# Data Mining in Teacher Evaluation System using WEKA

Fateh Ahmadi
Faculty of International Gheshm, Payam Noor University,
Gheshm, Iran.

M.E Shiri Ahmad abadi
Faculty of Mathematics and Computer, Amirkabir University of Technology,
Tehran, Iran.

## ABSTRACT

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential used in various commercial 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. In order to analyze opinion of students about their teachers  in Teacher Evaluation system, this paper surveys an application of data mining in Teacher Evaluation system & also present result analysis using WEKA tool. The large amount of data is stored in educational database, so in order to get required data & to find the hidden relationship; different data mining techniques are developed & used. 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 is analyze the performance of final Teacher Evaluation of a semester of a college & is presented the result which it is achieved using WEKA tool. 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?

## General Terms

Data mining, WEKA.

## Keywords

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

## 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].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].

One of the first studies on data mining applied in education was published in 1995 by Sanjeev and Zytkow. Researchers gathered the knowledge discovery as terms like "P pattern for data in the range R" from university database [3].Another study on data mining applied in education was published in 2000 by Becker and his friends who are performed for defining and understanding the impact of changes in curriculum on students at a university in Brasil[4]. A data mining application in which defining of student characteristics are used for measuring the satisfaction of students at higher education was performed by Luan in 2002 [5].Maltepe University students identifying characteristics had been clustered using K-means algorithm in 2005 by Erdoğan and Timor. In that study 722 students' data was used and the relationship between the university entrance exam results and achievements was examined [6].Vranić and Skoćır 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 [7].In  this paper is collected information and results  of  a survey about 104  teachers in Sanandaj Daughter Vocational Faculty  on  teacher's behaviors in classroom then with data mining  algorithms such  Association Rule and decision trees (j48) , it is  proceeded to analyze and predict acceptation of a teacher for continuing   the teaching   in faculty .There are  new rules and relations between selected parameters such as Evaluation's score, Teacher's degree, Degree's Type, Teaching experience, Acceptation to next semesters on teacher's evaluation  that is interested for education managers.

## 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 modeling 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.
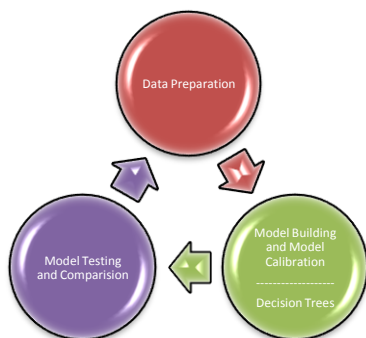


**Fig.1.Agraphical illustration of the methodology employed in this study**

## 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 variables used in this study**

| Variable Name | Data Type | Description |
|---|---|---|
| Evaluation score | Text | Evaluation's score |
| Teacher degree | Text | Teacher's degree |
| Degree type | Text | Teacher Degree's type |
| Teaching experience | Text | Teaching experience of 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 degree | Nominal | {Ba,MA,PHD} |
| Degree type | Nominal | {vocational,cultural} |
| 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 modeling, 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 preprocessing, 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 a*nd *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:

1. 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.

2. 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 830 students then it is prepared results of this survey for 104 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 "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-3.6 folder called "data". This contains a selection of data files in ARFF format.

## 3.2 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 Teacher_evalution.symbolic
@attribute Evaluation_score {Weak,Good,Excellent}
@attribute Teacher_degree {Ba,MA,PHD}
@attribute Degree_Type {vocational,cultural}
@attribute Teaching_experience {TRUE, FALSE}
```

```
@attribute Acceptation {yes, no}
@data
Weak,Ba,vocational,FALSE,no
Weak,MA,cultural,TRUE,no
Good,PHD,vocational,FALSE,yes
Excellent,MA,vocational,FALSE,yes
Excellent,Ba,cultural,FALSE,yes
Excellent,Ba,cultural,TRUE,no
Good,Ba,cultural,TRUE,yes
Weak,MA,vocational,FALSE,no
Weak,Ba,cultural,FALSE,yes
Excellent,MA,cultural,FALSE,yes
Weak,MA,cultural,TRUE,yes
Good,MA,vocational,TRUE,yes
Good,PHD,cultural,FALSE,yes
Excellent,MA,vocational,TRUE,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 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 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-3.6 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.
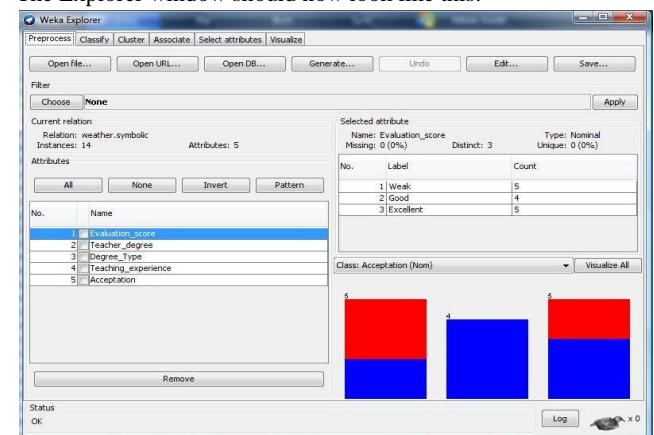
The Explorer window should now look like this:



