International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Using ID3 Decision Tree Algorithm to the Student Grade Analysis and Prediction

Khin Khin Lay, San San Nwe

Associate Professor, University of Computer Studies, Maubin, Myanmar

How to cite this paper: Khin Khin Lay | San San Nwe "Using ID3 Decision Tree Algorithm to the Student Grade Analysis and Prediction" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1392-1395, **https://doi.org/10.31142/ijtsrd26545**

IJTSRD26545

ABSTRACT

Data mining techniques play an important role in data analysis. For the construction of a classification model which could predict performance of students, particularly for engineering branches, a decision tree algorithm associated with the data mining techniques have been used in the research. A number of factors may affect the performance of students. Data mining technology which can related to this student grade well and we also used classification algorithms prediction. In this paper, we used educational data mining to predict students' final grade based on their performance. We proposed student data classification using ID3(Iterative Dichotomiser 3) Decision Tree Algorithm.

KEYWORDS: *Classification, ID3, Data Mining, Decision Tree, Predicting Performance*

I. INTRODUCTION

Educational data mining is an interesting research area which extracts useful, previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process (Surjeet & Saurabh, 2012). The main functionality of data mining techniques is applying various methods and algorithms in order to discover and extract patterns of stored data. These interesting patterns are presented to the user and may be stored as new knowledge in knowledge base.

Data mining has been used in areas such as database systems, data warehousing, statistics, machine learning, data visualization, and information retrieval.

Data mining techniques have been introduced to new areas including neural networks, patterns recognition, spatial data analysis, image databases and many application fields such as business, economics and bioinformatics. Some types of data mining techniques are: Clustering, Association Rule Mining, Neural Networks, Genetic Algorithms, Nearest Neighbor Method, Classification Rule Mining, Decision trees and many others. The outcome of their results indicated that Decision Tree model had better prediction than other models.

A decision tree is a flow-chart-like tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits, which test the value of an expression of the attributes. Arcs from an internal node to its children are labelled with distinct outcomes of the test. Each leaf node has a class label associated with it.

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions.

From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome (Surjeet & Saurabh, 2012).

In data mining, decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorization and generalization of a given set of data. The four widely used decision tree learning algorithms are: ID3, CART, CHAID and C4.5.

II. RELATED WORK

In order to predict the performance of students the researcher took into consideration the work of other 14 A Decision Tree Approach for Predicting Students Academic Performance researchers that are in the same direction. Other researchers have looked at the work of predicting students' performance by applying many approaches and coming up with diverse results.

Three supervised data mining algorithms, i.e. Bayesian, Decision trees and Neural Networks which were applied by [1] on the preoperative assessment data to predict success in a course (to produce result as either passed or failed) and the performance of the learning methods were evaluated based on their predictive accuracy, ease of learning and user friendly characteristics. The researchers observed that that this methodology can be used to help students and teachers to improve student's performance; reduce failing ratio by taking appropriate steps at right time to improve the quality of learning.

[2] compared four different classifiers and combined the results into a multiple classifier. Their research divided the data into three (3) different classes weighing the features and using a genetic algorithm to minimize the error rate improves the prediction accuracy at least 10% in the all cases of 2, 3 and 9-Classes. In cases where the number of features is low, the feature weighting worked much better than feature selection. The successful optimization of student classification in all three cases demonstrates the merits of using the LON-CAPA data to *predict* the students‟ final grades based on their features, which are extracted from the homework data. However, the research in this case was based on an online course as opposed to the regular classroom class that the present study considers.

Furthermore, [3] observed that in the problem of prediction of performance, it is possible to automatically predict students' performance. Moreover by using extensible classification formalism such as Bayesian networks, which was employed in their research it becomes possible to easily and uniformly integrate such knowledge into the learning task. The researchers‟ experiments also show the need for methods aimed at predicting performance and exploring more learning algorithms.

Also, [8] used Iterative Dichotomiser 3 (ID3) decision tree algorithm to predict the university students‟ grade of a university in Nigeria. A prediction accuracy of 79,556 was obtained from the model. They further suggested the use of other decision based model to predict student's performance.

