An Evolutionary Based Data Mining technique in Engineering Faculty Evaluation using weka

Lingaraj.K
Dept of CSE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Ranjana.V
Dept of ISE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Pruthvi.H.M
Dept of ISE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Abstract: Data mining, the extraction of hidden knowledge from large amounts of data repositories. Data mining is used in a vast area and numerous commercial data mining applications including retail sales, e-commerce, remote sensing, bioinformatics etc. Education is an essential element for the progress of country. Mining in educational environment is called Educational Data Mining. Educational data mining is concerned with developing new methods to discover knowledge from educational database. Educational data mining is concerned with developing new methods to discover knowledge from educational database. The main goal of this paper is gathering manageable experiences with data mining and also using of these experiences at E learning system and traditional education according to teacher evaluation. In this paper are verified hidden patterns of teacher evaluation by students and is predicted that which teachers will be invited to faculty classes and which teachers will be refusing and education managers due to evaluation reasons will cut the education contract with these teachers in next semesters? And what's effect of some items for examples Evaluation's score, Teacher's degree, Degree's type, Teaching experience, Acceptation to next semesters on teacher's evaluation?

Keywords: Data mining, WEKA , Classification, Clustering, Association rule, Data mining, Web mining.

1. INTRODUCTION

Data mining has attracted a great deal of attention 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 imminent 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].

Data Mining is a non-trivial process of identifying valid, novel, useful and ultimately understandable patterns in data. Alternative names for data mining are Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archaeology, data dredging, information harvesting, business intelligence, etc. Data mining can be used in various applications [12]:

Banking: loan/credit card approval, predict good customers based on old customers, view the debt and revenue changes by month, by region, by sector, and by other factors , access statistical information such as maximum, minimum, total, average, trend, etc.

Telecommunication industry: identify potentially fraudulent users and their atypical usage patterns, detect attempts to gain fraudulent entry to customer accounts, discover unusual patterns which may need special attention, find usage patterns for a set of communication services by customer group, by month, etc., promote the sales of specific services, improve the availability of particular services in a region.

Retail Industry: Identify customer buying behaviors, discover customer shopping patterns and trends, improve the quality of customer service, achieve better customer retention and satisfaction, enhance goods consumption ratios, design more effective goods transportation and distribution policies

DNA analysis: compare the frequently occurring patterns of each class (e.g., diseased and healthy), identify gene sequence patterns that play roles in various diseases

Now a day, large quantities of data is being accumulated. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Data could be large in two senses: in terms of size & in terms of dimensionality. Also there is a huge gap from the stored data to the knowledge that could be construed from the data. 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. Data Mining could help in a more in-depth knowledge about the data [2].

2. METHODOLOGY

Data mining is relatively a new technique to the world of information sciences. Successful implementation of this technique requires a sound methodology built on best practices. In this research study, has followed a popular data mining methodology called Cross Industry Standard Process for Data Mining (CRISP-DM), which is a six-step process [8]:

Problem description: Involves understanding project goals with business perspective, transforming this information into data mining problem description and

making project plan to reach the related goals .

Understanding the data: Involves identifying the sources of data, obtaining an initial set of data to assess the information coverage of the data for the problem on hand.

Preparing the data: Involves pre-processing, cleaning, and transforming the relevant data into a form that can be used by data mining algorithms.

Creating the models: Involves developing a wide range of models using comparable analytical techniques (i.e., selecting the appropriate modelling technique and setting the parameters related to the model to optimal values).

Evaluating the models: Involves evaluating and assessing the validity and the utility of the models against each other and against the goals of the study.

Using the model: Involves in such activities as deploying the models for use in decision making processes (i.e., making it a part of the decision support system/process).

A graphical representation of the methodology used in this study is shown in Figure 1.
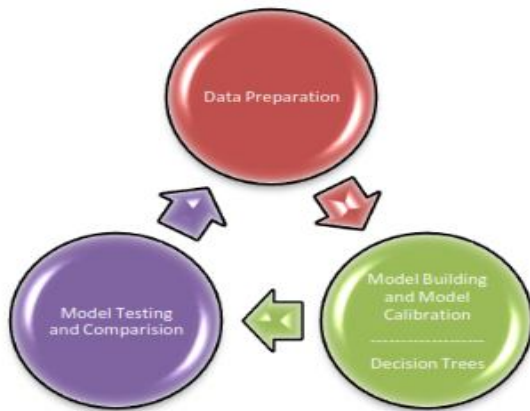

Fig.1.Agraphical illustration of the methodology employed in this study

2.1 DATA
In this study 104 records were used which is taken Sanandaj Daughter Vocational Faculty. Dataset have teacher' information such as Evaluation's score, Teacher's degree, Degree's type, Teaching experience, Acceptation.

