DATA MINING CLASSIFICATION OF TEACHING EVALUATION USING WEKA

Abstract- Teaching evaluation is an important process for every educational system. Identifying and understanding areas of pitfalls plays a vital role for the educational system development. Data mining techniques are helpful for this purpose. Preprocessing and classification are two important data mining techniques for knowledge discovery process that are widely used in various fields. Preprocessing transforms the raw data into a understandable format and classification classifies each item in a set of data into one of the predefined set of classes. In this paper, I have considered few constraints in the form of dataset for analyzing the teacher performance and the results are presented using weka 3.7 tool. Weka is one of the data mining tools which possess many machine learning algorithms. The paper also provides information about basics of preprocessing and classification techniques. The result provided by the tool will be very helpful for the educational managers to take decisions on the teacher’s contract in future. To propose the result, we have used selected classification techniques for our data set.

Keywords – Weka, Preprocessing, classification, Data mining

1. INTRODUCTION

In general, teacher evaluation refers to the formal process used to review and rate teacher’s performance and effectiveness in the classroom. Ideally, the findings from these evaluations are used to provide feedback to teachers and guide their professional development. Data mining techniques analyze data from different perspectives of teaching and summarize into important information to identify hidden patterns. Data preprocessing includes data cleaning, data integration, data selection and data transformation. Classification technique is used for gathering important information about data so that data sets can be placed into proper categories. One example of such classification of email will be either regular, acceptable email or as spam. Han and Kamber (2000) describes different data mining techniques which allow the users to analyze data from different dimensions, categorize it and summarize the relationships that are identified during the mining process and also explained clearly the process of changing data during processing [1]. Chein and Chen (2008) used several attributes in the form of groups to measure teacher quality. Some of them are cooperativeness, open-mindedness, reliability, efficiency, sense of humor, good voice etc. These also provide the performance measurement of teaching [2]. Pal et al., (2013) observed various aspects of teaching performance with few attributes. He designed a model which combines the knowledge and expertise of human experts which makes very easy for the educational institutes in decision making [3]. Khan (2005) conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with an objective to establish the prognostic value of different measures of personality, cognition and demographic variables for success at higher secondary level in science stream [4]. Cluster technique is used on the entire data set which gives different groups and from these groups random sample data is selected for the analysis. Ayesha et al., (2010) explained about various data mining techniques- k-means, apriori and k-nearest neighbor classification etc [5]. The paper analyzed the performance of students using attendance, lab work, counseling, internal examinations, parental interaction, university examination attributes. A sample data set with thirty students are considered for performance measurement. WEKA 3.7 software tool implemented in java environment which uses ARFF file [6]. "WEKA" stands for the Waikato Environment for Knowledge Analysis, which was developed at the University of Waikato in New Zealand. The GUI Chooser consists of four buttons and these buttons are used to start the following applications: Explorer: An environment for exploring data with WEKA (the rest of this documentation deals with this application in more detail) Experimenter: An environment for performing experiments and conducting statistical tests between learning schemes. Knowledge Flow: This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning. Simple CLI: Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.

2. METHODOLOGY

WEKA is an open source software that offers various machine learning and data mining algorithms for data preprocessing, classification, regression, clustering and association rules.

1 Assistant Professor, Department of Information Technology, Kavikulguru Institute of Technology and Science, Ramtek, Nagpur, Maharashtra, India
The individual process steps employed in data analysis will be: Data collection -> Preprocessing -> Classification -> Results

3. DATA COLLECTION
In our work, we considered sample data from a technical institute, located in the city of Nagpur. On the basis of data collected some attributes have been considered to predict the teacher teaching performance. The attributes used for judging teaching performance includes: 
Subject _Teaching, Communication_ Skills, Case_ Studies and Conduct

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Data Types</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject _Teaching</td>
<td>Nominal</td>
<td>Weak, Strong, Excellent</td>
</tr>
<tr>
<td>Communication_ Skills</td>
<td>Nominal</td>
<td>Good, Average, Poor</td>
</tr>
<tr>
<td>Case_ Studies</td>
<td>Nominal</td>
<td>Few, Large</td>
</tr>
<tr>
<td>Conduct</td>
<td>Nominal</td>
<td>Low, Better</td>
</tr>
</tbody>
</table>

