Maximizing Student Efficiency Using Data Mining Technique

PNV Syamala Rao

Assistant Professor, Department of CSE, BVRIT, Telangana, India

ABSTRACT

Data mining technology is a procedure for analyzing large amounts of available data and extracting useful information and knowledge to support critical decision-making processes. Data mining can be applied to various applications in the field of education to improve student performance. Educational data mining is developing rapidly and is an important technology for data analysis in the field of education. This paper shown different classification algorithms of data mining those are used for development of a data mining model for predictions of performances of students, on the basis of their personal demographic and academic information. This paper analyze and evaluate the students" performance by applying data mining classification algorithms in weka tool.

Keywords : Classification, Data Mining, Prediction, Performance, Knowledge, Information.

I. INTRODUCTION

Now the university operates in a very complex and competitive environment. The key challenge of modern universities is to analyze their performance in depth, to determine their uniqueness, and to develop strategies for further development and future action. The objective of the proposed research work is to find out whether any pattern in the available data may help to predict student academic performance based on individual and pre-university characteristics. Data mining may be a promising and thriving frontier in data analysis. Data mining can also be called data knowledge discovery (KDD). The system acts as a machine-driven or easy-to-extract model that represents knowledge that is implicitly saved or captured in large databases, data warehouses, Webs, data repositories, and information flows. Data mining is a multidisciplinary field covering information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information systems, artificial intelligence and data visualization. The application of data mining is prevalent in the educational system. Educational data mining is an emerging field that can be effectively applied in the field of education. Educational data mining uses several ideas and concepts, such as association rules mining, classification and clustering.

The emergence of knowledge can be used to better understand the student's promotion rate, student enrollment rate, student transfer rate and student success rate. Data mining system is crucial for measuring student performance improvement. Classification algorithms can be used to accurately classify and analyze student data sets. Students 'academic performance is influenced by parents' education, geography, economic status, attendance, gender, achievement and other factors. The main purpose of this paper is to use data mining methods to learn and analyze student performance. Data mining provides a number of tasks that can be used to study student performance. In this paper, the classification task is used to assess student performance and to deal with the accuracy of the various classification data mining algorithms provided. This paper is organization as per the following. Section I gives the introduction of research work, Section II describes the objective of the research, Section III shows the literature survey, Section IV specify the data collection and methodology, Section V describes the experiment result and discussion and section VI gives the conclusion of research. This research paper have compared the different data mining algorithms applicable to classification: rule-based, tree-based, function-based and Bayesian-based algorithms. This analysis is carried out after training and testing algorithms, and it is possible to draw conclusions about possible predictors

of student success. The purpose of our study is to find an optimal data mining algorithm to predict student performance.

II. LITERATURE SURVEY

We studied various articles regarding performance evaluation of Data Mining algorithms on different tools, some of them are described here, Mahendra Tiwari & Yashpal Singh (2012)[1] evaluated performance of 4 clustering algorithms on different datasets in WEKA with 2 test modes. We presented their result as well as about tool and data set which are used in performing evaluation.

Tanuja S, Dr. U. Dinesh Acharya, and Shailesh K R (20111) in their article "Comparison of different data mining techniques to predict hospital length of Stay" compared four data mining classification techniques MLP, Naïve Bayes, K-NN, J48 to predict length of stay for an inpatient in hospital on preprocessed dataset derived from electronic discharge summaries with 401 instances & 16 parameters. In result they found that MLP performs better than rest three classifiers with 87.8% correctly classified instances. Sonali Agarwal, G. N. Pandey, and M. D. Tiwari (2012) [3] describes the educational organizations are one of the important parts of our society and playing a vital role for growth and development of any nation. Data Mining is an emerging technique with the help of this one can efficiently learn with historical data and use that knowledge for predicting future behavior of concern areas. Growth of current education system is surely enhanced if data mining has been adopted as a futuristic strategic management tool. The Data Mining tool is able to facilitate better resource utilization in terms of student performance, course development and finally the development of nation's education related standards.

Monika Goyal and Rajan Vohra (2012) [4] applied data mining techniques to improve the efficiency of higher education institution. If data mining techniques such as clustering, decision tree and association are applied to higher education processes, it would help to improve students' performance, their life cycle management, selection of courses, to measure their retention rate and the grant fund management of an institution. This is an approach to examine the effect of using data mining techniques in higher education.

Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal (2012) [5] used decision tree classifiers are studied and the experiments are conducted to find the best classifier for retention data to predict the student's drop-out possibility.

Brijesh Kumar Baradwaj and Saurabh Pal (2011) [6] Used the classification task on student database to predict the students division on the basis of previous database.

K.Shanmuga Priya and A.V.Senthil Kumar (2013) [7] applied a Classification Technique in Data Mining to improve the student's performance and help to achieve the goal by extracting the discovery of knowledge from the end semester mark.

Bhise R.B, Thorat S.S and Supekar A.K. (2013) [8] used data mining process in a student's database using K-means clustering algorithm to predict students result.

Varun Kumar and Anupama Chadha (2013) [9] used of one of the data mining technique called association rule mining in enhancing the quality of students' performances at Post Graduation level.

Pallamreddy.venkatasubbareddy and Vuda Sreenivasarao (2010) [10] explained the Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal and use of decision trees is as a descriptive means for calculating conditional probabilities.

