

A Novel Approach for Professor Appraisal System In Educational Data Mining Using WEKA

¹Thupakula Bhaskar (Asst.Professor), ²G.Ramakrishna (Asst.Professor)

¹Department of Computer Engineering, ²Department of Information Technology
Sanjivani College of Engineering, Savitribai Phule Pune University
Kopergaon (T), Ahmad Nagar (D), Maharashtra, India

Abstract- Data mining, the concept of unseen predictive information from big databases is a powerful novel technology with great potential used in various commercial uses including banking, retail industry, e-commerce, telecommunication industry, DNA analysis remote sensing, bioinformatics etc. Education is a required element for the progress of nation. Mining in educational environment is called Educational Data Mining. Educational data mining is concerned with developing new methods to discover knowledge from educational database. In order to analyze opinion of students about their teachers in Professor Appraisal System, this paper surveys an application of data mining in Professor Appraisal System & also present result analysis using WEKA tool. There are varieties of popular data mining task within the educational data mining e.g. classification, clustering, outlier detection, association rule, prediction etc. How each of data mining tasks can be applied to education system is explained. In this paper we analyze the performance of final Professor Appraisal of a semester of a computer engineering department, Sanjivani College of engineering & is presented the result which it is achieved using WEKA tool. We have verified hidden patterns of Professor Appraisal by students and is predicted that which Professor will be invited to faculty classes and which Professor will be refusing and department heads due to Appraisal reasons will ask explanations with these Professor.

Keywords- Classification, Clustering, Association rule, Data mining, Appraisal, WEKA.

I.INTRODUCTION

Data mining has involved a great deal of responsiveness in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the forthcoming need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration [1]. Manual data analysis has been around for some time now, but it creates a bottleneck for large data analysis. The transition won't occur automatically; in this case, there is a need for data mining. [2]. Mining applied in education was published in 1995 by Sanjeev and Zytow. Researchers gathered the knowledge discovery as terms like "P pattern for data in the range R" from university database [3]. Vranić and Skočir was examined how to improve some aspects of educational quality with data mining algorithms and techniques by taking a specific course students as target

audience in academic environments [4]. In this paper I have collected information and results of a appraisal about 30

professors in Sanjivani College of Engineering, Department of Computer Engineering on professor's performances in classroom then with data mining algorithms such Association Rule and decision trees (j48), it is proceeded to analyze and predict acceptance of a professor for continuing the teaching in that subject. There are new rules and relations between selected parameters such as Teaching, Professor Degree, Preparation, Communication, Class Control, Teaching experience, Approved Staff to next semesters on professor appraisal system that is interested for Heads of Departments of Institution.

II. Methodology

In this research study, We have followed a popular data mining methodology called Cross Industry Standard Process for Data Mining (CRISP-DM), which is a six-step process [5]:

- **Problem explanation:** Comprises understanding development goals with business perspective.
- **Understanding the data:** Includes identifying the sources of data.
- **Formulating the data:** Includes pre-processing, cleaning, and transforming the relevant data into a form that can be used by data mining algorithms.
- **Creating the models:** Includes developing a wide range of models using comparable analytical techniques.
- **Assessing the models:** Includes evaluating and assessing the validity and the utility of the models against each other and against the goals of the study.
- **Using the model:** Includes in such activities as deploying the models for use in decision making processes.

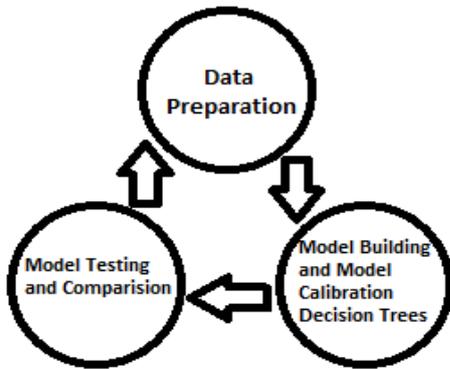


Fig.1.A graphical illustration of the methodology employed in this study

A. Data

In this study 34 records were used which is taken from feedback_2013_14_sem_1 of Department of computer engineering, Sanjivani College of Engineering. Dataset have professors' information such as Teaching, Preparation, Communication, Class Control along with this I have included Professor Degree, Professor Experience, Approved Staff.

