Survey on Students Academic Performance Prediction Techniques

Ms. Aswini S Prasad *
M.Tech Computer Science and Engineering
Vidya academy of Science & Technology
Thrissur, India.

Ms. Remya P.S
Assistant Professor,
Department of Computer Science and Engineering
Vidya academy of Science & Technology
Thrissur, India.

Ms. Aswathi S
M.Tech Computer Science and Engineering
Vidya academy of Science & Technology
Thrissur, India.

Ms. Silpa VM
M.Tech Computer Science and Engineering
Vidya academy of Science & Technology
Thrissur, India.

Abstract

Over the years student drops out from universities has increased tremendously. The Indian Institutes of Technology have seen over 9%(889) of drop-outs in the year 2016-17, reveals the survey from Ministry of Human Resources and development(MHRD). These estimates will rise when private colleges and government colleges are considered. If one could predict the academic performance of a student in advance, proper action could be taken to decrease the dropout rates. Such a system is helpful for the students as well as to the teachers for academic performance evaluation.

Keywords: Students' Academic Performance, Data Mining, Machine Learning, Classification.

INTRODUCTION

Over the years student drops out from universities has increased tremendously. The Indian Institutes of Technology have seen 9% (889) drops out over the 2016-17 academic year. Of which 71% (630) were PG students, 196 PhD scholars and 63 undergraduates, says the survey of MHRD. These estimates will rise when private colleges and government colleges are considered.

From another survey report it is revealed that in the year 2015-16, Out of 29988 students 3806 students got year out and 529 took TC from the private engineering college after first year. Sources said that in each private college, on an average, about five students drop out after the first year. The majority of them drop out because they got year out.

Students get drop out of colleges due to several reasons one of them is the poor performance of the students in academics which results in bad grades. One way to prevent such a situation is by early predicting the final grades of students in the middle of their studies from the previous results. If that could be done it will be helpful for the teachers as well as the students for improving the academic performance. By analysing the predictions teachers will get an overall view about the performance of a particular student and in what all areas he or she need to concentrate more in order to achieving good grades and students will get an overall idea about their studies.

Building such a model is a challenging task one could use data mining and machine learning techniques for this purpose. Data Mining provides powerful techniques for various fields including education. The research in the educational field is rapidly increasing due to the massive amount of student's data which can be used to discover valuable pattern pertaining students learning behaviour. This paper mainly discusses about such techniques in machine learning using which an prediction model could be build.

GENERAL METHOD

Data mining and machine learning are two important techniques that could be used for building the prediction model. Data mining technique is used to extract use full information such as hidden patterns and relation between the parameters from a huge volume of data. Machine learning is a method in computer science that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

The first step of any data mining project is the data collection phase that is the collection of required data for building the classification model it will be collected from the respective sources. After the data collection step, data pre-processing techniques such as data selection, data cleaning and data transformation is performed. The data after pre-processing is used for building the prediction model, using which the performance of students are analysed. The basic architecture of the prediction model is given in the figure 1.
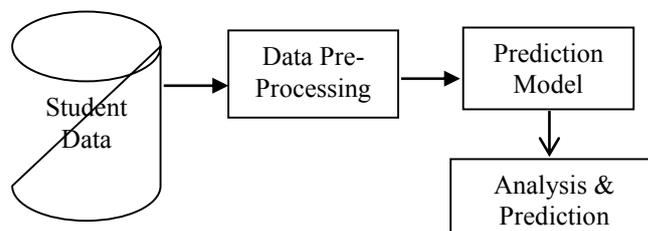

Figure 1: General Architecture of Prediction Model

LITERATURE REVIEW

Pauziah Mohd Arsad, et. al. Here the method of Artificial Neural Network (ANN) is used for the prediction of academic performance of students. The cumulative grade points (CGPA) is used as the measuring criterion. The data needed for the project is collected from electrical department of Teknologi MARA University, Malaysia. The first semester result of students is taken as the input predictor variable (Independent variable) and eighth semester grade points are taken as the output variable (Dependent variable). The study was done for two different entry points namely Matriculation and Diploma intakes. Performances of the models were measured using the coefficient of Correlation R and Mean Square Error (MSE). The outcomes from the study showed that fundamental subjects at semester one and three have strong influence in the final CGPA.

Midhun Mohan et. al. they mainly uses two type of techniques for the overall prediction of the students' performance over a huge volume of data. Those techniques are Learning Analytics and Predictive Analytics. Learning Analytics mainly deals with the data collection and data pre-processing phase where the required data for building the prediction model is collected from the CBSE schools, they have used MySQL server for storing the huge amount of data. Then in the data pre-processing step like data cleaning, data transformation etc. are done by using apache HIVE framework. From the pre-processed data needed information's are discovered out using MapReduce algorithm in the Hadoop framework. Then comes the Predictive analytics part where the actual predictions are made using the multiple linear regression model.

