Data Mining Analysis on Student's Academic Performance

ABSTRACT

Data mining is the evolving process of identifying and extraction the hidden information from a data warehouse. Data Mining is widely used in business, medical, engineering and educational areas. This study covers the application of data mining in education for predicting the academic performance of the students. Educational Data Mining(EDM) plays a dominant role in the data mining era. With the help of EDM we can predict the academic performance of the students using different techniques like Decision tree, Random Forest, Naive Bayesian and Multilayer Perceptron. In this paper, decision tree algorithm has been implemented for predicting the academic performance of the students, by not only collecting the academic grade but also the students background details, social details and academic details for predicting the performance of the students. The model can be used by the educational institution and teachers for improving the performance of the weak students. The decision tree implementation is done using the WEKA tool.

Keywords—Student Performance Prediction, Educational Data Mining, Academic Performance

Data Mining Analysis on Student's Academic Performance

ABSTRACT

Data mining is the evolving process of identifying and extraction the hidden information from a data warehouse. Data Mining is widely used in business, medical, engineering and educational areas. This study covers the application of data mining in education for predicting the academic performance of the students. Educational Data Mining(EDM) plays a dominant role in the data mining era. There is an essential need to identify effective algorithms for predicting the student's performance. With the help of EDM we can predict the academic performance of the students using different data mining techniques like Decision tree, Random Forest, Naive Bayesian and Multilayer Perceptron. Each technique has its own advantages and disadvantages. This paper comprises of what are the different types of EDM techniques used in EDM and what are the different tools used for implementing the EDM techniques in order to predict the performance of the student's. EDM helps in developing a warning system for identifying weak student's prior and give adequate training to improve the academic performance of the students.

Keywords—Student Performance Prediction, Educational Data Mining, Data Mining Technique, Academic Performance

LITERATURE SURVEY

Students' social activities and background details were used by Ching-Chieh Kiu [1] in the year 2018 for predicting the student's academic performance. This study used several data mining technique has been applied on predicting the accuracy of each dataset, of which decision tree J48 has given the best accuracy rate. The decision tree model has achieved 95% accuracy rate compared to other data mining technique. The dataset was divided into three subsets, students background details, student's social activities and student course work. The algorithms were implemented in WEKA tool. Moustafa, Hatim and Imad Elzein [2] stated that there are different techniques used for prediction the student's performance in the year 2018. The techniques are Classification, Clustering, Association and Decision tree. In classification technique educational data was classified into behavioral and educational. The dataset was obtained from Modern Academy for School Education (MASE) School, Abbassieh (South Lebanon) on the sampling method of students' data of year 2017-2018.

A comparative study on decision tree and random forest was done by Fergie Joanda and Reymon in the year 2018. The dataset was collected using questionaries' from the computer science students. The study proves that decision tree gives more accurate result of 66.9% compared to random forest which gave 61.44%. The comparison was implemented using WEKA tool [3]. According to the study of Rahul Patel classification methods gives more accurate result for predicting student's performance. The improved ID3 algorithm gives more accurate result of 74% compared to the ID3 which was not improved produced only 70% accuracy [4]. The Blended learning was implemented in the year 2018 by Owen H. T. Lu1, in the study Component Regression and Decision tree were blended and used for predicting the academic performance. Skewed dataset were used and decision tree was used for early warning system for collecting student's academic details which gave 95% accuracy rate and regression was used on student's grade which resulted 76% accuracy rate[5].

A comparative study was done on Naïve Bayes, Decision Tree, K-Nearest Neighbor and Discriminant Analysis by Samuel, Nor Bahiah and Siti Mariyam in the year 2019. The study was done to identify the best data mining technique for predicting the student's academic performance. It used 10 datasets from the University of California Irvine Repository. The decision tree out performed with the accuracy of 81.94% compared to other data mining techniques. The accuracy

of other techniques was Naïve Bayes-73.61 %, KNN-80.56 % and Discriminant Analysis -77.78% accuracy rate. The tool used in the study was WEKA. In future hybrid metaheuristics algorithms will be for feature selection on the student data [6]. According to the study of Nongnuch Ketui, Warawut Wisomka and Kanitha Homjun in year 2019, Gradient boosted trees has given the best accuracy compared to other classification data mining techniques like Decision Tree, Weighted Decision Tree, Iterative Dichotomiser 3 (ID3) and Random Tree. WEKA is the data mining tool which is used for implementing the data mining techniques. A raw dataset was collected from the Rajamangala University of Technology Lanna Nan. The gradient boosted tree and decision tree gave good accuracy rate of 92.31% and 91.03% compared to other techniques. The Weighted Decision Tree 84.14%, ID3-89.66% and Random Tree-84.14% accuracy rate. The classification technique is widely used in predicting the student's performance [7].

There are six different Educational Data Mining techniques which are used are: Visualization of Data, Classification, Regression, Clustering, Association rule mining, Sequential Pattern Matching by Nadia Anjum and Srinivasu Badugu in the year 2019. The Visualization of data methods used to focus on useful information and it supports decision making. In Classification the data are classified based on different algorithms like decision tree, Bayesian network. Regression is seen as a linear regression, it only describes the linear relationship between the dataset and if any non-linear relationship is found it would have poor working model. Clustering is used to identify common traits among the students. Association is used to identify the weak students by using different rules. In Sequential Pattern Matching it finds the relationships between Sequential events to know whether there is a sequence in their occurrences [8].