Table 1. The list of independent variables used in this study

Variable Name	Data Type	Description
Teaching	Text	Teaching Score
Professor Degree	Text	Professors Degree
Preparation	Text	Preparation Score
Communication	Text	Communication Score
Class Control	Text	Class Control Score
Teaching experience	Text	Teaching experience of Professor
Approved Staff	Text	Approved Professor or not

Table 2. The list of independent variables and values used in this study

Variable Name	Data Type	Values
Teaching	Text	{Excellent,Good,Satisfactory,Poor}
Professor Degree	Text	{BE,ME,PHD}
Preparation	Text	{Excellent,Good,Satisfactory,Poor}
Communication	Text	{Excellent,Good,Satisfactory,Poor}
Class Control	Text	{Excellent,Good,Satisfactory,Poor}
Teaching experience	Text	{TRUE,FALSE}
Approved Staff	Text	{Yes,No}

Teaching score of professors which are studying in Sanjivani College of Engineering, Computer Engineering Department Faculty are represented by the word system. Score ranges of these words are shown in Table 3.

Table 3. The output variable (Evaluation score) used in the study

Raw Score	Nominal Representation
Score < 60	Poor
60<=Score < 75	Satisfactory
75<=Score< 85	Good
85<=Score<= 100	Excellent

Table 4. The output variable (Teaching experience) used in the study

Raw –Years of Teaching	Nominal Representation
Years < 3	False
Years >= 3	TRUE

B. Background

In this research I have used WEKA and Data mining. The following subsections contain a summary of these topics.

a. WEKA

WEKA is a group of machine learning algorithms for data mining tasks. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and Predictive modeling, together with graphical user interfaces for easy access to this functionality [6].

b. Data Mining

Data mining is the method of defining interesting knowledge from big amount of data stored in database, data warehouse or other information sources. It includes various tasks such as classification, clustering, association rule etc.

c. Association Rule

Association rules are used to show the relationship between data items. Association rule generation consists of two separate steps: First, minimum support is applied to find all frequent item sets in a database. Second, these frequent item sets and the minimum confidence constraint are used to form rules [6]. Support & confidence are the normal method used to measure the quality of association rule. Association rule can be used in educational data mining and professor’s appraisal system for analyzing the learning data.

d. Classification

Classification is a data mining task that maps the data into predefined groups & classes. It is also called as supervised learning .It consists of two steps:

1. Model construction: It consists of set of predetermined classes. Each sample is assumed to belong to a predefined class. The set of sample used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae.
2. Model usage: This model is used for classifying upcoming or unidentified objects. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur [6].

e. Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. Clustering can be considered the most important unsupervised learning technique. In educational data mining and professor’s appraisal system, clustering has been used to group the professors according to their behavior e.g. clustering can be used to distinguish active professor from non-active professor according to their performance in activities.

III. ARCHTECTURE OF PROPOSED SYSTEM

In this paper, it is done a feedback of 2013_14 academic year and semester-1 survey from 319 students then it is prepared results of this survey for 30 professors.

a. The Explorer Interface of WEKA

Initially “preprocess” will have been selected. This is the tab you select when you want to tell WEKA where to find the data set that you want to use.

WEKA processes data sets that are in its own ARFF format. Conveniently, the download will have set up a folder within the WEKA-3.7 folder called “data”.

b. ARFF format files

You do not need to know about ARFF format unless you wish to convert data from other formats. However, it is useful to see the information that such files provide to WEKA.

```
@relation Professor_Appraisal.symbolic

@attribute Teaching {Excellent, Good, Satisfactory, Poor}
@attribute Professor_Degree {BE, ME, PHD}
@attribute Preparation {Excellent, Good, Satisfactory, Poor}
@attribute Communication {Excellent, Good, Satisfactory, Poor}
@attribute Class_Control {Excellent, Good, Satisfactory, Poor}
@attribute Teaching_Experience {TRUE, FALSE}
@attribute Approved_Staff {Yes, No}

@data
Excellent, ME, Excellent, Excellent, Excellent, TRUE, Yes
Excellent, ME, Good, Good, Excellent, TRUE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, No
Excellent, ME, Good, Excellent, Good, TRUE, Yes
Excellent, ME, Excellent, Excellent, Excellent, TRUE, Yes
Good, ME, Good, Good, Good, TRUE, No
Satisfactory, ME, Satisfactory, Good, Satisfactory, TRUE, Yes
Satisfactory, ME, Satisfactory, Good, Satisfactory, TRUE, Yes
Good, ME, Excellent, Good, Excellent, FALSE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, No
Excellent, PHD, Excellent, Excellent, Excellent, FALSE, No
Good, ME, Good, Good, Satisfactory, TRUE, Yes
Good, ME, Good, Good, Good, TRUE, Yes
Good, ME, Good, Excellent, Excellent, FALSE, No
Good, BE, Good, Satisfactory, Satisfactory, FALSE, No
Good, ME, Excellent, Excellent, Excellent, FALSE, No
Good, ME, Good, Good, Good, TRUE, No
Excellent, BE, Good, Excellent, Excellent, TRUE, No
Good, ME, Good, Good, Good, TRUE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, Yes
Excellent, BE, Excellent, Excellent, Good, TRUE, No
Good, ME, Good, Satisfactory, Good, TRUE, Yes
Excellent, ME, Excellent, Excellent, Excellent, FALSE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, Yes
Good, BE, Good, Good, Good, FALSE, No
Good, BE, Good, Good, Good, TRUE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, No
Excellent, ME, Excellent, Excellent, Excellent, TRUE, Yes
Good, ME, Good, Good, Poor, FALSE, No
Excellent, ME, Excellent, Good, Excellent, TRUE, Yes
Excellent, ME, Good, Satisfactory, Good, TRUE, Yes
Good, BE, Good, Excellent, Good, FALSE, No
Good, ME, Good, Good, Good, TRUE, No
Good, ME, Good, Satisfactory, Satisfactory, FALSE, No
```

