

A Machine Learning Framework for Identifying Learning Levels of Students in Higher Education

Qamar Rayees Khan¹, Muheet Ahmed Butt²

¹Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri (J&K), India

²Department of Computer Sciences, University of Kashmir, Srinagar, (J&K), India

ABSTRACT

The Educational data Mining (EDM) is one of the challenging and emerging fields in the area of computer sciences that analyze the data as per the requirement of the education/data analyst. The need of the EDM is of paramount importance wherein the data that pertains to the students are critically analyzed and the inferences are drawn. This paper will explore the area and propose a Machine Learning Framework for Identifying Learning Levels of Students in Higher Education. The main aim of this paper is to identify the slow learners in the education institutions by extracting the patterns from the data so that the appropriate strategies may be adopted thereafter to bridge the learning gap among the students.

Keywords : Education Data Mining (EDM), education/dataanalyst, Machine Learning, slow learners.

I. INTRODUCTION

The higher education in india comprises of universities and colleges that are span in all the states and Union Territories. Enrollment ratio in India is 82.3. There are about 760 universities and 38498 colleges providing education to the students. Jammu and Kashmir alone comprises of 09 state universities, 02 central universities, 01 Indian Institute of Technology (IIT), 01 Indian Institute of Management (IIM), 01 Indian Institute of Mass Communication (IIMC), 02 National Institute of Technology (NIT), 08 Medical Colleges, 09 Engineering colleges, 04 research institutions, 11 other prominent colleges and 142 private and government colleges.

The students after getting admission in the colleges are seen performing below their expected levels as has been revealed by the researchers while exploring and analyzing the educational data pertaining to the students in higher education. One of the biggest

challenge is to identify the learning level/ ability of the students so that a proper mechanism may be put in place to fill the learning gap among the slow learners. The problem needs to be addressed by analyzing the complete data of the students so that the parameters that affect their learning abilities may be tapped beforehand. The role of data mining and artificial Intelligence in pedagogical domain seem to be challenging[1]. From so many years, researchers try to analyze the education data and make an effort to extract new patterns and draw inferences from the same[2]. The new and emerging field of data mining ie: Education Data Mining (EDM) emerges where techniques are being explicitly used in the education data so as to formulate and derive the results[3][4].

II. Related Work

BK Bhardwaj et. al. in their paper focus on the data mining techniques for evaluating students performance enrolled in higher education . they have

proposed a model wherein they used various data mining techniques to handle the problem. They have used Decision tree algorithm for this classification problem. The results shows that the students performance is identified at the end of the semester examination. The key purpose of this research was to reduce the overall dropout ratio of the students and at the same time enhance the performance level of the students by appropriate counseling/advising.[5]

M. Ramaswamiet. al. in their study made a study of the educational Data Mining where they generate a database of the students the data of which was collected from students directly (primary data) using questionnaire and secondary data from other sources like schools, education officers of the district. The researchers pre-process the data and refine the same by removing null values, outliers, missing values etc . CHAID Predication model was used based on prediction rules . the efficiency and accuracy of this model was found satisfactory[6]

ParneetKaur et. al. in their paper have used predictive data mining model for identifying the slow learners enrolled in high school. The classification problem they have solved used dataset collected from high school. Various classification algorithms were used in this paper like Naïve Bayes, Multilayer Perceptron, SMO, REPTree and J48. The experimentation was performed using weka tool. The performance analysis of all the algorithms were evaluated and it emerged that Multilayer Perceptron performs better than others. [7]

Harwatia et. al. in their paper have used K-mean clustering algorithm to extract the hidden patterns from the data and classifying the student based on the demographic parameters and the average attendance in course. They used SPSS for the analysis and the results show four types of clusters based on six parameters.[8]

V.Rameshet. al. in their research paper identified the various factors that affect the performance of the

student in the end semester examinations. They used various Data Mining techniques to predict the performance of the students beforehand so that timely warning may be given to such students. The primary data was collected from the schools and it has revealed that type of the school does not influence students performance but parents occupation was seen as a contributing factor in students performance. [9].

