International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET2073124

Predict Students' Performance by Using J48 Algorithm

Myo Thandar Tun[*1], Yin Yin Htay[2]
[*1]Information Technology Supporting and Maintenance Department, University of Computer Studies, Mandalay, Myanmar
[2]Faculty of Information Science, University of Computer Studies, Magway, Myanmar

ABSTRACT

The critical issue to the academic community of higher education is to monitor the progress of students' academic performance. We can use data mining techniques for this purpose. J48 algorithm is one of the famous classification algorithms present today to generate decision trees in data mining technique. The data set used in this study is taken from University of Computer Studies (Mandalay). Weka machine learning tool is applied to make classification. In this work, we tested result classification accuracy was computed. This J48 classification algorithm give accuracy with 78.2%.

Keywords : Data Mining, Student Academic Performance, Classification, J48 algorithm.

I. INTRODUCTION

Data mining is one of the most cardinal areas in recent technologies for retrieving valid information from huge amount of unstructured and distributed data using parallel processing of data. Data mining techniques are applied in various fields to find the novel information from huge data set. Nowadays data mining techniques are mostly used in educational field. Most of the researcher has taken data mining techniques to find the useful information from educational field. Applying data mining in education field is the most prominent research area of today's researcher. Educational Data Mining (EDM) is attracting researchers for developing methods from educational institutions' data that can be used in the enhancement of quality of higher education. EDM uses the lot of information present in the educational databases about teaching-learning practices for the development of models which are benefits for all the participants in the educational process.

The prediction of students' performance in any higher education has become one of the most important needs of that institute to improve the quality of the teaching process of that institution. In this process, we get to know the needs of the students and hence we can fulfil those needs to get better results. Students who need special attention from the teachers can also be identified from this process. There are a number of algorithms available for predicting the performance of students such as Artificial Neural Network (ANN), Decision Trees, Clustering, Naive Bayes algorithm, Decision Trees being most commonly used.

II. RELATED WORK

Bhardwaj and Pal [1] conducted a performance analysis on 50 students and 8 attributes from VBS Purvanchal University, Jaunpur (Uttar Pradesh) with the objective to study student's performance using. To classify the data Decision tree method was used. This Study helped teachers to improve the result of the student.

Srimani and Kamath [2] used various methodologies to study the application of data mining algorithms for

the performance analysis of the learning model. Various classification models like Bayesian, Multilayer perceptron (MLP), Decision Tree using j48, Rule Based RIPPER were used to implement on dataset for class 1 to 7 that contains of 3500 data instances and 99 attributes to predict progress of each children. Results of all algorithms found to be accurate and hence the model is justified with accuracy (almost 100).

Yadav, Brijesh and Pal [3] conducted study to predict student's performance with 48 students and 7 attributes obtained whose records were taken from VBS Purvachal University, Janupur (UP), India on the sample method of computer Applications department of course MCA (master of Computer Applications) from session 2008 to 2011. Decision tree algorithms like ID3, C4.5, and CART were used for classification. Results were shown that CART is the best algorithm for the classification of data. This study will help to identify those students who need special attention and also this work will help to reduce fail ratio.

C. Marquez-vera and Venturia [4] used data mining techniques to predict school failure with 10 attributes and 670 middle-school Students' data from Zacatecas, Mexico classifying unbalanced data by rebalancing data and cost-sensitive classification approaches were used along with 10 classification algorithms and 10 fold-cross validation to compare and select best approach to predict students performance who might fail cost-sensitive classification approach is considered as best approach for classification technique.

Sembiring, Zarlis et al [5] applied the kernel method as data mining techniques to analyze the relationships between students' behaviour and their success by using Smooth Support Vector Machine (SSVM).

Nandini and Saranya [6] attempted to extract useful, reliable patterns for predicting pupil's performance by using ID3 algorithm on the database of Dr. NGP Arts and Science College, Coimbatore and founded that efficiency of ID3 algorithm is good.

Kabra and Bichkar [7] discussed the use of decision trees to predict the performance of students by using the data of 346 students of SGR Education Foundation's College of Engineering and Management and the results are only 60.46% accurate, that is, only 209 instances out of 346 were correctly classified.

Pandey and Sharma [8] applied many algorithms like J48, NBtree, Reptree and Simple cart on the dataset collected from Manav Rachna College of Engineering, Faridabad to predict the performance of students and found that J48 decision tree algorithm was the best suitable for model construction.

III. TECHNIQUES USED IN EDUCATIONAL DATA MINING (EDM)

A. Association Rule Mining

Association rule mining algorithms are intended to discover pertinent relationships between the variables of the data set. These associations are if…then rules. Association rules have a probability of occurrence, that is, if condition is met, and then there is a certain possibility of occurring result. Association rule mining algorithms are mine only strong rules. Strong rules satisfy a minimum support threshold and a minimum confidence threshold [9].

B. Clustering

Clustering is an unsupervised learning paradigm; it may reveal interesting unknown relations in the data. It is a system of grouping sets of objects in groups in a way that objects from a cluster have more similarities than objects from different clusters. Each cluster may be considered as a class with no label, and thus, clustering is sometimes referred to as automatic classification [9].

