Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 2, 4903 – 4907

DECISION TREE ACADEMIC PERFORMANCE MODEL FOR PRIMARY SCHOOL STUDENTS

[1]Dr. Manmohan Singh, [2]Dr. Monika Vyas, [3]Dr. Richa Chaudhary, [4]Urmila S Soni

[1]Department of Computer Science and Engineering, IES College of Technology, Bhopal, India, kumar.manmohan4@gmail.com
[2]Department of Civil Engineering, IES College of Technology, Bhopal, India, monikavyas06@gmail.com
[3]Amity Law School, Amity University Jaipur , Rajasthan India, richa.sa1123@gmail.com
[4]Deputy Director IES University, Bhopal , India, urmilas14@gmail.com

Abstract

Decision trees classifiers are easy and prompt data classifiers as supervised learning. Usually used in data mining to study the data and generate the tree and its rules that will be used to originate predictions. This study represents an implementation of a J48 algorithm on data collected from primary student. The aim of this study is developing a decision tree model to learning classification rules for primary students' data. This study is an attempt to use the data mining processes, particularly classification, to help in enhancing the quality of the primary educational system by evaluating student data to study the main attributes that may affect the student performance in primary classes. For this purpose, the data mining is used for mining student related academic data over the previous year. The classification rule generation process is based on J48 classifier.

Keywords: Data, Decision Tree, Data Mining, WEKA.

1. INTRODUCTION

Data mining is the process of autonomously extracting useful information or knowledge from large datasets. It involves the use of complicated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. Data mining is a step of KDD Process. Knowledge Discovery in Databases (KDD) is the process of extracting models and patterns from large databases. Data Mining refers to the process of applying the discovery algorithm to the data. This research has important contribution. Our results provide insight into the entire process of applying data mining tools to real-world data sets. In the following section we describe the overall methodology of the research, from selection of a data-mining algorithm to create a modeling of the academic performance prediction problem for primary school students. Next, we give the brief description of decision tree and Data mining tools WEKA. Finally, we discuss the practical importance of this research and our conclusions (Kumar, N.V.A. and G.V. Uma, 2009).

The various techniques of data mining like classification, clustering and rule mining can be applied to bring out various hidden knowledge from the educational data. Prediction can be classified into: Classification, regression and density estimation. In classification, the predicted variable is a binary or categorical variable. Some popular classification methods include decision trees, logistic regression and support vector machines. In regression, the predicted variable is a continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks and support vector machine regression. Classification techniques like decision trees, Bayesian networks can be used to predict the student's behavior in an educational environment, his interest towards a subject or his outcome in the examination (Bhardwaj, B.K. and S. Pal, 2011) .

1.1. Decision Tree

The concept of decision trees was developed and refined over many years by (Bhardwaj, B.K. and S. Pal, 2011) starting with A Decision tree is a classification schemes which generate a tree and a set of rules, representing the model of different classes from a given dataset. As per (Han, J. and M. Kamber, 2006)

Decision tree is a flow chart like structure, where each internal node denotes a test on the an attribute, each branch represents an outcome of the test and leaf nodes represent the classes or class distributions We have used J48 in WEKA to do the prediction analysis. Decision trees are generated from the training data in a top-down direction. The root node of a decision tree is the trees initial state-the first decision node. Each node in a tree contains some data. On a basis of an algorithm some calculations are completed and the decision tree node is been split into two or more branches. In some cases, the node cannot be split, in this case it will be the final decision node.

2. METHODOLOGY

This section describes the process we followed to collect and analyze the academic performance. We discuss our selection of a data-mining tool, followed by the difficult task of preparing the data for analysis. We present our model of the academic performance prediction problem.

2.1. Source of Database and Description

Database has collected by filling the questionnaires by concerning student or teacher or student parent. The survey was designed to gather information pertaining to the perceived educational status of parents and demographic information of student such as name, address, age, sex, education. The survey consisted of 26 questions. Some questions were to be answered yes or no, but generally respondents were provided with more options to answer the questions. The data was originally represented in excel data format in the form of two dimensional table consisting of 373 instances with each data point corresponding to the responses of an individual's, the dataset was converted into Attribute Relation File Format (ARFF) for effective and efficient usage WEKA system. Table 1 shows the description of each attributes of database(García, E.P.I. and P.M. Mora, 2011.)

2.2. Preparing the Data and Selecting the Relevant Attribute

In the data preparation phase we selected the relevant attributes from the available data, created meaningful groups within the attributes and derived new attributes from our knowledge of the domain.

The information gain with respect to a set of examples is the expected reduction in entropy that results from splitting a set of examples using the values of that attribute. This measure is used in Decision Tree induction and is useful for identifying those attributes that have the greatest influence on classification. The aim of data preprocessing is to improve the quality of the data which will help in improving "the accuracy and efficiency of the subsequent mining process" (García, E.P.I. and P.M. Mora, 2011). Often, outliers decrease the accuracy and efficiency of the models. Data preprocessing allows transforming the original data into a suitable shape to be used by a particular mining algorithm. So, before applying the data mining algorithm, a number of general data preprocessing tasks have to be addressed (Ramesh et al., 2011). Normally in data mining process preprocessing is one of the important stages where relevant data's are grouped and cleaned, this can be done with any of the classification algorithms and in this study we take J48 classifier with the help of WEKA software.