II. Proposed Methodology

The Intelligent Machine Learning Model is proposed for identifying the learning level of the students enrolled in different semesters. Since, our concern is more specific to identify slow learners, so that the institutions may be able to bridge the learning gap among them for better performance afterwards. There are various phases in the proposed model

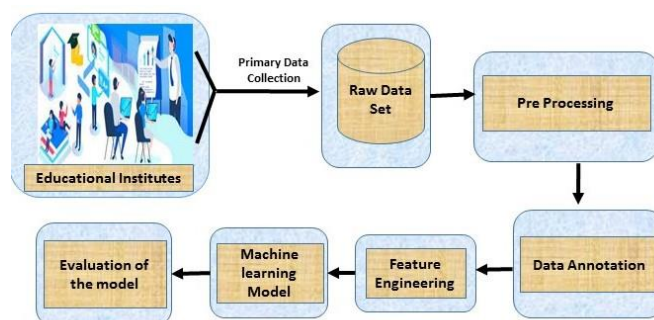


Figure. 1 : Proposed Machine Learning Model

1. Data Collection Phase: Primary data is collected from the college(s) by taking the various important / deciding parameters that may have influence on the learning levels of the students.
2. Pre-Processing Phase: The data collected from the source is raw and need to be refined/polished so that null values, missing values, outliers, consistency issues if any are resolved.
3. Data Annotations: Before the refined data is feed into the Machine Learning Model, data annotations need to be carried out so that the model shall be trained with the training data. The annotations are carried out by labeling the class labels after consulting the education experts.

Annotation scheme is to be framed and the data is annotated as per the scheme.

4. Feature Engineering Phase: The refined data that is the output from the Data Annotation phase is given to the specified algorithms for the feature selection. The attributes that show the influence is selected and the attribute that does not contribute to the results are eliminated. By using these algorithms, we shall be able to reduce the number of attributes that shall reduce the time for training the model.
5. Machine Learning Techniques: The data after proper annotations is used by the Machine Learning Algorithms for training the model by using the various classifiers. The model is tested by using the testing data. The appropriate results are generated for necessary validation of the model.

III. Research Outcome

The outcome of this proposed model is to identify the slow learners in the educational institutions so that the institutions may be able to bridge their learning gap for better performance in their course.

IV. Conclusion

In this paper, various techniques of Machine Learning are used for identifying learning levels of the students in higher education. The paper focuses on the slow learners so that their learning ability may be enhanced by adopting the measures and strategies to bridge their learning gap. In future, we shall implement the framework by using the primary data that need to be collected from various higher educational institutions.

V. REFERENCES

- [1]. Suhas G. Kulkarni, Ganesh C. Rampure, Bhagwat Yadav, —Understanding Educational Data Mining (EDM)‖, International Journal of Electronics and Computer Science Engineering, 2013.
- [2]. Siti Khadijah Mohamad and ZaidatunTasir, "Educational data mining: A review", Procedia Social and Behavioral Sciences, vol. 97, pp. 320-324, November 2013.
- [3]. Cristobal Romero and 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, vol. 40, no. 6, pp. 601-618, November 2010.
- [4]. K. Umamaheswari and S. Niraimathi, "A Study on Student Data Analysis Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 8, August 2013.
- [5]. BK Bhardwaj, S. Pal, " Mining Educational Data to Analyze Students Performance", (IJACSA), Vol. 2, No. 6, 2011
- [6]. M. Ramaswami and R. Bhaskaran (2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, pp 10-18.
- [7]. ParneetKaura, ManpreetSinghb ,Gurpreet Singh Josanc "Classification and Prediction based DataMining Algorithms to Predict Slow Learners in Education Sector" Science Direct Procedia Computer Science 57 (2015) 500 –508 2015 (ICRTC- 2015).
- [8]. Harwatia, ArditaPermataAlfiania, FebrianaAyuWulandari," Mapping Student's Performance Based on Data Mining Approach", Science Direct Agriculture and Agricultural Science Procedia3(2015) 173 – 177.
- [9]. V.Ramesh Assistant Professor Department of CSA, SCSVMV University Kanchipuram India "Predicting Student Performance: A Statistical and Data Mining Approach", International Journal of Computer Applications (0975 – 8887) Volume 63– No.8, February 2013.

Cite this article as :

Qamar Rayees Khan, Muheet Ahmed Butt, "A Machine Learning Framework for Identifying Learning Levels of Students in Higher Education", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 3 Issue 1, pp. 582-584, January-February 2017.
Journal URL : <https://ijsrset.com/IJSRSET218251>