

# A Review on Student Performance Predication Using Data Mining Approach

Rahul Misra<sup>1</sup>, Dr. Ramkrishan Sahay<sup>2</sup>

<sup>1</sup>PhD Scholar, Department of CSE, Mahamaya Technical University, Noida, India

<sup>2</sup>Professor, Department of CSE, Mahamaya Technical University, Noida, India

[misra.rrahul@gmail.com](mailto:misra.rrahul@gmail.com)

**Abstract-** The success of an academic institution can be measured in terms of quality of education provides to its students. In the education system, highest level of quality is achieved by exploring the data relating to redirection about students performance. Predicting the performance of a student is a great concern to the higher education managements. In this article give the detailed overview about the predication of the student performance based on the data mining approach.

**Keywords-** Data Mining, Education, Student, Performance, predication Models.

## I. INTRODUCTION

Measuring of academic performance of students is challenging since students academic performance hinges on diverse factors like personal, socio-economic, psychological and other environmental variables. The scope of this paper is to predict the student marks and what are the factors that influence the performance of the students [1].

Student's performance is a fundamental part in higher learning foundations. This is on the grounds that one of the rules for a great college depends on its superb record of scholarly accomplishments [2]. There are a ton of definitions on students performance in view of the past writing. Usamah et al. (2013) expressed that students performance can be acquired by estimating the learning appraisal and co-educational program [3]. In any case, the greater part of the examinations referenced about graduation being the proportion of student's achievement. Generally, most of higher learning institutions in Malaysia used the final grades to evaluate student's performance. Final grades are based on course structure, assessment mark, final

exam score and also extracurricular activities [3]. The evaluation is important to maintain student's performances and the effectiveness of learning process. By analyzing student's performance, a strategic program can be well planned during their period of studies in an institution [4].

Besides, educators could also monitor their student's achievements. Students could improve their learning activities, allowing the administration to improve the systems performance. Thus, the application of data mining techniques can be focused on specific needs with different entities. In order to encounter the problems, a systematically review is proposed. The proposed systematically review is to support the objectives of this study, which are:

- To study and identify the gaps in existing prediction methods.
- To study and identify the variables used in analyzing students performance.
- To study the existing prediction methods for predicting students performance.

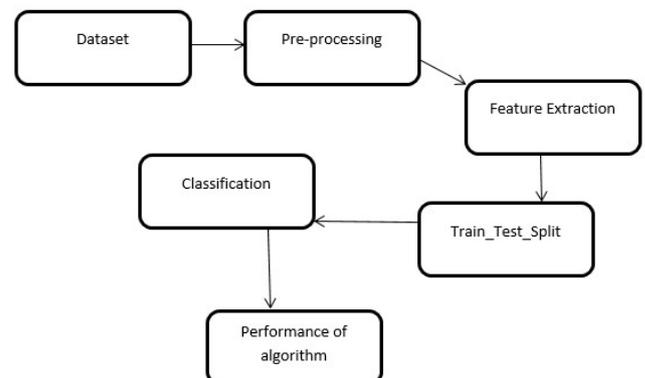


Fig. 1: Example of Student Performance prediction Model

## II. LITERATURE REVIEW

Various researchers done the work on the student performance predication using different data mining techniques and models. Here some of the work done by the different researchers presented below.

In [5] presented the used CHAID prediction model to analyze the interrelation between variables that are used to predict the outcome of the performance at higher secondary school education. The features like medium of instruction, marks obtained in secondary education, location of school, living area and type of secondary education were the strongest indicators for the student performance in higher secondary education. The CHAID prediction model of student performance was constructed with seven class predictor variable.

Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme [6] have applied machine learning techniques to improve the prediction results of academic performances for two the real case studies. Three methods have been used to deal with the class imbalance problem and all of them show satisfactory results. They first re balanced the datasets and then used both cost-insensitive and sensitive learning with SVM for the small datasets and with Decision Tree for the larger datasets. The models are initially deployed on the local web.