TABLE 1: The List of Independent Variable used in the Study

Variable Name	Data Type	Description
Evaluation Score	Text	Evaluation's Score
Teacher's Degree	Text	Teacher's Score
Degree type	Text	Teacher's Degree type

Teaching Experience	Text	Teaching Experience of a Teacher
Acceptation	Text	Acceptation for a Teacher

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

Variable Name	Data Type	Values
Evaluation Score	Nominal	{ Weak,Good,Excellent }
Teacher's Degree	Nominal	{B.E,M.TECH,PHD}
Degree type	Nominal	{Full-Time, Part-Time}
Teaching Experience	Nominal	{ True, False}
Acceptation	Nominal	{ Yes, No }

Evaluation score of teachers which are studying in Sanandaj Daughter Vocational 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	Weak
60<=Score<75	Good
75<=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

2.2 Background
In this research are used WEKA and Data mining (Classification, clustering & Association Algorithm).The following subsections includes a brief overview of these topics

2.2.1 Weka

WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code [10]. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and Predictive modelling,

together with graphical user interfaces for easy access to this functionality [11].

It is freely available software. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any platform. WEKA has several standard data mining tasks, data pre-processing, clustering, classification, association, visualization, and feature selection.

2.2.2 Data Mining

Data mining is the process of discovering interesting knowledge from large amount of data stored in database, data warehouse or other information repositories. It includes various tasks such as classification, clustering, association rule etc.

2.2.3 Association Rule

Association rules are used to show the relationship between data items. Mining association rules allows finding rules of the form: If antecedent then (likely) consequent where antecedent and consequent are item sets which are sets of one or more 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 [11]. Support & confidence are the normal method used to measure the quality of association rule. Support for the association rule X->Y is the percentage of transaction in the database that contains XUY [12]. Confidence for the association rule is X->Y is the ratio of the number of transaction that contains XUY to the number of transaction that contain X [7]. Association rule can be used in educational data mining and teacher's evaluation system for analyzing the learning data.

2.2.4 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:

Model construction: It consists of set of predetermined classes. Each tuple /sample is assumed to belong to a predefined class. The set of tuple used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae.

Model usage: This model is used for classifying future or unknown 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 [11].

2.2.5 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 teacher's evaluation system, clustering has been used to group the teachers according to their behavior e.g. clustering can be used to distinguish active teacher from non-active teacher according to their performance in activities.

3. ARCHTECTURE OF PROPOSED SYSTEM

In this paper, it is done a web base survey from 3000 students then it is prepared results of this survey for 201 teachers

3.1 The Explorer Interface of Weka

In WEKA application issue, this is probably the most confusing part of becoming familiar with WEKA because you are presented with quite a complex screen.

Initially "pre-process" 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-3.6 folder called "data". This contains a selection of data files in ARFF format.

