International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

A STUDY ON STUDENT CAREER PREDICTION

Anooja S K[1], Dileep V K[2]

[1]*Student Dept. of Computer Science and Engineering, LBSITW, APJ Abdul Kalam University, Kerala, India*
[2]*Assistant Professor Dept of Computer Science and Engineering, LBSITW, Kerala, India*

Abstract - *Educational Data Mining (EDM) and machine learning has become an inevitable technologies in past years. Most of the educational systems has adapted many technologies to improve the performance of students. Nowadays the rate of failures are increasing. In order to improve the performance of students educational institutions adapt many techniques. In this article, two important factors are focused on: Firstly, to identify the major factors which affect the student performance and secondly to find the algorithm which is mostly used for the prediction techniques and to check the accuracy levels obtained by each classification techniques.*

Key Words: Educational Data Mining, Student's performance, prediction, Machine Learning, Naïve Bayes, Clustering, Classification, Artificial Neural Network

1. INTRODUCTION

Academic performance of students has always been a major factor for determining the student's career and the prestige of the Institutions. For this purpose Education Data Mining (EDM) is used. The applications such as model development helps to predict student performance in their academics. Therefore, the researchers had to dig deep into various methods in data mining to improve existing method. The applications of Machine Learning methods to predict students' performance based on the background of student's performance and evaluation marks. It will leads to the detection of high caliber students in the institution and help them for providing scholarships. Machine learning algorithms such as Decision Tree [10] and Naive Bayes [9] is highly used in Educational Data Mining. But they had certain limitations stated by Havan Agrawal [11] when input is provided in a continuous range to Bayesian classification the accuracy of the models reduces. Such classification works better with discrete data. Also stated that a Neural Network outperforms when given a continuous data.

Data mining is one of the most important technique adapted by most of the researchers. It will discover the

data automatically for large repositories and give the better result. Naïve Bayes, Regression, Classification, K -means etc. are some of the algorithms used by the previous scholars. Among these neural network gives more accurate result. This review is for finding the methods used for the prediction of student's performance and to check the attributes which are commonly used.

2. LITERATURE REVIEW

A study conducted by Ioannis E. Livieris, et al. [1] to predict the performance of students in Mathematics use ANN(Artificial Neural Network)[17]. It can be found that the modified spectral Perry trained artificial neural network performs better classification compared to other classifiers.

S. Kotsiantis, et al. [2] investigated in distance learning of machine learning techniques [18] for dropout prediction of students. Important contribution was made by this study was a pioneer and helped to carve the path. Machine learning techniques were first applied by him and his team in an academic environment. An algorithm was fed on demographic data and several project assignment rather than class performance data to make prediction of students.

Moucary, et al. [3] applied a hybrid technique on K Means Clustering [19] and Artificial Neural Network for students who are pursuing higher education. Firstly, Neural Network was used to predict the performance of student and then it will be fitted to a particular cluster such as K-Means algorithm. This clustering helped for the instructors to identify a student capabilities during their performance in the academics.

A prediction model for students' performance Amrieh, et al. [12] proposed based on data mining methods. In addition to the previous work he include the behavioral condition of the student. The classifiers such as Naïve Bayesian [20], Artificial Neural Network and Decision tree [21] are used for classification. The ensemble methods such as Random Forest, Bagging and Boosting [22] were used to improve the performance. The model achieved up to 22.1% more in accuracy compared when behavioral features were removed. It increased up to 25.8% accuracy after using the ensemble methods.

Ramaswamy and Rathinasabapathy, et.al [10] used Bayesian network approach to predict the overall performance of student. In this study the data contain 35 attributes with 5650 records of HSC grade. Based on the two-case (pass, fail), three-case (very good, good, poor) the feature selection is performed. Using the software WEKA [16](data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data or called from the Java code) they estimate the algorithm. The result showed that the Bayesian network models with Network Augmented with Tree search algorithm achieves better performance.

A study on the prediction of student performance by Angeline D M, et.al [11] used apriori algorithm which extracts the set of rules specific to every category and analyze the given knowledge to classify assignment, internal assessment, group action etc. It categorize the given set into average, good, below average.

Jeevalatha, Ananthi and SaravanaKumar, et.al [14] presented the performance analysis for placement selection. They work with decision tree algorithm using the factors like HSC, UG marks and communication.

Dinesh A and Radhika V, et.al [13] targeted on the techniques and strategies of institutional data processing for data discovery. This paper suggest the relation mining between 1995 and 2005 and in 2008 to 2009. During this period 45% papers are for prediction. The prediction model acts as a warning system to improve the performance.

Mueen et.al [8] studied educational datamining to predict student performance. The algorithm such as decision tree, back-propagation etc. are used for the measurement and comparison. The data can be collected from the university GPA and it is to be performed in WEKA tool. It shows the number of instances is much smaller than the number of instances in other class. And it has a prediction accuracy of 86%.

