IMPACT: International Journal of Research in
Engineering & Technology (IMPACT: IJRET)
ISSN(E): 2321-8843; ISSN(P): 2347-4599
Vol. 2, Issue 7, Jul 2014, 127-132
© Impact Journals

DETECTION OF ABNORMAL VALUES IN THE RESULT SHEETS USING EDUCATIONAL DATAMINING

BHAVNA KHATRI[1], VIMAL N. PANDEY[2] & PRADEEP CHOUKSEY[3]

[1]Research Scholar, AISECT University, Bhopal, Madhya Pradesh, India

[2]Professor, H.K. Arts College, Ahmedabad, Gujarat, India

[3]Vice Principal, TIT Group of Institutions, Bhopal, Madhya Pradesh, India

ABSTRACT

Educational Datamining (EDM) is an emerging dicipline, concerned with developing methods for exploring the unique types of data that come from educational system and using those methods to better understand students, and the system which they learn in.

This paper is designed to justify the capabilities of datamining techniques in context of higher education by offering a datamining model for higher education system in technical institution. In this, we are proposing a detection techniques for detecting abnormal values in the student's result sheets. For this we are applying dataminig techniques like classification, decision tree etc. on the huge educational data, for finding errors in the sheets with respect to score or grade or any calculation mistakes.

KEYWORDS: Decision Tree, Educational Data Mining (EDM), Classification, WEKA

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 technical education 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.

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.

Decision Tree

The concept of decision trees was developed and refined over many years by (Han, J., & Kamber, M. 2006) starting with (Rud, 2001). 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 and Kamber (2000) 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.

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.

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.

Table 1: Description of Datasets

S. No.	Attribute Name	Description
1	College_Code	College code
2	Name_Place	Place of college
3	Name_Block	Name of block
4	City	Khandwal (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	College_Type	(Govt., Private)
14	Location_College	Rural of urban
15	No_Faculty	Number of faculty's in college

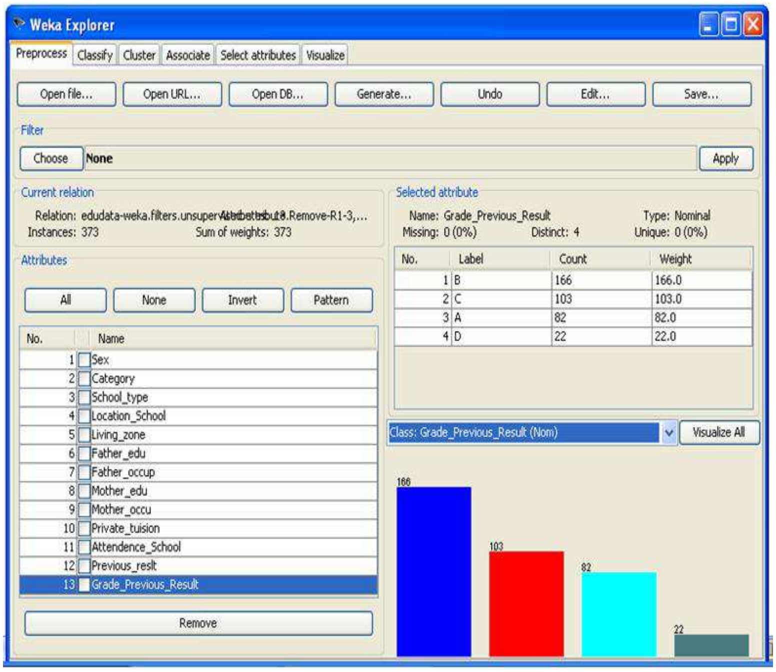
Table 1: Contd.,

16	Family_Size	Number of members of in astudent family
17	Living_Zone	Residential area of student
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_College	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

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” (Han and Kamber 2007). 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 (V. Ramesh, at all 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.

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.



Filter: None

Relation: edudata-weka.filters.unsupervised.attribute.Remove-R1-3,...
Instances: 373 Sum of weights: 373

No.		Name
1	□	Sex
2	□	Category
3	□	School_type
4	□	Location_School
5	□	Living_zone
6	□	Father_edu
7	□	Father_occup
8	□	Mother_edu
9	□	Mother_occu
10	□	Private_tuision
11	□	Attendence_School
12	□	Previous_reslt
13	□	Grade_Previous_Result

Name: Grade_Previous_Result Type: Nominal
Missing: 0 (0%) Distinct: 4 Unique: 0 (0%)

No.	Label	Count	Weight
1	B	166	166.0
2	C	103	103.0
3	A	82	82.0
4	D	22	22.0

Class: Grade_Previous_Result (Nom)

Figure 1: View of Class Attribute

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.