3.2 ARFF Format

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 Engineering_Faculty

@ATTRIBUTE Evaluation_Score
     {Weak,Good,Excellent}

@ATTRIBUTE Teacher_Degree
     {BE,MTECH,PHD}

@ATTRIBUTE Degree_Type      {FT,PT}

@ATTRIBUTE Teaching_Experience
     {True,False}

@ATTRIBUTE Acceptation
     {Yes,No}

@DATA

Weak,BE,PT,False,No

Weak,BE,PT,True,No

Good,BE,PT,False,No

Good,BE,PT,True,No

Excellent,BE,PT,False,No

Excellent,BE,PT,True,No

Weak,BE,FT,False,No
```

```
Weak,BE,FT,True,No

Good,BE,FT,False,Yes

Good,BE,FT,True,Yes

Excellent,BE,FT,False,Yes

Excellent,BE,FT,True,Yes

Weak,MTECH,PT,False,No

Weak,MTECH,PT,True,No

Good,MTECH,PT,False,Yes

Good,MTECH,PT,True,Yes

Excellent,MTECH,PT,False,Yes

Excellent,MTECH,PT,True,Yes

Weak,MTECH,FT,False,No

Weak,MTECH,FT,True,No

Good,MTECH,FT,False,Yes

Good,MTECH,FT,True,Yes

Excellent,MTECH,FT,False,Yes

Excellent,MTECH,FT,True,Yes

Weak,PHD,PT,False,Yes

Weak,PHD,PT,True,Yes

Good,PHD,PT,False,Yes

Good,PHD,PT,True,Yes

Excellent,PHD,PT,False,Yes

Excellent,PHD,PT,True,Yes

Weak,PHD,FT,False,Yes

Weak,PHD,FT,True,Yes
```

```
Good,PHD,FT,False,Yes

Good,PHD,FT,True,Yes

Excellent,PHD,FT,False,Yes

Excellent,PHD,FT,True,Yes
```

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 Weak. The @attribute lines declare the attributes of the examples in the data set (Note that this will include the classification attribute). 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.

3.3 This Opening a DataSet

In the Explorer window, click on "Open file" and then use the browser to navigate to the 'data' folder within the WEKA-3.6 folder. Select the file called Teacher_evaluation.arff. (This is in fact the file listed above).

This is a 'Teacher evaluation' data set, like the ones used in class for demonstration purposes. In this case, the normal usage is to learn to predict the 'Acceptation' attribute from four others providing information about the   Teacher evaluation.
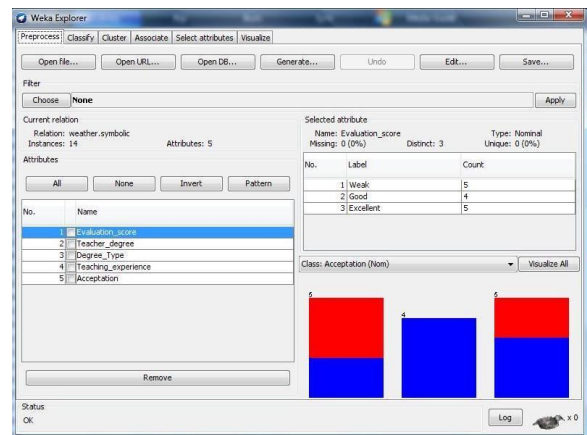

Fig3. Opening 'Teacher evaluation' Data Set in Weka

Most of the information it displays is self-explanatory: it is

a data set containing 14 examples (instances) each of which has 5 attributes. The 'Acceptation' 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 ('Evaluation Score').

This shows that it has 3 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 'Evaluation Score'.

If you click on 'Teacher Degree' in the panel on the left, the information about the 'Evaluation Score' attribute will be replaced by the corresponding information about the Teacher Degree attribute.

3.4 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.
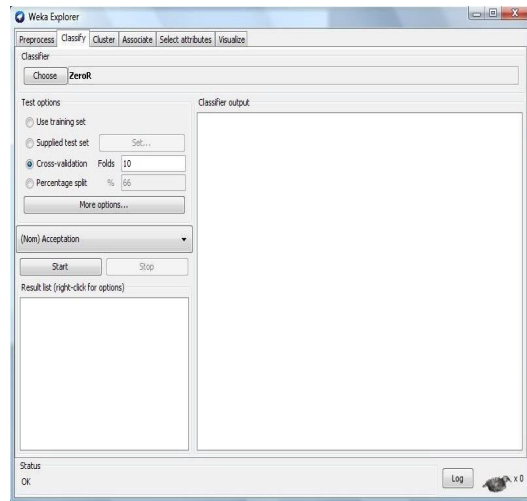
The window should now look like this:

Fig4. Selecting Classifying algorithm in Weka for case study

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.

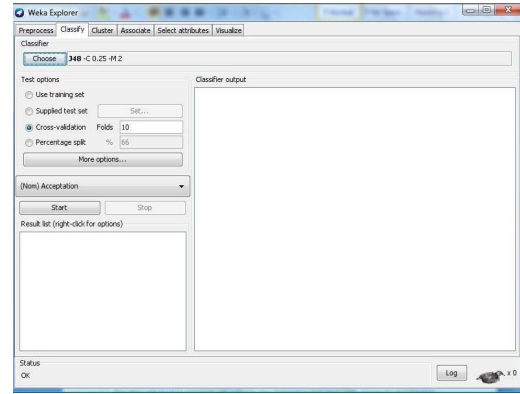The Explorer window now looks like this indicating that J48 has been chosen

Fig5. Decision Tree with J48 tree with Weka.

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

3.5 Choosing the Experimental Procedure
The panel headed 'Test options' allows the user to choose the experimental procedure. This paper shall has 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 'Acceptation'. (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)

3.6 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 large for trees:

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:

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.