Al.Radaideh et.al [6] predict the performance of technology and computer science faculty who took C++. Based on the attributes such as gender, age, department etc. it will analyze the result. It is to be worked on WEKA. It indicate the collected samples and attribute were not sufficient to generate a classification model of high quality.

An early identification of student dropout by Baradwaj and Pal el.al [5] use decision tree in the information like attendance, class test, semester and assignment marks. It showed or predict the performance into average, poor and good.

Mythili M S and Shanavas A R et.al [4] applied the algorithms such as J48, Randomforest, Multilayer Perceptron and Decision tree which is collected from student management system and is analysed and evaluate to get the maximum satisfied output. It is worked under the platform of WEKA.

Noah Barida and Egerton et.al [15] studied and evaluate the performance of student by grouping the grading into various classes using CGPA. They used the methods like neural network, regression and K-means to identify the weak performers in the group of data. It will obtain an accuracy of 83.6%.

Ramesh, Parkavi and Yasodha et.al [7] conduct a study on the prediction using the algorithms such as Naïve Bayes, Multilayer perceptron, SMO, J48 and REP on the placement details. From the result it can be concluded that MLP is more suitable than other algorithm.

Bhise, Thorat and Supekar et.al [9] present the method using K-means clustering algorithm. It mainly focused on the drop out ratio of the students and improve it by considering the evaluation factors like midterm and final exam assessment test. Using different clustering techniques namely hierarchical, partitions and categorical.

Bendangnuksung and Dr. Prabu P et.al [17] present the prediction of student performance using deep neural network. In this paper they predict the performance of students whether they fall under fail category or pass category through logistic classification analysis. The proposed deep neural network model achieved up to 84.3% accuracy and outperforms other machine learning algorithms in accuracy.

Lubna Mahmoud Abu Zohair et.al [16] suggested prediction of student's performance in educational entities and institutes. In this paper they proposed that, in order to help at-risk students and assure their retention, providing the excellent learning resources and to improve the results. So, the main aim of this project is to prove the possibility of training and modeling a small dataset size and the feasibility of creating a prediction model with credible accuracy rate. This research explores as well the possibility of identifying the key indicators in the small dataset, which will be utilized in creating the prediction model, using visualization and clustering algorithms. Best indicators were fed into multiple machine learning algorithms to evaluate them for the most accurate model. Among the selected algorithms, the results proved the ability of clustering algorithm in identifying key indicators in small datasets. The main outcomes of this study have proved the efficiency of support vector machine and learning discriminant analysis algorithms in training small dataset size and in producing an acceptable classification's accuracy and reliability test rates.

Table1—Comparison of various machine learning techniques used for student performance prediction

COMPARATIVE STUDY			
YEAR	AUTHOR	TITLE & METHOD	REMARKS
2008	S Kotsiantis	Preventing student dropout in distance learning systems using machine learning techniques Applied Artificial Intelligence	Machine learning technique was implemented by him and his colleagues
2011	V.Ramesh, P.Parkavi and P.Yasodha	Performance analysis of data mining	Conclude MLP get more

		techniques for placement chance prediction	accurate result that J48, RMO
2012	Ramaswami,M and R.Rathinasabapathy	Student prediction	The result showed that the Bayesian network models with Network Augmented with Tree search algorithm achieves better performance.
2013	Bhise R B, Thorat S.S and Supekar A.K	Importance of data mining in higher education system	Mainly focused on the drop out ratio of the students and improve it by considering the evaluation factors like midterm and final exam assessment test
2016	Amrieh, E.A., Hamtini, T. and Aljarah	Mining Educational Data to Predict Student's academic Performance using Ensemble Methods	The model achieved up to 22.1% more in accuracy compared when behavioral features were removed. It increased up to 25.8% accuracy after using the ensemble methods.
2018	Bendangnuksung and Dr. Prabu	Students' Performance Prediction Using Deep Neural Network	Achieved up to 84.3% accuracy and outperforms other machine learning algorithms in accuracy.
2019	Lubna Mahmoud Abu Zohair	Prediction of Student's performance by modelling small dataset size	Proved the efficiency of support vector machine and learning discriminant analysis algorithms

3. CONCLUSION AND FUTURE WORK

The machine learning and datamining techniques used in related research work doesn't provide an accuracy of above 87%. And the possibility of misprediction is also occur. In order to overcome this situation the future works can be implemented in deep neural networks. Since it is a multi-hidden layer network the result can be of greater accurate than the previous ones and the fitting techniques can be

easily done. The previous ones and the fitting techniques can be easily done in DNN

REFERENCES

[1] Livieris, Ioannis, Tassos Mikropoulos, and Panagiotis Pintelas. "A decision support system for predicting students' performance." *Themes in Science and Technology Education*9.1 (2016): 43-57.

