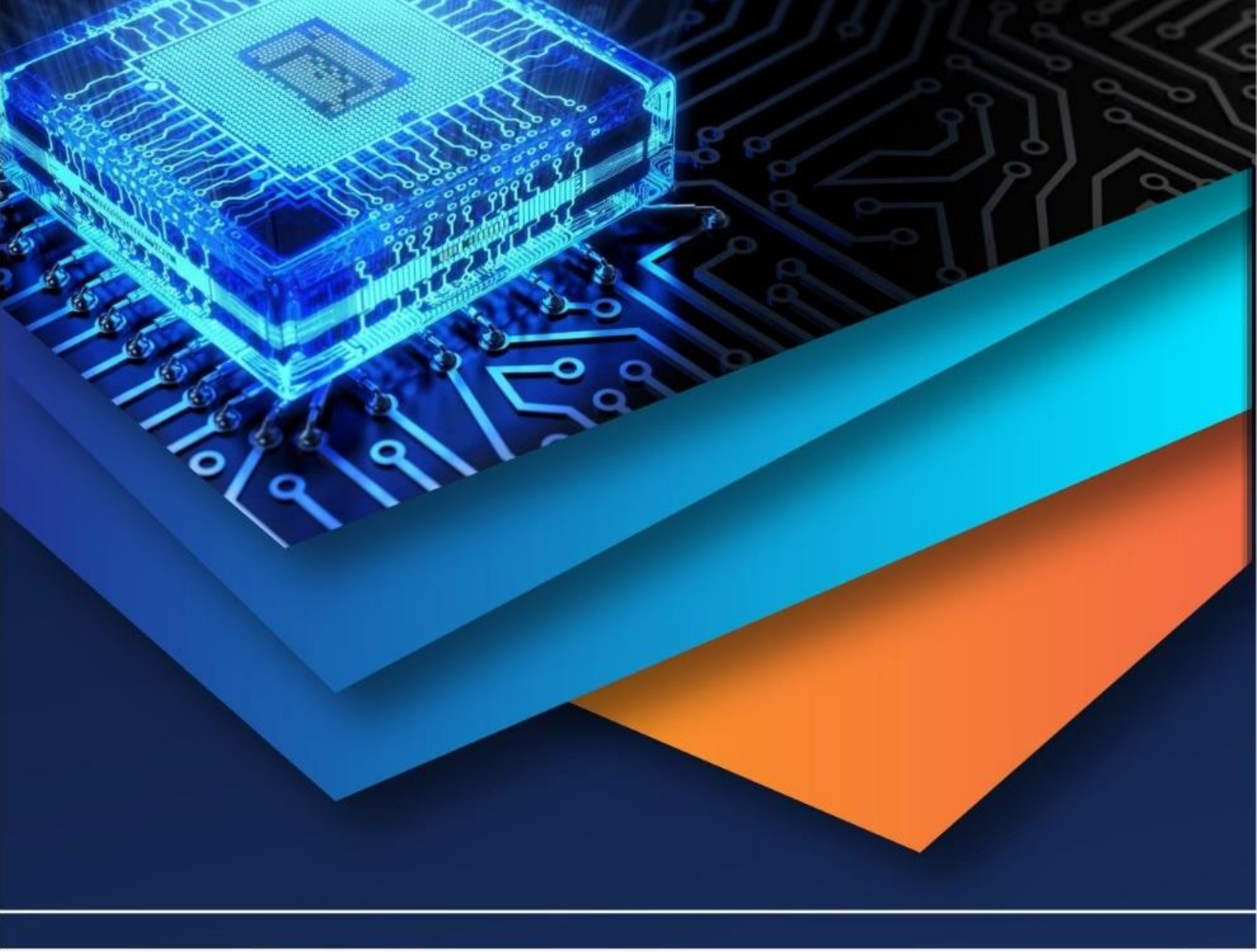
III. CONCLUSION

Mostly paper discussed here presented the different data mining technique in order to improve the student's performance prediction. These papers are related to academic drop out. The main aim of all papers is to predict that how many students get success or failure. For this, researcher collected the previous academic record, social factors of student. The advantage of educational data mining is that student can identify if they are at risk point or not. This is also beneficial for universities in order to provide the extra tutorial to weak students. Different papers used different technique to perform classification and prediction. The accuracy of different classifier are compared. The algorithm which has achieved high accuracy are selected to perform prediction. Some researchers performed prediction at first semester while some researchers performed at middle semester and end semester. So, the difference is only in duration. All papers used GPA as a main attribute to perform prediction.

REFERENCES

[1] Berhanu, F., & Abera, A.," Students' Performance Prediction based on their Academic Record", International Journal of Computer Applications,vol.131(5), pp. 27-35, 2015.
[2] Buniyamin, N., Mat, U. bin, & Arshad, P. M., "Educational data mining for prediction and classification of engineering students achievement" IEEE 7th International Conference on Engineering Education (ICEED), pp. 49-53, 2015.
[3] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior", vol.73, pp.247–256, 2017.
[4] Devasia, T., Vinushree T P, & Hegde, V., "Prediction of students performance using Educational Data Mining", International Conference on Data Mining and Advanced Computing (SAPIENCE), pp. 91-95, 2016.
[5] Amjad Abu Saa, " Educational data mining & students performance prediction. International journal of advanced computer science & application" vol.7(5), pp.212- 220, 2016
[6] Amirah Mohamed Shahiri, Wahidah Husain, et al., "A review on predicting student's performance using data mining techniques", Procedia computer Science vol.72 pp. 414-422, 2015
[7] Thi-Oanh Tran, Hai-Trieu Dang, Viet-Thuong Dinh, Xuan-Hieu Phan, et al, "Performance prediction for students: A multi strategy approach" Cybernatics and Information Technologies, vol.17(2) ,164-182, 2017.
[8] Surjeet Kumar Yadav and Saurabh Pal., " Data mining: A prediction for performance improvement of engineering students using classification", world of computer science and information technology(WCSIT) vol.2 pp. 51-56, 2012).
[9] Arsad, P. M., Buniyamin, N., & Manan, J. A., "Prediction of engineering students academic performance using artificial neural network and linear regression: A comparison", IEEE 5th conference on Engineering Education(ICEED), pp. 43-48, 2013.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)