

Students Performance Prediction through Educational Data Mining - An Uncomplicated Review

V. Sathya Durga¹, J. Thangakumar²

¹Research Scholar, Department of Computer Science, Hindustan Institute of Technology and Science, India

²Associate Professor, Department of Computer Science, Hindustan Institute of Technology and Science, India

Abstract - Educational Data mining (EDM) is a new active research area. It focuses on applying data mining methods to the educational dataset to discover useful information. One of the most interesting tasks of EDM is to predict the performance of students. Many researchers all over the world have experimented and published their research work on students' performance prediction. The main objective of this survey paper is to bring out the methodology used by the researchers and the key findings of those research work in a simple and an uncomplicated way.

Key Words: Data Mining, Educational Data Mining, Performance Prediction, Neural Network, Decision Tree

1. INTRODUCTION

Educational data mining is a prominent research area which harvests data coming in from academic set up and integrates methods from multidisciplinary areas like data mining, machine learning and statistics [1]. Knowledge discovered through EDM activities can be used by a wide spectrum of users like students, teacher, course outliners, management and all those integrated into the educational stream.

Typical Educational Data Mining Activities include,

- Profiling and categorizing students
- Determining students learning style
- Finding courses that are taken together
- Predicting student's future performance

This paper surveys scholastic works of fellow researchers on predicting the academic performance of students in a simple and elegant manner.

2. LITERATURE REVIEW

1. Anjali B Raut and Ankita A Nichat (2017) present a student performance system which is based on the common classification technique, the Decision Tree. This system models an online examination, where students take up an exam and instant result (Pass/Fail) along with weak concepts are displayed. Generalization sequential pattern mining algorithm is used for result determination. The decision tree generated by C4.5 is used to analyze the performance of students and classify them based on their marks. The author's states that this analysis using data mining will help the

management to identify weak students and provide extra coaching before the final exam. [2]

2. Ali Daud (2017) develop a prediction model based on learning analytics which predicts student's academic performance. Attributes like family expenditure, family income and family assets were included in this research work. To evaluate metrics like Precision Rate, Recall and F1-score were used. Five fold cross validation was used. Classifiers used are Naïve Bayes, SVM, C4.5 and Decision Trees in this work. Data was collected from graduate and undergraduate courses from different universities of Pakistan. Results reveal that SVM classifier outperformed other classifiers. Some interesting revelation of this research work is as follows; family expenditure affects student's performance, married students perform better than bachelor students and house condition of students influences the performance of the students. [3]

3. Tismy Devasia, Vinushree T P and Vinayak Hegde (2016) conduct a study to predict students' performance in the upcoming exam based on academic history. A Web-based application is developed. Nineteen attributes of 700 students are taken as input. When student's marks are entered, it is compared with existing student's attributes and Naïve Bayes classifier is applied to predict the final grade. It is noted mother's qualification and income of family have a high correlation with the performance of students. Generating data sources, identifying factors affecting performance, constructing a predictive model and validating the model are steps proposed in developing a model for academic prediction. The authors states that this model will help in minimizing the failure ratio and aids in taking acceptable action towards low performer. [4]

4. Hashmia et.al (2016) develop a new mining model to predict the performance of students. It is stated that EDM aims in discovering hidden knowledge and patterns from student's performance. Fuzzy genetic algorithm and decision tree algorithm were applied independently on selected students attribute like internal mark, admission score and session marks. A decision tree was implemented with C4.5 algorithm on the dataset which consists of marks secured by 120 under graduate and 48 post graduate students. From the results, we can infer that students were classified into at-risk and safe categories. Furthermore, the authors talk about education as a tool for developing a country and prediction tools like these will help teachers take early actions to improve students' performance. [5]

5. Author Kavipriya (2016) reviews different classification methods which are commonly used in prediction. These methods intake previous exam data and predict the performance of the upcoming exam. Classification based techniques are discussed in this work. Comparing the accuracy rate of these techniques, it was found out Support Vector Machine had high accuracy. The author states that training a model takes 60% of the time, which can be improved by using ensemble clustering. Better results can be achieved if the models can be trained with real-time data. [6]

