National Conference on "Recent Trends in Mathematical Modeling, Simulation Methods, Computations, and Physical Sciences"
(MMSCP-2019), Organised by Hislop College, Nagpur, Maharashtra, India
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 4 | Issue 4 | ISSN : 2456-3307

Predicting Students' Performance using J48 Decision Tree

Mehta Smruti Hemantkumar[1], Dr. Ashish Adholiya[2]

[1]Research Scholar, Pacific Academy of higher Education and Research University, Udaipur, Rajasthan, India
[2]Assistant Professor of IT of Marketing, Pacific Academy of higher Education and Research University, Udaipur, Rajasthan, India

ABSTRACT

Data is necessary in any industry which can be processed for getting useful information. Previously the data mining techniques are used in business to earn more profits and increase the business. In academic and educational field, Data Mining is a leading tool for predicting performance of students. Performance of student in university courses plays an important role to the higher education institutions. A society develops the quality of their citizens on the bases of education. There are many techniques available in data mining such as classification, clustering, association, etc., which are useful in extracting the hidden knowledge and useful information. Here, we used classification technique. In this paper, Educational Data Mining is used to predict students performance based on their marks in an examination. For these we have used WEKA tool. After pre-processed the data, we applied the J48 decision tree algorithm to discover classification rules.

Keywords: Educational Data Mining, Classification, WEKA, J48 Decision Tree

I. INTRODUCTION

"A secret of success is to know something that nobody else knows" said by Aristotle Onassis [1].

Data mining is the process of discovering and scrutinizing the large amount of data to find out the valid and useful information in database and using it to take essential decisions.

Knowledge is extracted from the hidden database and searches the information according to the need. For these, various algorithms of data mining are used. At last, the data is filtered by passing through various steps of data mining and then user used it according to their need. It is an essential part of education institute.

The data comprise of various formats like text, audio, images, animated scripts, video etc. which are stored in warehouses. The data mining is a interdisciplinary research field, which includes areas like machine learning, statistics, artificial intelligence, data retrieval, information technology, neural networks, information based systems and data visualization [2]. Nowadays, the researchers are paying more attention in the field of education data mining because the educational data mining is becoming the most promising field. The data can be discovered from various colleges, universities and websites. The result can be helpful to instructors, teachers or professors in improving the teaching, e-learning systems, co-curriculums activities and other benefits.

The outline of this paper is organized as follows: Section 2 presents the description of EDM. Section3

presents the related work in educational data mining. Section 4 presents model construction of classification. Section 5 describes how WEKA tool is used for classifying the student's data. Section 6 concludes this paper with a summary and future work.

II. Educational Data Mining

The Educational Data Mining(EDM) community website, www.educationaldatamining.org, (Baker and Yacef (2009)) [3], defines educational data mining as follows: "Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in."

EDM converts the primary data into the useful information which comes from educational domain and has a greater impact on it.

EDM is a promising research field with psychological and computational method and research advances to understand the learning process of students. EDM applies various data mining tools and techniques related to education.

The various data mining tasks are available within EDM e.g. classification, clustering, prediction, association rule etc. EDM predicts the behavior of students, student's future enhancement skills, prediction of scholarship and other areas of student which exaggerate the quality of their skill and attitudes.

III. Related Work

Prerna Kapoor and Reena Rani [4] presented that decision tree s are most extensively researched area in Knowledge Discovery. J48 Decision Tree algorithm had been developed and is used for solving everyday classification tasks. The author aimed to improve the predictive performance of these algorithms by extenuating three of their major disadvantages by Pruning Trees. The dataset is taken from Supermarket. The algorithm cuts the Tree Size as well as the Finished Entropy Mean and Absolute. The decrease in error results in accurate classification. In the classified tree the Total Entropy of the Decision Tree, Randomness in J48 is much higher in J48.

Jai Ruby, Dr. K. David [5] measured the quality of education based on the academic performance of students and the results produced. Many factors are influencing the academic performance of the students. This study is mainly focusing on analyzing the prediction accuracy of the academic performance of the students by using only influencing factors by Multi Layer Perceptron algorithm. This paper helps the institutions to know the academic performance of their students in advance and can concentrate on weak students to improve their academic results.

Chaman Verma [6] has introduced Data Science as a collective field that joints skills from various domains such as Software engineering, Statistics and Data mining. It focuses on storing, transforming, cleaning and processing of unstructured data. The author described the execution of popular data mining algorithm named Apriori using WEKA tool and finds the association rules between data set transactions with the help of support and confidence measurements in WEKA tool. The algorithm discovers only one association rule. It discloses that presence of A with C and C with A in transaction data set. It also reveals that transaction T1 and T2 are associated with each other well due to occurrence of both frequent characters A and C.