2.3. Building the Classification Model

The next step is to build the classification model using the decision tree method. The decision tree is a very good and practical method since it is relatively fast and can be easily converted to simple classification rules. The decision tree method depends mainly on using the information gain metric which determines the attribute that is most useful. The information gain depends on the entropy measure.

2.4. Experimental Setup

This section present the class attributes details and which parameters have taken in during creating a decision tree model. Class attribute

consist four classes as shown in Fig. 1 and parameter setting is shown in Fig. 2.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -R -N 3 -Q 1 -B -M 2

Relation: edudata-weka.filters.unsupervised.attribute.Remove-R1-3,5-8-weka.filters.unsupervised.attribute.Remove-R1,15-weka.filters.unsupervised.attribute.Remove-R3-weka.filters.unsupervised.attribute.Remove-R1,7-weka.filters.unsupervised.attribute.Remove-R5

Instances: 373

Attributes: 13

Test mode: evaluate on training data

=== Classifier model (full training set) ===

=== Summary ===

Correctly Classified Instances 228 61.126%

Incorrectly Classified Instances 145 38.874%

Table 1. *Description of datasets*

S. No.	Attribute name	Description
1	School_Code	School dice code
2	Name_Place	Place of school
3	Name_Block	Name of block (three blocks of Batul District))
4	District	Batul (M.P.)
5	Scholer_Number	Student scholar number
6	Name_Student	Name of student
7	Student_Father_Name	Student father name
8	Student_Mother_Name	Mother's name
9	Age	Age of student (06-10 years)
10	Sex	Gender (M, F)
11	Class	(III, IV, V)
12	Category	Category (SC, ST, Gen, OBC)
13	School_type	(Govt., Private)
14	Location_School	Rural of Urban
15	No_teachers	Number of Teacher's in school
16	family_size	Number of members of in a student family
17	Living_zone	Residential area of student (Rural, Urban)
18	Father_edu	Father's Education
19	Father_occup	Occupation of student father
20	Mother_edu	Mother; Education
21	Mother_occu	Occupation of Student's Mother
22	Family_income	Family income
23	Private_tuision	Are student take private tusion?
24	Attendence_School	Attendence of student's in a class
25	Previous_result	Previous year result of student in percentage
26	Grade_Previous_Result	Previous year result of student in Grad