Fig.2. ARFF file format for dataset in this paper.

It consists of three parts. The @relation line gives the dataset a name for use within WEKA. The @attribute lines declare the attributes of the examples in the data set. Each line specifies an attribute’s name and the values it may take. In this paper the attributes have nominal values so these are listed explicitly. In other cases attributes might take numbers as values and in such cases this would be indicated as in the following example:

```
@attribute Teacher_Degree numeric
```

The remainder of the file lists the actual examples, in comma separated format; the attribute values appear in the order in which they are declared above.

c. Opening a data set.

In the Explorer window, click on “Open file” and then use the browser to navigate to the ‘data’ folder within the WEKA-3.7 folder. Select the file called Professor_Appraisal.arff. In this case, the normal usage is to learn to predict the ‘Approved_Staff’ attribute from four others providing information about the Professor evaluation.

The Explorer window should now look like this:

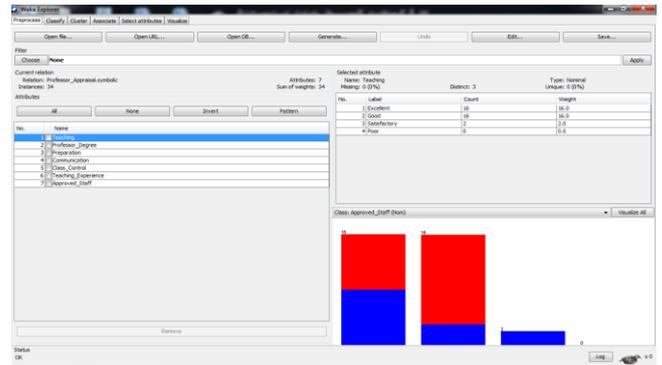


Fig.3. Opening ‘Professor appraisal’ data set in WEKA

Most of the information it displays is self-explanatory: it is a data set containing 34 examples (instances) each of which has 7 attributes. The ‘Approved_Staff’ attribute has been suggested as the class attribute (i.e. the one that will be predicted from the others). Most of the right hand of the window gives you information about the attributes. Initially, it will give you information about the first attribute (‘Teaching’). This shows that it has 4 possible values tells you how many there are of each value. The bar chart in the lower right shows how the values of the suggested class variable are distributed across the possible values of the ‘Teaching’. If you click on ‘Professor_Degree’ in the panel on the left, the information about the ‘Teaching’ attribute will be replaced by the corresponding information about the Teacher Degree attribute.

d. Choosing a classifier

Next it is necessary to select a machine learning procedure to apply to this data. The task is classification so click on the ‘classify’ tab near the top of the Explorer window. By default, a classifier called ZeroR has been selected. A different classifier is desired so click on the Choose button. A hierarchical popup menu appears. Click to expand ‘Trees’, which appears at the end of this menu, then select J48 which is the decision tree program. The Explorer window now looks like this indicating that J48 has been chosen.

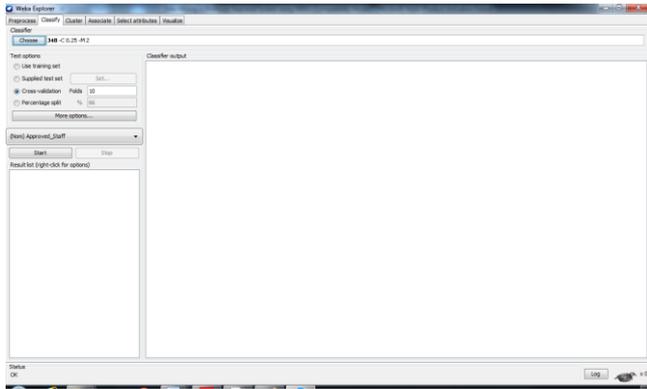


Fig.4. Decision tree with J48 tree in WEKA.