Arockiam et al. [7] used FP Tree and K-means clustering technique for finding the similarity between urban and rural students programming skills. FP Tree mining is applied to sieve the patterns from the dataset. K-means clustering is used to determine the programming skills of the students. The study clearly indicates that the rural and the urban students differ in their programming skills. The huge proportions of urban students are good in programming skill compared to rural students. It divulges that academicians provide extra training to urban students in the programming subject.

Cortez et. al. [8] attempted to predict failure in the two core classes (Mathematics and Portuguese) of two secondary school students from the Alentejo

region of Portugal by utilizing 29 predictive variables. Four data mining algorithms such as Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Support Vector Machine (SVM) were applied on a data set of 788 students, who appeared in 2006 examination. It was reported that DT and NN algorithms had the predictive accuracy of 93% and 91% for two-class dataset (pass/fail) respectively. It was also reported that both DT and NN algorithms had the predictive accuracy of 72% for a fourclass dataset.

## III. PREDICATION MODELS

### *Decision Tree (DT)*

A decision tree is a tree in which each branch node will represent a choice between several alternatives and each leaf node will represent a decision [9]. A decision tree is commonly used for obtaining information so as to fulfil the purpose of decision making. Decision tree starts from a root node which is there for users to take actions. From root node users split each and every node recursively into different nodes according to decision tree learning algorithm. The final result is a decision tree where each branch represents a possible context of the decision and its outcome.

### *Neural Network*

The next predictor method used by the researcher to estimate student performance is the neural network. The researcher uses Multi-layer perceptron algorithm to predict student performance. Among seven papers three of them have been used this neural network techniques.

### *Naive Bayes (NB)*

Naive Bayes algorithm is actually based on the probability theory, i.e. the Bayesian theorem [10] and is a simple classification method. It is named as naive because it solves problems based on two critical assumptions: it assumes that there are zero hidden components that will affect the process of analysing and it supposes that the prognostic components are conditionally independent with similar classification.

This classifier provides an efficient algorithm for data classification and it represents the promising approach to the discovery of knowledge.

### ***Support Vector Machine (SVM)***

Support Vector Machine is used for classification which is also a supervised learning method. There are three research papers that have used Support Vector Machine algorithm as their technique to analyse student's performance to review it thoroughly.

### ***K-Nearest Neighbors (KNN)***

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm

## IV. CONCLUSION

Predicting the performance of a student is a great concern to the higher education managements. In this paper given a detail overview about the student performance predication using data mining. Also discussed the related work present by the different researchers and academics and given a brief overview about the different predication models.

## V. REFERENCES

- [1] V.Ramesh, P.Parkavi and K.Ramar, "Predicting Student Performance: A Statistical and Data Mining Approach", International Journal of Computer Applications, Vol.-63, No.8, pp. 35-39, February 2013.
- [2] Jagannath Mohanty, "Modern Trends in Indian Education , Second Revised & Enlarged Edition", Deep &Deep Publication Pvt. Ltd., New Delhi, 2004.
- [3] U. bin Mat, N. Buniyamin, P. M. Arsad and R. Kassim, "An overview of using academic

analytics to predict and improve students' achievement: A proposed proactive intelligent intervention," 2013 IEEE 5th Conference on Engineering Education (ICEED), pp. 126-130, 2013.

- [4] Z. Ibrahim, D. Rusli, Predicting students academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.
- [5] M. Ramaswami and R.Bhaskaran, "A Child Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, No. 1, January 2010.
- [6] Nguyen Thai-Nghe, Andre Busche, and Lars SchmidtThieme, "Improving Academic Performance Prediction by Dealing with Class Imbalance", 2009 Ninth International Conference on Intelligent Systems Design and Applications.
- [7] L.Arockiam, S.Charles, I.Carol, P.Bastin Thiyagaraj, S. Yosuva, V. Arulkumar, "Deriving Association between Urban and Rural Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, pp. 687-690, 2010.
- [8] P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), pp.5-12, 2008.
- [9] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms", International Journal of Computer Science and Management Research, vol.1, 2012.
- [10] V. Ramesh, "Predicting student performance: A statistical and data mining approach", International Journal of Computer Applications, vol. 63, no. 8, 2013.