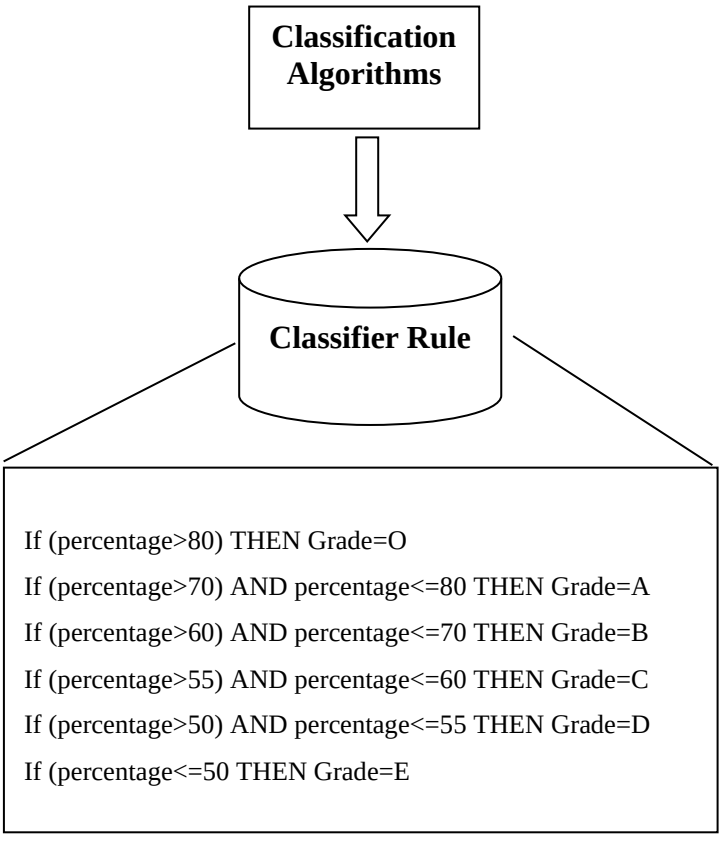
Mashael A. Al-Barrak and Muna Al-Razgan [7] described Educational data mining as a process of applying data mining tools and techniques to analyze

data at educational institutions. The author used EDM to predict students' final GPA based on their grades in previous courses. After pre-processed the data, J48 decision algorithm is applied to discover the classification rules.

Dineshkumar B Vaghela, Priyanka Sharma [1] have discussed various data mining applications like Retail industries, Telecommunication industries, Financial data analysis, Biological data analysis, Intrusion detection etc. The author presented new algorithm with Binary Search Tree which stores the global rules by consolidating the local rules generated at each site. This Global Rule Binary Search Tree (GRBST) can then be used in prediction of Students' admission to college.

IV. Classification

Classification is a data mining technique which is based on machine learning. It classifies each item in a data set and maps it into predefined classes or groups to develop a model. It is also known as supervised learning. There are various algorithms of classification technique such as Decision tree, Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression etc. Here, we used Decision Tree algorithm. Because, Decision Tree algorithm is used in statistical data mining and machine learning.



WEKA as EDM tool

In this study, we select WEKA (Waikato Environment for Knowledge Analysis) software which was developed at the University of Waikato in New Zealand. With the help of WEKA, we apply data preprocessing, cleaning and handle missing values. This software used some tools like data pre-processing, classification, association

rules, visualization, clustering and regression. It is open source and runs on any platform. It is used in many application areas, particularly in education and research.

In this paper, we have collected data set of BCA students. Figure 2 shows the result remarks of students in an examination. 139 students fail, 69 students pass, 84 students got first class, 68 students got second class and 16 students got first class distinction.