[2] Kotsiantis, Sotiris B., C. J. Pierrakeas, and Panayiotis E. Pintelas. "Preventing student dropout in distance learning using machine learning techniques." *International conference on knowledge-based and intelligent information and engineering systems*. Springer, Berlin, Heidelberg, 2003.

[3] Moucary, C. El, Marie Khair, and WalidZakhem."Improving student's performance using data clustering and neural networks in foreign-language based higher education." *The Research Bulletin of Jordan ACM* 2.3 (2011): 27-34.

[4] Khasanah, Annisa Uswatun. "A review of student's performance prediction using educational data mining techniques." *Journal of Engineering and applied sciences* 13: 5302-5307.

[5] Bhardwaj, B. K., and S. Pal. "Data Mining: A prediction for performance improvement using classification (IJCSIS) *International Journal of Computer Science and Information Security 9*." (2011).

[6] Al-Radaideh, Qasem A., Emad M. Al-Shawakfa, and MustafaI. Al-Najjar. "Mining student data using decision trees." *International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan*. 2006.

[7] Ramesh, V., P. Parkavi, and P. Yasodha. "Performance analysis of data mining techniques for placement chance prediction." *International Journal of Scientific & Engineering Research* 2.8 (2011): 1.

[8] Mueen,A,B.Zafar and U.Manzoor,Modeling and predicting academic performance using data mining techniques.*International Journal of Modern Education & Computer Science* (2016):36-42

[9] Bhise, R. B., S. S. Thorat, and A. K. Supekar. "Importance of data mining in higher education system." *IOSR Journal of Humanities And Social Science (IOSR-JHSS)* 6.6 (2013): 18-21.

[10] Ramaswami,M and R.Rathinasabapathy,2012.Student performance prediction *International Journal of Computational Intelligence and Informatics* (2012):231-235.

[11] Alangari, Njoud, and Raad Alturki. "Association Rule Mining in Higher Education: A Case Study of Computer." *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies* (2019): 311.

[12] Amrieh, Elaf Abu, Thair Hamtini, and Ibrahim Aljarah. "Mining educational data to predict student's academic performance using ensemble methods." *International Journal of Database Theory and Application* 9.8 (2016): 119-136.

[13] Kumar, A. Dinesh, and Dr V. Radhika. "A survey on predicting student performance." *International Journal of Computer Science and Information Technologies* 5.5 (2014): 6147-6149.

[14] Jeevalatha, T., N. Ananthi, and D. Saravana Kumar. "Performance analysis of undergraduate students placement selection using decision tree algorithms." *International Journal of Computer Applications* 108.15 (2014).

[15] OTOBO Firstman Noah, BAAH Barida and Taylor Onate Egerton, "Evaluation of student performance using data mining over a given data space", *International Journal of Recent Technology and Engineering* (2013):4

[16] Zohair, Lubna Mahmoud Abu. "Prediction of Student's performance by modelling small dataset size." *International Journal of Educational Technology in Higher Education* 16.1 (2019): 27.

[17] Zohair, Lubna Mahmoud Abu. "Prediction of Student's performance by modelling small dataset size." *International Journal of Educational Technology in Higher Education* 16.1 (2019): 27.

[18] Singhal, Swasti, and Monika Jena. "A study on WEKA tool for data preprocessing, classification and clustering." *International Journal of Innovative technology and exploring engineering (IJItee)* 2.6 (2013): 250-253.

[19] Zurada, Jacek M. *Introduction to artificial neural systems*. Vol. 8. St. Paul: West, 1992.

[20] Ramab, Admir, et al. "Distance learning at biomedical faculties in Bosnia & Herzegovina." *Connecting Medical Informatics and Bio-informatics: Proceedings of MIE2005: the XIXth International Congress of the European Federation for Medical Informatics.*"" Vol. 116. IOS Press, 2005.

[21] Kaur, Navjot, Jaspreet Kaur Sahiwal, and Navneet Kaur. "Efficient k-means clustering algorithm using ranking method in data mining." *International Journal of Advanced Research in Computer Engineering & Technology* 1.3 (2012): 85-91.

[22] Lewis, David D. "Naive (Bayes) at forty: The independence assumption in information retrieval." *European conference on machine learning*. Springer, Berlin, Heidelberg, 1998.

[23] Myles, Anthony J., et al. "An introduction to decision tree modeling." *Journal of Chemometrics: A Journal of the Chemometrics Society* 18.6 (2004): 275-285.

[24] Hastie, Trevor, et al. "Multi-class adaboost." *Statistics and its Interface* 2.3 (2009): 349-36