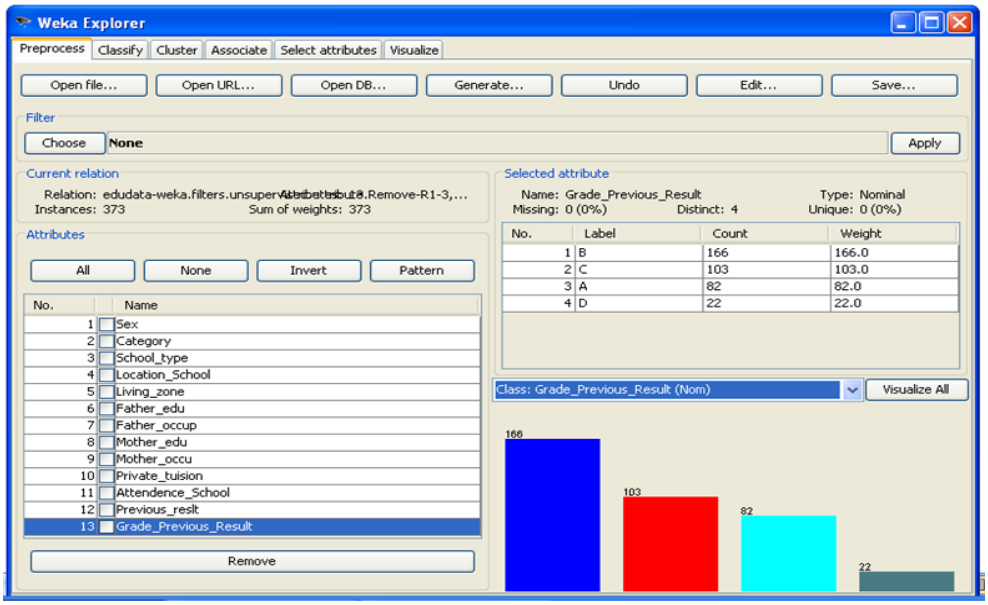

Fig. 1. *View of class attribute*

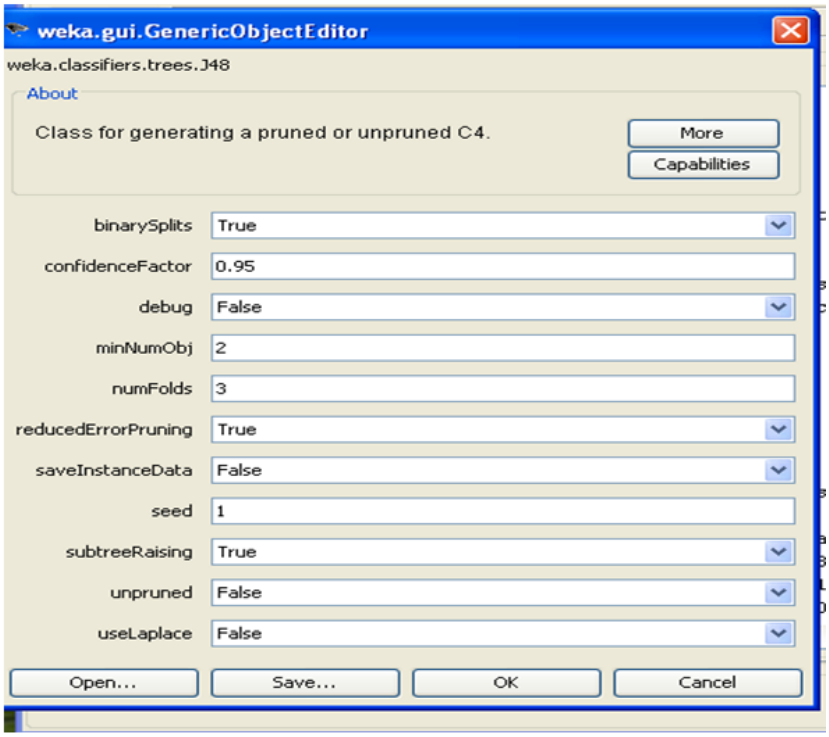

Fig. 2. *Parameter setting of experiment*

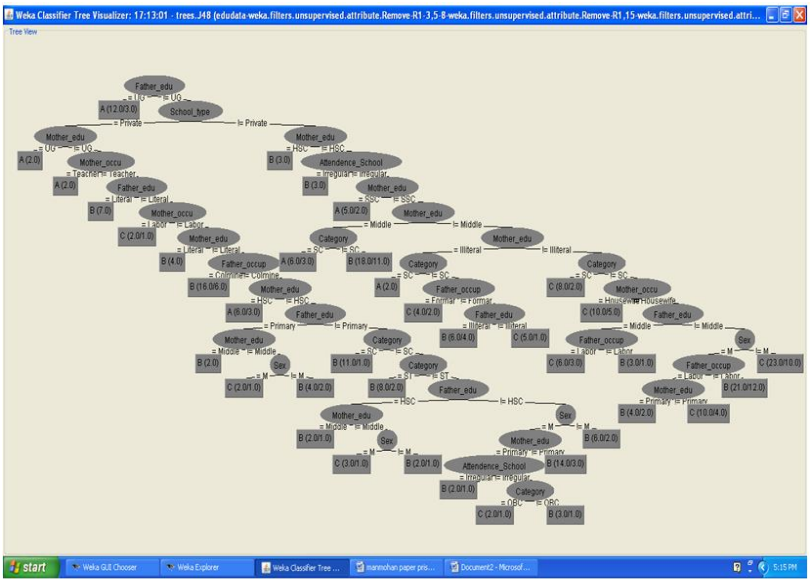

Fig. 3. *Generated decision tree with J48 classifier*

3. CONCLUSION

This study we have generated decision tree model which is shown in Fig. 3. We can easily extract if…..then rules from decision tree. Our aim is to generate some valuable if…then rules from student data. These rules may be useful for taking decisions to improve academic performance of primary school student data.

Reference

[1] Bhardwaj, B.K. and S. Pal, 2011. Data mining: A prediction for performance Improvement using classification. Int. J. Comput. Sci. Information Security, 9.

[2] García, E.P.I. and P.M. Mora, 2011. Model prediction of academic performance for first year students. Proceeding of the 10th Mexican International Conference on Artificial Intelligence, pp: 169-174.

[3] Han, J. and M. Kamber, 2006. Data Mining: Concepts and Techniques, 2nd Edn., Boston, MA: Elsevier.

[4] Han, J. and M. Kamber, 2008. Data Mining Concepts and Techniques. 2nd Edn., Morgan Kaufmann Publishers.

[5] Kumar, N.V.A. and G.V. Uma, 2009. Improving academic performanceof students by applying data mining technique. Eur. J. Sci. Res., 34: 526-534.

[6] Ramesh, V., P. Parkavi and P. Yasodha, 2011. Performance analysis of data mining techniques for placement chance prediction. Int. J. Sci. Eng. Res., 2.

[7] Youssef, A.B. and M. Dahmani, 2010. The impact of ICT's on students' performance in Higher Education: Direct effects. Indirect effects and Organizational change.