C. Classification

Classification is used to classify the data based on the training set and then uses that pattern to classify the new data which is known as the training set. It is called as supervised learning technique because the classes are predefined before extracting patterns on the target data [10]. In Educational Data Mining (EDM), some popular classification methods are Decision Trees, Bayesian Classifier, Artificial Neural Network, Support Vector Machine, K-Nearest Neighbor, Linear Regression and Density Estimation.

D. Linear Regression

Linear regression is a prediction technique that predicts a numeric value. Various attributes like sales, age and weight. It is statistical methodology [10]. The aim of the task is to achieve a function of the independent variables that allows computing conditional expectations of a dependent variable for prediction.

E. Support Vector Machine

Support Vector Machine (SVM) is a classification method of both linear and nonlinear data. To transform the original training data into a higher dimension it uses a nonlinear mapping. The significant benefit of SVM is the features of global optimization and high generalization ability. In addition, it removes over fitting issues and provides a sparse solution when comparing with traditional approaches like Artificial Neural Network (ANN) [11].

IV. PROPOSED WORK

In this work, we collected data from the students of University of Computer Studies, Mandalay. We have taken only 500 data set from totally 700 data set collected from the student, from this for predicting the performance of student. To predict the student performance, we applied the data set in weka tool.

The data set we used for better prediction is given in the Table 1.

TABLE 1. ATTRIBUTE DESCRIPTION

Attribute	Description
Major	Major (CS-Computer Science or CT-Computer Technology
Age	student's age (from 16 to 22)
Address	student's home address type ('U' - urban or 'R' - rural)
Famsize	family size ('LE3' - less or equal to 3 or 'GT3' - greater than 3)
Medu	mother's education (0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
Reason	reason to choose this University (close to 'home', University 'reputation', 'course' preference or 'other')
Grade	Grade from previous Academic (Excellent, Good, Average or Poor)

A. Data Preprocessing

Predicting the student performance student.arff data was used in the weka machine learning. This contains 17 attributes. These attributes were applied to data preprocessing for better prediction. we applied CfsSubsetEval from attribute selection method in data preprocessing stage to obtain better classification result. Based on the above preprocessing step, 7 attributes are selected for classification step. Fig.1 shows the attribute selection method using CfsSubsetEval.

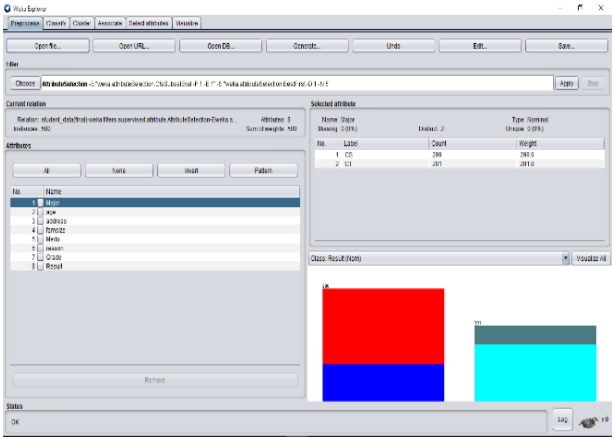

Fig. 1. Attribute Selection

B. CfsSubsetEval

CfsSubsetEval estimates the importance of a subset of features by taking into individual predictive ability of each attribute along with the degree of redundancy between them [12]. Using these attribute evaluation step proper attributes was selected for the classification step.

C. J48 algorithm

J48 algorithm is developed by Quinlan Ross that generates the decision trees which can be used for classification problems. It is the successor of ID3 algorithm by dealing with both categorical and continuous attributes to build a decision tree. It is also based on Hunt's algorithm. Based on the selected threshold, J48 algorithm splits the attribute values into two partitions such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values [13]. Fig. 2 shows classification using J48 algorithm.

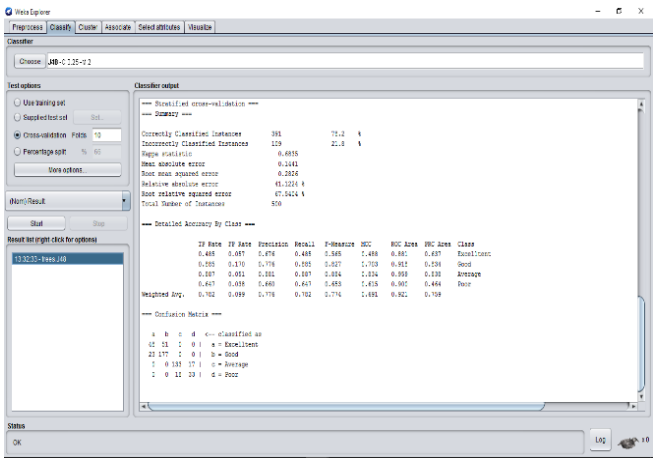

Fig 2. Classification with J48 algorithm

Table 2 shows the experimented result. The J48 classification algorithm classified 391 instances are correctly and 109 instances are incorrectly. The classification accuracy was 78.2%. So, J48 algorithm accurately predicted the given data set. This classification accuracy will helpful to predict the student performance of future data set.