The other information alongside J48 indicates the parameters that have been chosen for the program. This paper will ignore these.

e. Choosing the experimental procedures

The panel headed ‘Test options’ allows the user to choose the experimental procedure. This paper shall have more to say about this later in the course. For the present exercise click on ‘Use training set’. (This will simply build a tree using all the examples in the data set). The small panel half way down the left hand side indicates which attribute will be used as the classification attribute. It will currently be set to ‘Approved_Staff’. (Note that this is what actually determines the classification attribute – the ‘class’ attribute on the pre-process screen is simply to allow you to see how a variable appears to depend on the values of other attributes).

f. Running the decision tree program

Now, simply click the start button and the program will run. The results will appear in the scrollable panel on the right of the Explorer window. Normally these will be of great interest but for present purposes all this paper needs to notice is that the resulting tree classified all 14 training examples correctly. The tree constructed is presented in indented format, a common method for large trees:

```

J48 pruned tree
-----

Teaching_Experience = TRUE
|  Professor_Degree = BE: No (3.0)
|  Professor_Degree = ME: Yes (21.0/8.0)
|  Professor_Degree = PHD: Yes (0.0)
Teaching_Experience = FALSE: No (10.0)

Number of Leaves :    4
Size of the tree :    6
    
```

Fig.5. Decision tree with J48 tree in WEKA.

The panel on the lower left headed ‘Result list (right-click for options)’ provides access to more information about the results. Right clicking will produce a menu from which ‘Visualize Tree’ can be selected. This will display the decision tree in a more attractive format:

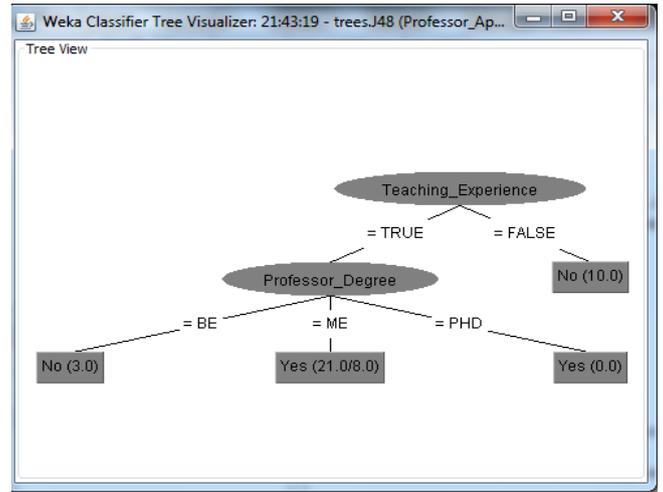


Fig.6. Visualize tree with J48 tree in WEKA.

Note that this form of display is really only suitable for small trees. Comparing the two forms should make it clear how the indented format works.

IV. CONCLUSION

At Professor’s appraisal, teaching score of students is very important factor that many colleges/universities gather this information on performance of professor. New rules by using data mining and J48 tree as a decision tree in this paper are results that heads of institutes could use these rules in future decisions to submit new professors and continue with elected old professors. For example is discovered these rules as you see in Fig.6:

- 1- IF(Teaching=GOOD)THEN (Approved_Staff is Yes means next semester who continue his/her teaching)
- 2- IF(Teaching=Excellent)AND (Teaching_experience=FALSE means is low)THEN(Approved_Staff is Yes means next semester who continue his/her teaching)

And etc.

Correctness of this rules depending variety of datasets and statistical instances can vary. But data mining tools such as WEKA as is showed in this paper can conclude variety results that help head of the departments in Engineering Colleges. These results will be used by HOD’s in decision-making.

V. REFERENCES

1. Jiawei Han. and Micheline Kamber, Jian Pei "Data Mining: Concepts and Techniques", 3rd edition.
2. Sunita B Aher, Mr. LOBO L.M.R.J. Data Mining in Educational System using WEKA, *IJCA*,2011
3. A. P. Sanjeev ve J. M, Zytkow. “Discovering Enrollment Knowledge in University Databases,” 1th Conference on KDD (Montreal.20-21 August 1995).
4. M.Vranić, D. Pintar, Z.Skoćir, “The Use of Data Mining in Education Environment,” ConTEL 2007 (Zagreb 13-15 June 2007), 243.
5. C. Shearer, “The CRISP-DM model: The new blueprint for data mining” *Journal of Data Warehousing*, (2000). 5: 13-22.
6. International Educational Data Mining, available at <http://www.educationaldatamining.org/>

AUTHORS PROFILE



T. Bhaskar received a M.Tech(CSE) from JNTU Hyderabad. He is currently working as Asst. Professor in Computer Engg, Sanjivani College of Engineering, Kopargaon, Maharashtra India. His research interest includes data mining, network security.



G. Ramakrishna is currently working as Asst. Professor in IT Dept, Sanjivani College of Engineering, Kopargaon, Maharashtra India. His research interest includes data mining, Bigdata analytics.