**Fig.3. 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.

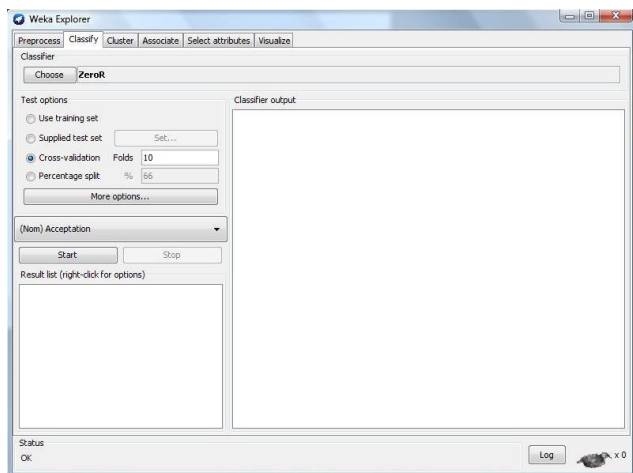The window should now look like this:



**Fig.4. Selecting classify algorithm in WEKA for this 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.

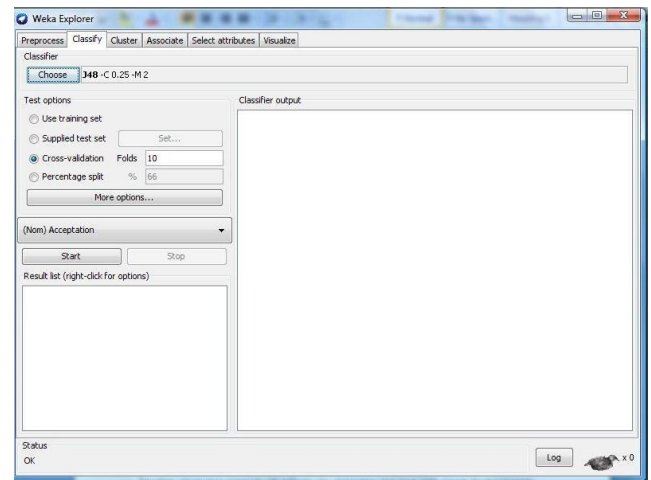The Explorer window now looks like this indicating that J48 has been chosen.



**Fig.5. 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.

## 3.5   Choosing the experimental procedures

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 for large trees:

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

Evaluation_score = Weak
|   Degree_Type = vocational: no (2.0)
|   Degree_Type = cultural: yes (3.0/1.0)
Evaluation_score = Good: yes (4.0)
Evaluation_score = Excellent
|   Teaching_experience = TRUE: no (2.0)
|   Teaching_experience = FALSE: yes (3.0)
Number of Leaves  :            5
Size of the tree :     8
```

**Fig.6. 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:
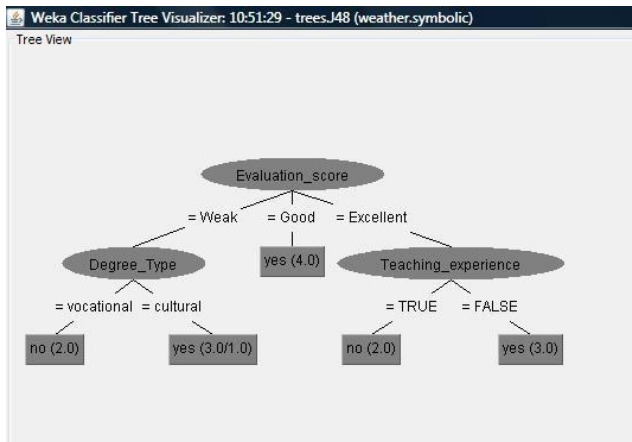


**Fig.7. Visualize tree with J48 tree in WEKA.**

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.

## 4. Conclusions

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. For example is discovered these rules as you see in Fig.6:

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

## AUTHOR'S PROFILE

**Fateh Ahmadi** received the B.E degree in Computer Science & Engineering in 2000 from Azad University, Tehran, Iran and pursuing the M. E. degree in Computer Science in Gheshm International faculty, Gheshm, Iran. He is doing the dissertation work under the guidance of **Dr M.E Shiri Ahmad abadi**, Associate Professor, Department of Computer science, Faculty of Mathematics and Computer, Amirkabir University of Technology Tehran, Iran.

**Dr M.E Shiri Ahmad abadi,** Associate Professor , Department of Computer science, Faculty of Mathematics and Computer, Amirkabir University of Technology Tehran, Iran. He has registered for PhD in Computer Science and Engineering , Amirkabir University of Technology Tehran University , Iran. He is presently working as an Associate Professor , Department of Computer science ,Faculty of Mathematics and Computer, Amirkabir University of Technology Tehran. His research interests include Evolutionary Computation, IA ,Genetic Algorithms and Data Mining.