III. OUR PROPOSED METHOD

A. The ID3 Decision Tree

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric-information gain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric - information gain. To find an optimal way to classify a learning set we need some function which provides the most balanced splitting. The information gain metric is such a function. Given a data table that contains attributes and class of the attributes, we can measure homogeneity of the table based on the classes. The index used to measure degree of impurity is Entropy [2]. The Entropy is calculated as follows: Splitting criteria used for splitting of nodes of the tree is Information gain. To determine the best attribute for a particular node in the tree we use the measure called Information Gain.

B. Advantage of ID3

- Understandable prediction rules are created from the training data.
- Builds the fastest tree.
- Builds a short tree.
- Only need to test enough attributes until all data is classified.
- Finding leaf nodes enables test data to be pruned, reducing number of tests.

C. Disadvantage of ID3

- Data may be over-fitted or over classified, if a small sample is tested.
- Only one attribute at a time is tested for making a decision.
- Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum.

IV. Data Preparation

The first step in this paper is to collect data. It is important to select the most suitable attributes which influence the student performance. We have training set of 30 under graduate students. We were provided with a training dataset consisting of information about students admitted to the first year in Table I.

TableI Training Data Set

Sr. no.	Roll no.	Attend-ance	Apti- tute	Assign-ment	Test	Presentation	Grade
1	IT1	Good	Avg	Yes	Pass	Good	Excellent
2	IT2	Good	Avg	Yes	Pass	Good	Excellent
3	IT 3	Good	Avg	Yes	Pass	Good	Excellent
4	IT4	Good	Avg	Yes	Pass	Good	Excellent
5	IT5	Good	Avg	Yes	Pass	Good	Excellent
6	IT6	Avg	Avg	Yes	Pass	Avg	Good
7	IT7	Poor	Good	Yes	Pass	Avg	Good
8	IT8	Avg	Good	Yes	Pass	Avg	Good
9	IT9	Avg	Good	Yes	Pass	Avg	Good
10	IT10	Poor	Poor	No	Fail	Poor	Fail
11	IT11	Poor	Poor	No	Fail	Poor	Fail
12	IT12	Avg	Age	Yes	Pass	Age	Good
13	IT13	Good	Good	Yes	Pass	Good	Excellent
14	IT14	Good	Good	Yes	Pass	Good	Excellent
15	IT15	Good	Good	Yes	Pass	Good	Excellent

16	IT16	Good	Good	Yes	Pass	Good	Excellent
17	IT17	Good	Avg	Yes	Pass	Good	Excellent
18	IT18	Good	Avg	Yes	Pass	Good	Excellent
19	IT19	Good	Avg	Yes	Pass	Good	Excellent
20	IT20	Good	Poor	Yes	Pass	Good	Excellent
21	IT21	Good	Poor	Yes	Pass	Good	Excellent
22	IT22	Good	Poor	Yes	Pass	Good	Excellent
23	IT23	Good	Poor	Yes	Pass	Good	Excellent
24	IT24	Good	Poor	Yes	Pass	Good	Excellent
25	IT25	Poor	Poor	No	Fail	Poor	Fail
26	IT26	Avg	Good	Yes	Pass	Avg	Good
27	IT27	Poor	Good	No	Fail	Poor	Fail
28	IT28	Good	Good	Yes	Pass	Good	Excellent
29	IT29	Good	Good	Yes	Pass	Good	Excellent
30	IT30	Good	Good	Yes	Pass	Good	Excellent

To work out the information gain for A relative to S, we first need to calculate the entropy of S(Grade). Here S(Grade) is a set of 30 examples are 20"Excellent(Ex)", 6 "Good(G)" and 4 "Fail(F)".