6. Masheal A. Al-Barak and Muna Al-Razgan (2016) present a system which uses data mining and predicts students' performance. In this work, the classification technique is used and the final grade is predicted based on the grades obtained in a few courses. Courses which impacts final GPA is identified. The dataset consists of 236 records of students from King Saud University. J48 decision tree algorithm was applied on the dataset and final grades of all students were obtained. Interestingly it was found out that marks obtained in courses like Software Engineering, DBMS, Information Security had a high impact on student's final grade. [7]

7. K. Prasada Rao, M.V.P. Chandra Sekhara and B. Ramesh (2016) proposes a classification based learning model for predicting students' performance. This works is of two parts, the first part focuses on predicting the performance of Computer Science & Engineering students using classification techniques. Students are classified as excellent, good, average and slow learner. The next part, analyses the accuracy and model building time of J48, Naïve Bayes and Random Forest algorithm. Two hundred records of rural-based Computer Science & Engineering students formed the dataset of the study. Results showed that Random Forest has much better accuracy than other algorithms but it requires more time to construct the model. [8]

8. Praneet Kaur, Manpreet Singh and Gurpreet Singh Josan (2015) uses a classification algorithm to identify slow learners among students and display them using predictive data mining models. Through an extensive search of literature, authors identify factors which influence the performance of students. These factors were taken as input variables. Five classification algorithms MLP, Naïve Bayes, SMO, J48 and Reptree were applied on datasets of high school students. It was found out MLP outperformed other classifiers with 75% accuracy. Results reveal those students who had a computer and internet at home performed well in exams. Authors feel that in future if educational data mining is integrated with DBMS and E-Learning accuracy of prediction method can be enhanced easily. [9]

9. A.P.Tribhuvan and Gade J.G (2015) implements ID3 Decision Tree and Naive Bayes techniques to predict student's academic performance. They try to explore the accuracy of the two techniques with respect to prediction. The dataset consists of marks of 100 UG students of Information Technology Department from a private educational institution. The collected data is checked for missing value as ID3 can't handle missing values effectively.

After pre-processing, the predictive model is built. Results show that ID3 achieves 92% accuracy for 50 instances which is relatively higher than Naive Bayes classifier. [10]

10. Anuradha and T. Velmurugan (2015) have developed a new system to predict final exam results of students. The proposed system uses data collected from three engineering college students. They apply classification techniques for prediction. The output of their experiment revel Naïve Bayes classifier performs well than other classifiers. Author's states that data mining will improve the status of educational institute and the performance of students. [11]

11. Jai Ruby and David (2015) build a prediction model based on Multi-Layer perceptron algorithm. Datasets consisted of 165 records with academic, personal and economic attributes. The average accuracy achieved with all attributes was 52% and with selected attributes was 33%. [12]

12. G.Narasinga Rao and Srinivasan Nagaraj (2014) present a new method to predict final exam marks of students using midterm exam marks. They use straight line regression to build the prediction model. The dataset consists of 49 records of students from reputed institutions. Method of least square was applied on the dataset and final exam marks were predicted. By comparing the actual and the predicted value through a graph, we can infer that error rate is very less. This paper justifies that line regression is an effective method for performance prediction. [13]

13. Azwa et.al (2013) develop a prediction model which predicts first-year computer science students performance. They use Naïve Bayes classifier to build their prediction model. Data for this research work was provided by the Information Technology Center, UniSZA. Results reveal that Naïve Bayes classifier gives 57.4% accuracy. Analyzing the result, it was found that female students of low family income performed better in their studies. [14]

14. Mahendra Tiwari, Randhir Singh and Neeraj Vimal (2013) use different data mining technique to analyze student's academic performance. Datasets of engineering students are used in this study. Using associative analysis it is interpreted 19% of students are poor in attendance and GPA. They use classification technique and classify student's grade as excellent, good, average, poor. Clustering algorithm, K Means was used and a simple predictive model with 5 clusters was developed. The authors state that the goal of their work is to develop faith in data mining techniques . [15]

15. Ioannis et.al (2012) develop an prediction model based on Artificial Neural Network. They evaluate the classification capabilities of four training algorithm. Dataset of this study consist of the performance of students in mathematics. A simple user-friendly interface is developed to load data from the user and to display results to them. Ten fold cross validation was used. Experimental results show that MSP trained Neural Network exhibits better classification accuracy. [16]

3. INFERENCES

The following inferences can be made after the comprehensive study of selected literary works.

- Mostly classification based techniques are used for performance prediction because of their simple nature.
- Neural Network, Naive Bayes, Decision Tree are the popular techniques in line for prediction.
- Prediction model achieves accuracy of more than 75%. ID3 reaches 98.5 % accuracy.
- Socio - Economic factors are influential on students performance.