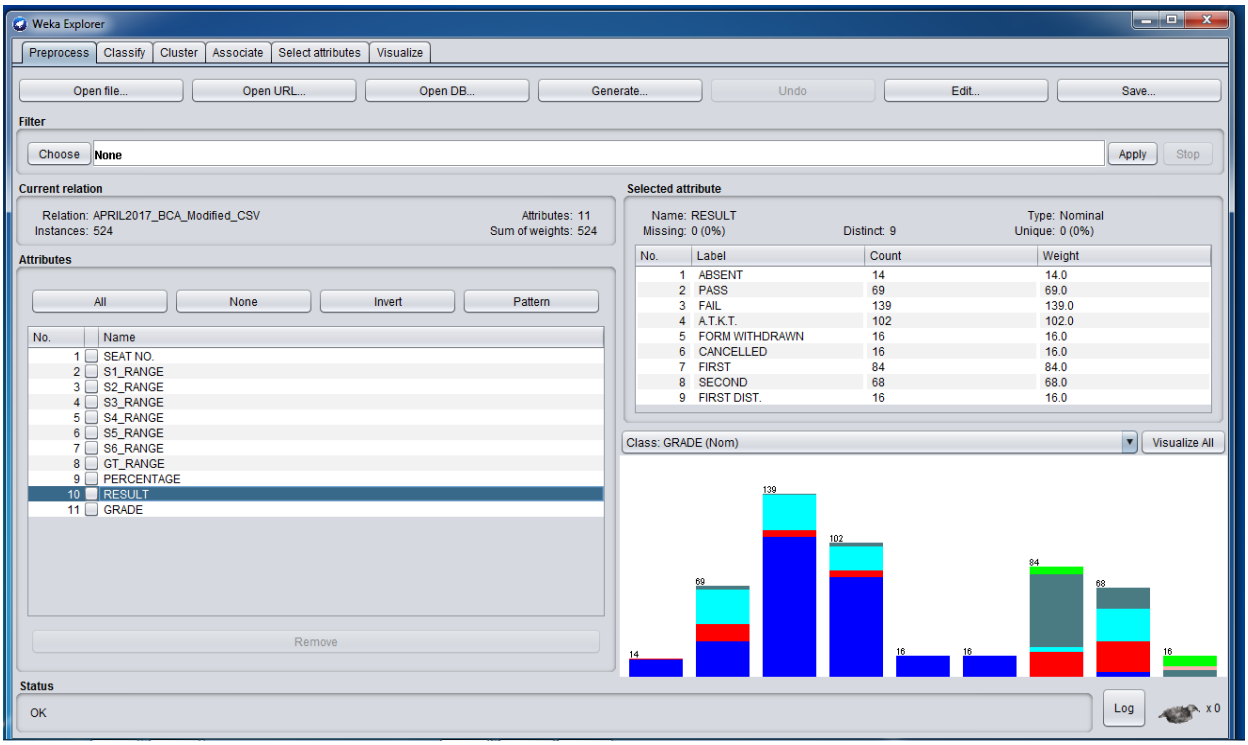


Figure 5.1: WEKA Explorer window open with dataset

In figure 3, the accuracy of correctly classified instances after selecting attributes are more than before selecting attributes. The mean absolute error and root mean squared error are 0.0048 and 0.057 which are relatively low. So, it will give more accurate result.

	10 Fold Cross Validation		Training and Test set	
	Before	After	Before	After
Correctly Classified Instances	47.081%	99.058%	52.919%	99.2467%
Incorrectly Classified Instances	52.919%	0.942%	47.081%	0.7533%
Kappa statistic	0.3018	0.985	0.381	0.988
Mean absolute error	0.1433	0.0048	0.1337	0.004
Root mean squared error	0.2718	0.057	0.2586	0.0448
Relative absolute error	80.5221%	2.2716%	75.1925%	1.914%
Root relative squared error	91.2219%	17.6437%	86.7936%	13.8601%

Figure 5.2 Percentage of Classifier (Before and After selecting attribute)

V. CONCLUSION & FUTURE WORK

This paper shows that how to improve students' performance on education related data. We used data mining techniques to discover the hidden knowledge. We used classification technique using J48 algorithm, which is used to predict the performance of students. In testing process, the accuracy is more than 90%, means it is well suited for classifying large sets of data. In future we will improve classification accuracy by using some other data mining techniques like K-Nearest Neighbor, Navie Bayesian etc.

VI. REFERENCES

[1] Dineshkumar B Vaghela, Priyanka Sharma, "Students' Admission Prediction using GRBST with Distributed Data Mining", Communications on Applied Electronics (CAE), Volume 2 – No.1, June 2015.

[2] Dr. Pranav Patil, "A Study of Student's Academic Performance Using Data Mining Techniques", International Journal Of Research In Computer Applications And Robotics, Vol.3, Issue 9, ISSN 2320-7345, September 2015.

[3] Sen, Umesh Kumar, "A Brief Review Status of Educational Data Mining", International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), Vol. 3, Issue 1 (Jan. – Mar. 2015).

[4] Prerna Kapoor and Reena Rani, "Efficient Decision Tree Algorithm Using J48 and Reduced Error Pruning", International Journal of Engineering Research and General Science, Volume 3, Issue 3, May-June,2015.

[5] Jai Ruby, Dr. K. David, "Analysis of Influencing Factors in Predicting Students Performance Using MLP – A Comparative Study", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2, February 2015.

[6] Chaman Verma, "Finding Pattern using Apriori Algorithm through WEKA", International Conference on Recent Innovations in Science, Agriculture, Engineering and Management, ISBN: 978-93-86171-80-1, 20th November 2017.

[7] Mashael A. Al-Barrak and Muna Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study", International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.