TABLE 2. EXPERIMENTED RESULTS

Total Number of Instance	Instances Classified Correctly	Instances Classified Incorrectly	Classification Accuracy
500	391	109	78.2%

V. CONCLUSION AND FUTURE WORK

In this paper, we used J48 classification algorithm for the prediction of student performance. Prediction of student performance is one of the most prominent researches filed in educational data mining. J48 classification algorithms were applied for predicting the student performance. The classification accuracy was computed by J48 classification is 78.2%. In future we wish to apply different classification algorithm to improve the classification accuracy. There are numerous researches has to be done in the field of education to increase the classification accuracy using different classification algorithm. This helps to the teacher as well as institution to take proper decision to improve the performance of the student.

VI. ACKNOWLEDGEMENTS

I would like to special thank Rector and Pro-Rector of University of Computer Studies (Mandalay) and also especially like to thank Head of Department of Information Technology Supporting and Maintenance (ITSM) and everyone. Finally, I am grateful to my family, thank you for your love, support and energy, and appreciation for this.

VII.REFERENCES

[1]. B.K. Bharadwaj and S. Pal. "Mining Educational Data to Analyze Students Performance", International Journal of Advance Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp.63-69, 2011.

[2]. P.K. Srimani and Annapurna S Kamath." Data Mining Techniques for the Performance Analysis of a Learning Model-A case study" International Journal of Computer Applications (0975-8887), volume 53-No 5 September 2012.

[3]. Surjeet Kumar Yadav, Bhardwaj and S Pal." Data Mining Applications: A Comparative Study for Predicting Student's performance" International Journal of Innovative Technology & Creative Engineering (ISSN;2045-711) Vol.1 no.12 December.

[4]. C. MARQUEZ-VERA, C. ROMERO and S. VENTURA "Predicting School Failure Using Data Mining" 2011

[5]. Sajadin Sembiring, M. Zarlis ET.AL." Prediction of Student Academic Performance by an Application of Data Mining Techniques"2011 International Conference on Management and Artificial Intelligence IPEDR vol.6(2011) IACSIT press, Bali, Indonesia.

[6]. K. Nandhiini and S. Saranya "ID3 Classifier for pupils Status Prediction" International Journal of Computer Application (0975-8887) vol 57-No.3, November 2012.

[7]. R.R. Kabra and R.S. Bichkar." Performance of Engineering Students using Decision Trees". International Journal of Computer Applications (0975-8887) volume 36-No.11, December 2011.

[8]. Mrinal Pandey and Vivek Kumar Sharma. "A Decision Tree Algorithm Pertaining to the Student Performance Analaysis and Prediction". International Journal of Computer Applications (0975-8887) volume 61-No.13, January 2013.

[9]. Ricardo Mendes and Joao P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics and Applications", IEEE,2017.

[10]. A. Dinesh Kumar, R. Pandi Selvam, K. Sathesh Kumar, "Review on Prediction Algorithms in Educational Data Mining", International Journal of Pure Applied and Mathematics (IJPAM), Volume-118.

[11]. J. John Kennedy, R. Pandi Selvam, "Cloud-Centric IoT based Decision Support System for Gestational Diabetes Mellitus using Optimal Support Vector Machine", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May 2019.

[12]. Maryam Zaffar, Manzoor Ahmed, K. S. Savita, "Performance Analysis of Feature Selection Algorithm for Educational Data Mining", IEEE Conference on Big Data and Analytics (ICBDA), 2017.

[13]. Jiawei Han, Micheline Kamber, Jian Pei," Data Mining Concepts and Techniques", 3rd Edition.

[14]. Athanasios S. Drigas, P. Lelipoulous, "The Use of Big Data in Education", Internal Journal of Computer Science Issues, Vol-11, September 2014.

[15]. R. Swathi, N. Pavan Kumar, L. KiranKranth, "Systematic Approach on Big Data Analytics in Education Systems", International Conference on Intelligent Computing and Control Systems (ICICCS), 2017.

[16]. A. Dinesh Kumar, R. Pandi Selvam, V. Palanisamy, "Prediction of Student Performance using Hybrid Classification", International Journal of Recent Technology and Engineering (IJRTE), November 2019.

[17]. Ramanathan L, Saksham Dhanda, Suresh Kumar D, "Predicting Students' Performance using Modified ID3 Algorithm", International Journal of Engineering and Technology (IJET), 2013.

[18]. https://www.softwaretestinghelp.com/data-mining-tools

Cite this article as :

Myo Thandar Tun, Yin Yin Htay, " Predict Students' Performance by Using J48 Algorithm, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 7, Issue 3, pp.578-582, May-June-2020. Available at

doi : https://doi.org/10.32628/IJSRSET2073124

Journal URL : http://ijsrset.com/IJSRSET2073124