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 Figure 1 and parameter setting is shown in Figure 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 ===

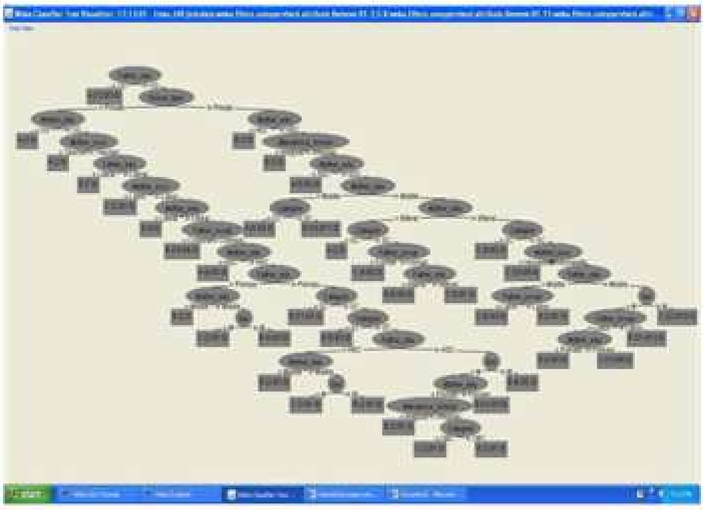
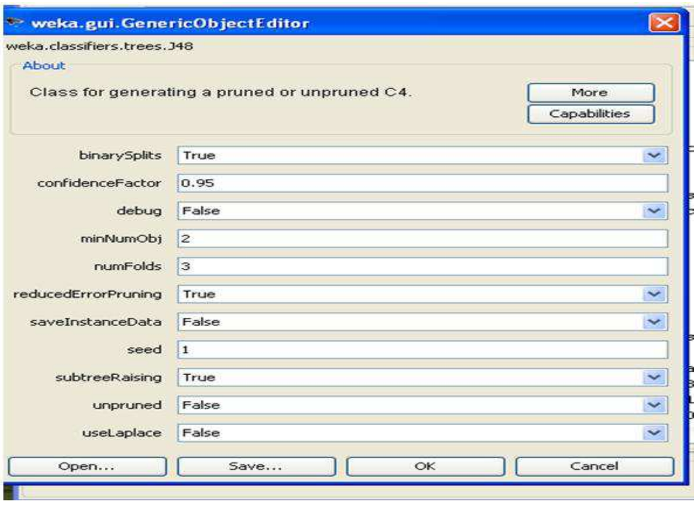


Figure 2: Parameter Setting of Experiment

Table 2

Correctly Classified Instances	228	61.126%
Incorrectly Classified Instances	145	38.874%



weka.gui.GenericObjectEditor

weka.classifiers.trees.J48

About

Class for generating a pruned or unpruned C4.

binarySplits	True
confidenceFactor	0.95
debug	False
minNumObj	2
numFolds	3
reducedErrorPruning	True
saveInstanceData	False
seed	1
subtreeRaising	True
unpruned	False
useLaplace	False

Figure 3: Generated Decision Tree with J48 Classifier

CONCLUSIONS

This study we have generated decision tree model which is shown in Figure 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 technical college student data.

REFERENCES

1. Osman N. Darcan and Bertan Y. Badur, 2012 “Student Profiling on Academic Performance Using Cluster Analysis” IBIMA Publishing Journal of e-Learning & Higher Education Article ID 622480, 8 pages DOI: 10.5171/2012.622480

2. Brijesh Kumar Bhardwaj and Saurabh Pal, 2011” Data Mining: A prediction for performance Improvement using classification” (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April.

3. V. Ramesh, P. Parkavi, P. Yasodha, 2011, “Performance Analysis of Data Mining Techniques for Placement Chance Prediction” International Journal of Scientific & Engineering Research Volume 2, Issue 8, August, ISSN 2229-5518.

4. Ernesto Pathros Ibarra García, Pablo Medina Mora, 2011"Model Prediction of Academic Performance for First Year Students," micai, pp.169-174, 2011 10th Mexican International Conference on Artificial Intelligence.

5. Adel Ben Youssef and Mounir Dahmani, “The impact of ICT’s on students’ performance in Higher Education: Direct effects, Indirect effects and Organizational change”, 2010.

6. N. V. Anand Kumar, G. V. Uma (2009), Improving Academic Performance of Students by Applying Data Mining Technique, European Journal of Scientific Research ISSN 1450-216X Vol.34 No.4, pp.526-534.

7. Jaiwei Han and Micheline Kamber, 2008 “Data Mining Concepts and Techniques”, Second Edition Morgan Kaufmann Publishers.
8. M. Bray 2007, the Shadow Education System: Private Tutoring And Its Implications For Planners, (2nd ed.), UNESCO, PARIS, France.
9. J. A. Moriana, F. Alos, R. Alcala, M. J. Pino, J. Herruzo, and R. Ruiz, 2006 “Extra Curricular Activities and Academic Performance in Secondary Students”, Electronic Journal of Research in Educational Psychology, Vol. 4, No. 1, pp. 35-46.
10. Pardos Z.; Heffernan N.; Anderson B.; and Heffernan C., 2006. Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks. In Proc. of 8th Int. Conf. on Intelligent Tutoring Systems. Taiwan.
11. Han, J., & Kamber, M. 2006. Data mining: Concepts and techniques (2nd ed.). Boston, MA: Elsevier.