III. DATA COLLECTION METHODOLOGY

For this study, we have collected student data from the various higher educational institute by preparing the question bank and submitting it among the students. During this phase we have collected total 3558 student recordset. After collecting the data, we have converted it to the suitable format for further analysis into the WEKA tool.

Data was collected based on the parameters such as gender, Hsc percentage, stream, father qualification, mother qualification, Father occupation, mother occupation, no_of_sublings, family_annual_income, overall attendance, weekly library usage, weekly lab

usage, daily_reading_hours, assignment_marks and Final result.

IV. EXPERIMENT RESULT AND DISCUSSION

This research study used WEKA used weka tool for analysis. We have used 10-fold cross validation method to test the accuracy of algorithm.

This research study used various classification algorithms to analyse the student's performance such as **Rule based classification**: Decisiontable, Jrip, Oner, Part, and Zeror, **trees-basedalgorithms:** Decisionstump, J48, Randomforest, Random tree, **functions-based algorithms**: Logistic, Multilayerperceptron, and Smo, **bayes-based algorithms**: Bayesnet and naivebayes. This research study found result as per the following table 1.1 and figure 1.1.

Table 1.1 Analysis result of various algorithm using WEKA

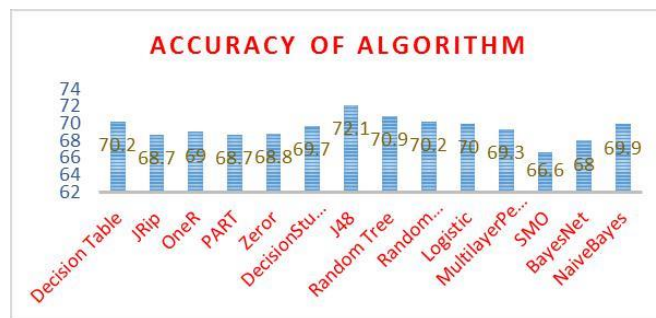| Classification Techniques	Algorithms	Accuracy of Algorithm
Rules-based Algorithm	Decision Table	70.2
	JRip	68.7
	OneR	69
	PART	68.7
	Zeror	68.8
Trees-based Algorithm	DecisionStump	69.7
	J48	72.1
	Random Tree	70.9
	Random Forest	70.2
Functions-based	Logistic	70
Algorithms	MultilayerPerceptron	69.3
	SMO	66.6
Bayes-based Algorithms	BayesNet	68
	NaiveBayes	69.9

Figure 1. Chart for analysis result of various algorithm

V. RESULT ANALYSIS

We have collected student dataset and converted it to the format which is understandable by WEKA tool. After that we have applied various classification algorithm on collected dataset to predict the student's performance ratio using the WEKA tool. As per the result shown in table 1.1 and figure 1.1 we found that J48 algorithm gives the highest accuracy among all other algorithms.

VI. CONCLUSION AND FUTURE SCOPE

Data mining perform the very important role for analysis of student's performance. Data mining methods and techniques helps to educators for taking important in decision-making process.In this research paper, we have applied various classification algorithms of data mining for predicting the performances of students. To perform the data mining analysis we have collected student dataset using demographic and academic information. In this research study we have used various classification algorithm to perform the analysis such as **Rule based classification**: Decisiontable, Jrip, Oner, Part, and Zeror, **trees-basedalgorithms:** Decisionstump, J48, Randomforest, Random tree, **functions-based algorithms**: Logistic, Multilayerperceptron, and Smo and **bayes-based algorithms**: Bayesnet and naivebayes. The analysing result shows that tree based J48 algorithm is the best algorithm among all other

algorithm for predicting the student's performance. As a future scope we planned to develop a hybrid algorithm to get the higher accuracy as compared to J48.

VII. REFERENCES

[1]. Mahendra Tiwari, Yashpal Singh, "Performance Evaluation of Data Mining clustering algorithms in WEKA", Global Journal of Enterprise Information System, vol 4, Issue I, pp.24-29, Nov. 2012.

[2]. Tanuja S, Dr. U. Dinesh Acharya, and Shailesh K R "Comparison of different data mining techniques to predict hospital length of Stay, Journal of Pharmaceutical and Biomedical Sciences (JPBMS)", Vol. 07, Issue 07, Oct. 2011.

[3]. Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", 2012.

[4]. Monika Goyal ,Rajan Vohra2, "Applications of Data Mining in Higher Education", 2012.

[5]. Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal, "Mining Education Data to Predict Student's Retention: A comparative Study", 2012.

[6]. Brijesh Kumar Baradwaj, Saurabh Pal, "Mining Educational Data to Analyze Students Performance", 2011.

[7]. K.Shanmuga Priya, A.V.Senthil Kumar, "Improving the Student's Performance Using Educational Data Mining", 2013.

[8]. Bhise R.B, Thorat S.S, Supekar A.K, "Importance of Data Mining in Higher Education System", 2013.

[9]. Varun Kumar, Anupama Chadha, "Mining Association Rules in Student's Assessment Data", 2012.

[10]. Pallamreddy.venkatasubbareddy, Vuda Sreenivasarao, "The Result Oriented Process for Students Based On Distributed Data Mining", 2010.