4. DATA PROCESSING
One of the important steps in data mining is data pre-processing. Data pre-processing is used to fill missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies. We used the data in terms of percentage for the attributes considered in the data set.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Range</th>
</tr>
</thead>
</table>
| Subject _Teaching  | Subject _Teaching: 90% = Excellent  
75% >= Subject _Teaching: 90% = Strong  
75% < Subject _Teaching: 90% = Weak |
| Communication_ Skills | Communication_ Skills: 85% = Good  
70% >= Communication_ Skills: 85% = Average  
70% < Communication_ Skills: 85% = Poor |
| Case_ Studies      | Case_ Studies: 90% = Large  
70% >= Case_ Studies: 90% = Poor |
| Conduct            | Conduct: 70% = Better  
70% < Conduct: 70% = Low |

5. EXPERIMENTAL WORKS AND RESULTS
The primary data required is designed in an excel sheet and this data is applied to the weka 3.7 environment for data classification (Fig. 1). As said, I have used weka tool for analysis.. The algorithms can be applied directly to the data set. When we start the weka tool, the following screen appears (Fig. 2). The figure 2 has of four different applications: Explorer, Experimenter, Knowledge Flow and Simple CLI. (i) Explorer, is an environment for exploring data; (ii) Experimenter, is an environment for performing experiments and conducting statistical tests between learning schemes; (iii) Knowledge Flow, is a Java-Beans-based interface for setting up and running machine learning experiments. For the exercises in this tutorial; (iv) Simple CLI, provides a simple command-line interface and allows direct execution of Weka commands. Now start the explorer application by a click and select the preprocess button followed by open file in the application. Whenever open file is selected, it will ask about the input data set .Provide the appropriate input data set for preprocess (Fig. 3). Figure 3 represents the results of preprocess. Left panel has the attributes of the data set and right panel has the attribute selection with count and weight measures and also it shows the analysis in the form of bar charts. There is also an option to select the class that to be considered. I have considered conduct class (Fig. 4). After preprocess, next select the classify button which provides the classification of the attributes. By default ZeroR technique will be applied. User has the option to select their own required technique.Apart from techniques, test options are also present. I have used cross validation with 10 folds and J48 classifier for our data set. After selecting J48, click on start. Then the result analysis will be displayed on the right panel. The result information contains about the algorithm used, number of folds, relation name, attributes, classifier model, correctly classified and in correctly classified instances etc.

== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: TeachingDataSet1
Instances: 14
Attributes: 4
Subject_ Teaching
Communication_Skills
Case_ Studies
Conduct
Test mode: 10-fold cross-validation

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

J48 pruned tree

Subject_Teaching = Weak: Poor (5.0/2.0)
Subject_Teaching = Strong: Better (4.0/1.0)

Subject_Teaching = Excellent
  | Case_Studies = Few: Better (3.0)
  | Case_Studies = Large: Poor (2.0)

Number of Leaves : 4
Size of the tree : 6
Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 6 42.8571 %
Incorrectly Classified Instances 8 57.1429 %

Kappa statistic -0.2727
Mean absolute error 0.5288
Root mean squared error 0.5915
Relative absolute error 104.7714 %
Root relative squared error 116.1374 %
Coverage of cases (0.95 level) 92.8571 %
Mean rel. region size (0.95 level) 89.2857 %
Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class
0 0.25 0 0 0 0.271 Poor
0.75 1 0.5 0.75 0.6 0.271 Better

Weighted Avg. 0.429 0.679 0.286 0.429 0.343 0.271

=== Confusion Matrix ===

a b <--- classified as
0 6 | a = Poor
2 6 | b = Better

By right click on the result list provides a menu which has an option “visualize tree”. Select the visual tree. This will display the classification by using a decision tree format (Fig 5 and 6).
Data Mining Classification Of Teaching Evaluation Using Weka

Figure 2. Weka GUI chooser

Figure 3. Pre-processing data set

Figure 4. Classification using cross-validation with run information
6. CONCLUSION
Teacher evaluation is very important factor that to be considered in the education field to strengthen the country. The results provided by the weka 3.7 tool provides analysis of teaching performance and visualized by the decision tree for classification on subject teaching into classes poor and better. The results very helpful for the education managers to take a decision on the service contracts of teachers in the educational institutions.

7. REFERENCES