Madhav S. Vyas et. al. uses a decision tree model for predicting the academic performance of students. The data needed for constructing the model is collected and data pre-processing is performed where the continuous values are converted to discrete values and the null values are eliminated. Then by using CART algorithm to the data the decision tree prediction model is being build. And the students with poor performance are predicted out.

Alana M. et. al. Here they are mainly concentrating on two approaches Clustering and regression. Clustering is used for data pre-processing and regression is used for making the predictions. In clustering similar data's are grouped together, though which the data classification are being done and using the regression method an model is being for doing the predictions.

Huda Al-Shehri et. al. This work presented two prediction models for the estimation of student's performance in final examination. The work made use of the popular dataset provided by the University of Minho in Portugal, which relate to the performance in math subject and it consists of 395 data samples. Forecasting the performance of students can be useful in taking early precautions, instant actions, or selecting a student that is fit for a certain task. The need to explore better models to achieve better performance cannot be overemphasized. Most of earlier work on the same dataset used K-Nearest Neighbor algorithm and achieved low results, while Support Vector Machine algorithm was rarely used, which happens to be a very popular and powerful prediction technique. To ensure better comparison, we applied both Support Vector Machine algorithm and K-Nearest Neighbour algorithm on the dataset to predict the student's grade and then compared their accuracy. Empirical studies outcome indicated that Support Vector Machine achieved slightly better results with correlation coefficient of 0.96, while the K-Nearest Neighbor achieved correlation coefficient of 0.95.

David de la Peña et. al. Focuses on a logistic regression model. The collected data are preprocessed and data cleaning is performed after that a reference model is built for each of the course using logistic regression and are stored in the database which are then used for classification. Here the whole student details are being used to classify into two classes namely dropout or non-dropout.

Febrianti Widyahastuti et. al. In here they have done a analysis of linear regression and multilayer perceptron models using WEKA tool kit. They have used undergraduate student data in information system management. The dataset consist of 50 data records. The data's are processed in WEKA tool itself and the algorithms needed for the model construction are also available in the tool named WEKA. Linear regression uses the concept of one dependent variables effect on one or more independent variable, here final grades are chosen as dependent variable and posting and attendance as independent variables. The relations between the variables are plotted using a scatter plot graph. In multilayer perceptron model the concept of neural networks are being used, which is based on weighted connections. Each node will have a weight associated with it which is ten multiplied by the input node to generate output prediction. They have concluded that the multilayer perceptron has more accuracy than the linear regression model. But it takes more time for processing larger datasets.

Xiaofeng Ma et. al. Here they have used decision tree based model. The data needed for the model construction and testing are taken out from the UCI Machine learning Repository. After collecting the data they are being normalized and are used for constructing the model. For the construction of a decision tree model the attribute with the highest information gain is chosen as the root node for the initial partition of the data. The information gain is calculated using the concept of entropy. The partition continues until there are no data left the leaf node contains the label of the data. After the initial model construction an test data is used to test the model.

CONCLUSION

As the data growth progresses it had become difficult for people to analyse and process large amount of data but with the help of data mining and machine learning techniques the analysis of such an huge amount of data has become easier. These techniques are not only helpful in analysing the data's but also helped in predicting the future scope using which huge progress could be made.

REFERENCES

[1] Pauziah Mohd Arsad, Norlida Buniyamin and Jamalul-lail Ab Manan. "A Neural Network Students' Performance Prediction Model (NNSPPM)" IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), 26-27 November 2013.

[2] Midhun Mohan M G, Siju K Augustin and Dr. Kumari Roshni V S "A BigData Approach for Classification and Prediction of Student Result Using MapReduce", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015

[3] Madhav S. Vyas and Reshma Gulwani. "Predicting Student's Performance using CART approach in Data Science", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[4] Alana M. de Morais, Joseana M. F. R. Araújo and Evandro B. Costa. "Monitoring Student Performance Using Data Clustering and Predictive Modelling", 978-1-4799-3922-0/14/$31.00 ©2014 IEEE

[5] Huda Al-Shehri, Amani Al-Qarni, Leena Al-Saati, Arwa Batoaq, Haifa Badukhen, Saleh Alrashed, Jamal Alhiyafi and Sunday O. Olatunji. "Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)

[6] David de la Peña, Juan A. Lara, David Lizcano, María and Concepción Burgos, María L. Campanario. "Mining activity grades to model students' performance", ICEMIS2017, Monastir, Tunisia

[7] Febrianti Widyahastuti and Viany Utami Tjhin. "Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron", 978-1-5090-4688-1/17/$31.00 ©2017 IEEE

[8] Xiaofeng Ma and Zhurong Zhou. "Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree", 978-1-5386-4649-6/18/$31.00 ©2018 IEEE.

[9] Sagardeep Roy and Anchal Garg, "Analyzing Performance of Students by Using Data Mining Techniques A Literature Survey", 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)