```
J48 pruned tree
-----------------

Teacher_Degree = BE
|   Degree_Type = FT
|   |   Evaluation_Score = Weak: No (2.0)
|   |   Evaluation_Score = Good: Yes (2.0)
|   |   Evaluation_Score = Excellent: Yes (2.0)
|   Degree_Type = PT: No (6.0)
Teacher_Degree = MTECH
|   Evaluation_Score = Weak: No (4.0)
|   Evaluation_Score = Good: Yes (4.0)
|   Evaluation_Score = Excellent: Yes (4.0)
Teacher_Degree = PHD: Yes (12.0)

Number of Leaves  :     8

Size of the tree :     12


Time taken to build model: 0.02 seconds
```

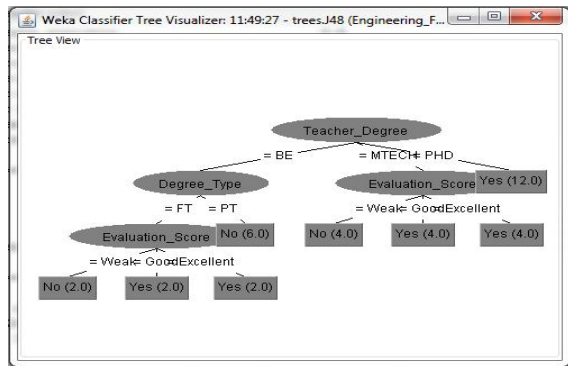**Fig6. Decision Tree with J48 tree with Weka.**

Fig7. Visualize Tree with J48 tree with 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 'Cost/Benefit Analysis' can be selected. This will display the decision tree in a more attractive format
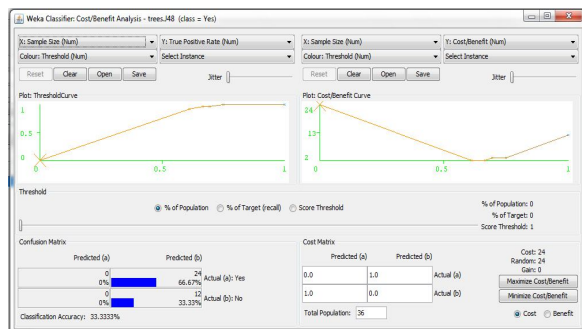

Fig8. Cost/Benefit Analysis with J48 tree with Weka

4. CONCLUSION

At teacher's evaluation, evaluation's score of students is very important factor that many universities gather this information on performance of teachers. New rules by using data mining and J48 tree as a decision tree in this paper are results that education managers could use these rules in future decisions to submit new teachers and continue with elected old teachers. 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 education managers in universities. These results will be used by managers in decision-making.

5. REFERENCES

[1] Han,J. and Kamber, M., "Data Mining: Concepts andTechniques", 2nd 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), 246.

[4] K. Becker, C. Ghedini ve E.L. Terra, "Using KDD to analyse the impact of curriculum revisions in a Brazilian university," SPIE 14th Annual International Conference (Orlando. April 2000), 412.

[5] J. Luan, "Data Mining, Knowledge Management in Higher Education, Potential Applications", 42nd Associate of Institutional Research International Conference (Toronto,Canada: 2002), 1.

[6] Ş.Erdoğan, M. Timor, "A Data Mining Application in a Student Database," Havacılık ve Uzay Dergisi. Cilt No 2,Sayı 2: 57-64, (July 2005), 57.

[7] M.Vranić, D. Pintar, Z.Skoćır, "The Use of Data Mining in Education Environment," ConTEL 2007 (Zagrep 13-15 June 2007), 243.

[8] C. Shearer, "The CRISP-DM model: The new blueprint for data mining" Journal of Data Warehousing, (2000). 5: 13-22.

[9] "Data Mining Introductory and Advanced Topics" by Margaret H. Dunham

[10] WEKA (2007). http://www.cs.waikato.ac.nz/ml/WEKA/.

[11] International Educational Data Mining, available at http://www.educationaldatamining.org/

[12] Sunita B Aher and Lobo L.M.R.J.. Data Mining in Educational System using WEKA. IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT) (3):20-25, 2011. Published by Foundation of Computer Science, New York, USA (ISBN: 978-93-80864-71-13)

[13] Fateh Ahmadi and M.E.Shiri Ahmad abadi "Data mining by teacher evaluation using Weka" IJCA ISSN:-0975-8887,page.No-14 to 18 Volume 63-No-10, Feb 2013