$$\text{Entropy}(S) = - P_{Ex}\log_2(P_{Ex}) - P_G\log_2(P_G) - P_F\log_2(P_F) \quad (1.1)$$

$$= - [20/30]\log_2[21/30] - [6/30]\log_2[6/30]$$

$$- [4/30]\log_2[4/30]$$

$$= 1.241946$$

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A in Table II, relative to a collection of examples S,

$$\text{Gain}(S, \text{Attendance}) = \text{Entropy}(S) - \frac{|S_G|}{|S|}\text{Entropy}(S_G)$$

$$- \text{Entropy}(S) - \frac{|S_{Avg}|}{|S|}\text{Entropy}(S_{Avg})$$

$$- \text{Entropy}(S) - \frac{|S_{Poor}|}{|S|}\text{Entropy}(S_{Poor}) \quad (1.2)$$

$$= 1.241946 - 0.1203213$$

$$= 1.1216247$$

Table II Information Gain Value Table

Gain	Value
Gain(S, Attendance)	1.1216247
Gain(S, Aptitude)	0.234518
Gain(S, Assignment)	0.5665102
Gain(S, Test)	0.5665095
Gain(S, Presentation)	1.241946

Therefore, "Presentation" attribute is the decision attribute in the root node. "Presentation" as root node has three possible values – Good, Average, Poor. as shown in figure 1.

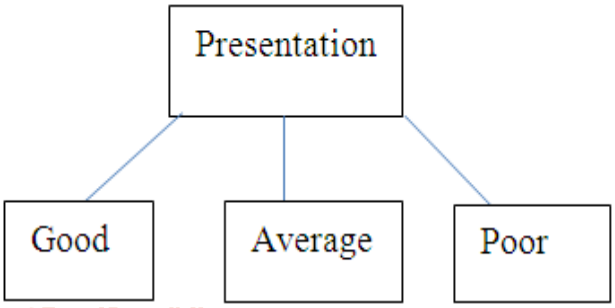

Figure1. Presentation as rood node

This process goes on until all data classified perfectly or run out of attributes. The knowledge represented by decision tree can be extracted and represented in the form of IF-THEN rules in figure II.

IF Presentation = "Good" AND Attendance = " Good" THEN Grade = "Excellent"
IF Presentation = "Average" AND Test = " Pass" THEN Grade = "Good"
IF Presentation = " Poor" AND Test = " Fail" THEN Grade = "Fail"

Figure2. Rule Set generated by Decision Tree

V. CONCLUSIONS

A classification model has been proposed in this study for predicting student's grades particularly for IT under graduate students. In this paper, the classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Aptitude, Presentation and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester.

REFERENCES

[1] Osmanbegovic E., Suljic M. "Data mining approach for predicting student performance" Economic Review-Journal of Economics and Business. Volume 10(1) (2012)

[2] Behrouz, M, Karshy, D, Korlemeyer G, Punch, W. "Predicting student performance: an application of data Mining methods with the educational web-based system" Lon-capa. 33rd ASEE/IEEE Frontiers in Education Conference. Boulder C.O. USA, (2003).

[3] Bekele, R., Menzel, W. "A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students". Journal of Information Science (2013).

[4] Kovacic, Z. "Early prediction of student success: Mining student enrollment data" Proceedings of Informing Science & IT Education Conference. (2010).

[5] Surjeet K, Yadav, Bharadwaj, B. Pal B." Data Mining Applications: A comparative Study for Predicting Student's performance." International journal of innovative technology & creative engineering. Volume 1(12). (2012).

[6] Nnamani, C. N, Dikko, H. G and Kinta, L. M. "Impact of students‟ financial strength on their academic performance: Kaduna Polytechnic experience". African Research Review 8(1), (2014).

[7] Ogunde A.O., Ajibade D.A. "A data Mining System for Predicting University Students F=Graduation Grade Using ID3 Decision Tree approach", Journal of Computer Science and Information Technology, Volume 2(1) (2014).

[8] Undavia, J. N., Dolia, P. M.; Shah, N. P. "Prediction of Graduate Students for Master Degree based on Their Past Performance using Decision Tree in Weka Environment". International Journal of Computer Applications; Volume 74 (21), (2013). https://www.wekatutorial.com/