4. CONCLUSION

This paper surveys and presents various research works on predicting the academic performance of students. This survey justifies the capabilities of data mining application in the educational sector. It is concluded through this survey that without any doubts, educational data mining will definitely improve the quality of education and the educational system.

REFERENCES

- [1] Crist'obal Romero and Sebastián Ventura. 2010."Educational Data Mining: A Review of the State of the Art". IEEE Transactions on Systems, Man, and Cybernetics, Vol. 40, No. 6, pp.601-618.
- [2] Anjali B Raut, Ankita A Nichat. 2017. "Students Performance Prediction Using Decision Tree Technique". International Journal of Computational Intelligence Research, Volume 13, Number 7, pp. 1735-1741.
- [3] Ali Daud, Naif Radi Aljohani, Rabeeh Ayaz Abbasi and Miltiadis D. Lytras. 2017. "A Neural Network Approach for Students' Performance Prediction". Proceedings of International World Wide Web Conference Committee, April 3-7, 2017, Perth, Australia.
- [4] Tismy Devasia, Vinushree T P and Vinayak Hegde .2016. "Prediction of Students Performance using Educational Data Mining". Proceedings of International Conference on Data Mining and Advanced Computing, pp. 91-95.
- [5] Hashima Hamsa, Simi Indiradevi, Jubilant J. Kizhakkethottam, 2016. "Students academic performance prediction model using decision tree and fuzzy genetic algorithm", Proceedings of Global Colloquium in Recent Advancement and Effectual Reaserches in Engineering, Science and Technology.
- [6] P. Kavipriya. 2016. "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques". International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 12, pp.101-105.
- [7] Mashael A. Al-Barak and Muna Al-Razgan. 2016. "Predicting Students Final GPA Using Decision Trees: A Case Study". International Journal of Information and Education Technology, Vol. 6, No. 7, pp.528-533.
- [8] K. Prasada Rao, M.V.P. Chandra Sekhara and B. Ramesh. 2016. "Predicting Learning Behavior of Students using Classification Techniques", International Journal of Computer Applications, Volume 139 – No.7, pp.15-19.
- [9] Praneet Kaur, Manpreet Singh and Gurpreet Singh Josan 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. In Proceedings of 3rd International Conference on Recent Trends in Computing, pp.500 – 508.
- [10] Tribhuvan A.P, Tribhuvan P.P, and Gade J.G. 2015. "Applying Naive Bayesian Classifier for Predicting Performance of a Student Using Weka". Volume 7, Issue 1, pp.239-242.
- [11] C. Anuradha and T. Velmurugan. 2015. A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance". Indian Journal of Science and Technology, Vol 8(15).
- [12] Jai Ruby and David. 2015. "Analysis of Influencing Factors in Predicting Students Performance Using MLP- A Comparative Study". Vol. 3, Issue 2. pp.1085-1092.
- [13] G.Narasinha Rao and Srinivasan Nagaraj .2014. "A Study on the Prediction of Student's Performance by applying straight-line regression analysis using the method of least squares". International Journal of Computer Science Engineering, Vol. 3, No.01, pp 43-45.
- [14] Azwa Abdul Aziz, Nur Hafieza Ismail, Fadhilah Ahmad, Hasni Hassan .2013."A Framework for Students' Academic Performance Analysis using Naïve Bayes Classifier", Jurnal Teknologi, Vol 80, No 5.
- [15] Mahendra Tiwari, Randhir Singh and Neeraj Vimal .2013. "An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education". International Journal of Computer Science and Mobile Computing, Vol. 2, Issue. 2, pp.53 – 57.
- [16] Ioannis E. Livieris, Konstantina Drakopoulou, Panagiotis Pintelas .2012. "Predicting students' performance using artificial neural networks". www.math.upatras.gr/~kdrak/Papers/LDP12.pdf

BIOGRAPHIES



V. Sathya Durga is a Research Scholar of Hindustan Institute of Technology and Science, Chennai. She is pursuing her Ph.D. in the area of Educational Data Mining.



Dr. J. Thangakumar is an Associate Professor in the Department of Computer Science, Hindustan Institute of Technology and Science, Chennai. His Ph.D. involved Mobile Adhoc Networks and Network Security.