

Data Mining for Staff Recruitment in Education System using WEKA

***Mr. Ashish P. Joshi**, Assistant Prof., BCA Department, V.P. & R.P.T.P. Science College, Vallabh Vidya Nagar, Gujarat, India joshiashish_mca@rediff.com

****Mr. Ronak K. Panchal**, Assistant Prof., BCA Department, Vidyabharati Trust College of BBA and BCA, Umraikh, Bardoli, Gujarat, India, ronuresearch@gmail.com

ABSTRACT: Education system is the essential component of the country development. It plays a major role to development of country. Thus the teaching staff who can develop the youth, they must be the qualified, nationalizes and dedicated. This analysis is based on the recruitment of the teaching staff. In order to analyze the rules given for assistant professor and associate professor by University Grant Commission, this paper generates an application of data mining for staff recruitment in education system. This application generates the result analysis using WEKA tool based on the candidate's different parameters. Different data mining techniques are developed & used to digging out required data & to find the hidden relationship from the large amount of data stored in database. In this paper how data mining task can be applied, analyze and explained from the varieties of popular task like clustering, association, classification etc. is discussed. The objective of this paper is to identify whether the candidate is applicable for the post of assistant professor or associate professor on the basis of some parameter like candidate's degree, qualification, experience, API score.

Keywords: Data mining, WEKA, cluster, classify, association, regression, J48, decision tree

1. INTRODUCTION OF DATA MINING

Digging out some meaningful data from the large amount of data it is called data mining.

We can also say that extracting data or extracting hidden pattern about data is called data mining. It is great potential and powerful technology used in various commercial applications including games, business, medical, retail sales, e-commerce, Web and Semantic Web, remote sensing, financial, Eco System Disturbances, bioinformatics, etc.

Processed data becomes information, processed information becomes knowledge. In the case of data mining techniques the information is already available that is stored in a proper manner in the data warehouse, its need some process or apply some anticipated pattern or algorithm that makes it knowledge.

Data mining involves six common classes of tasks:^[2]

- *Anomaly detection* – The identification of unusual data records, that might be interesting or data errors that require further investigation.
- *Association rule learning* – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

- *Clustering* – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- *Classification* – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".
- *Regression* – attempts to check a function which models the data with the optimum (least) error.
- *Summarization* – providing a more compact representation of the data set, including visualization and report generation.

2. INTRODUCTION OF WEKA

WEKA (Waikato Environment for Knowledge Analysis) is a popular suite of [machine learning](#) software written in [Java](#), developed at the [University of Waikato, New Zealand](#).^[3]

The vital features behind the success of WEKA:

- It provides many different algorithms for data mining and machine learning
- It is open source and freely available
- It is platform-independent
- It is easily useable by people who are not data mining specialists
- It provides flexible facilities for scripting experiments
- It has kept up-to-date, with new algorithms being added as they appear in the research literature.

The techniques which are provided by the WEKA:

- Data preprocessing and visualization

- Attribute selection
- Classification (OneR, Decision trees)
- Prediction (Nearest neighbor)
- Model evaluation
- Clustering (K-means, Cobweb)
- Association rules

3. METHODOLOGY

Data mining is technique that use in variety of fields. It may give successful result, if the proper methodology is applied. In this research study, below methodology is applied which is five step process.

3.1 Problem analysis

It identifies the existence of problem. Decide objective of project with knowledge perspective.

3.2 Collecting and Understanding the data

There is need of collection of data to solve out the problem. It may be available or it may be created. First find out the source of data and If data is available then understand. If data is not available then create it.

3.3 Preparing the data

This includes the preparing the data for model. It includes pre-processing or extracts important data. Create proper format file of the data like in weka .arff file format is preferable.

3.4 Design the models

Select the type of model from available wide range of model for the comparison perspective.

3.5 Analyze result and implementation

Check whether obtained result is as per anticipation or not. If the obtained result is as per our expectation, apply the model in the tool for use in decision making process.

4. APPLICATION AND EXPERIMENT

The following steps describe how to apply data into the WEKA and get desired result.

4.1 create Data File

The first step is required to collect the data and prepared it as per specific format. The most preferable format for the weks is .arff (Attribute Relation File Format) format which can be written like below.

```

staff recruitment - Notepad
File Edit Format View Help
@relation Staff_Selection.symbolic

@attribute api_score{good,best}
@attribute experience {yes,no}
@attribute degree {master,doctorate}
@attribute designation {adhyapak,assistant,associate}

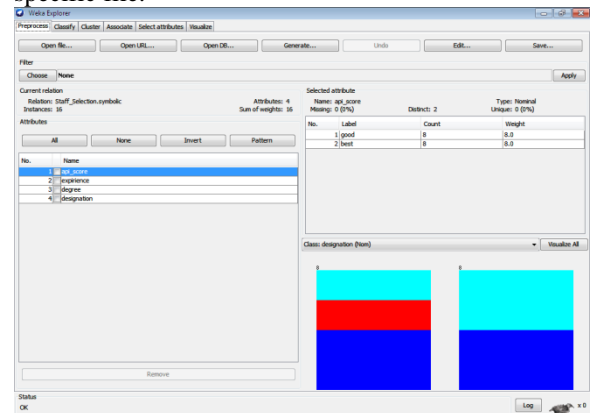
@data
good,no,doctorate,assistant
good,no,doctorate,assistant
good,no,master,adhyapak
good,no,master,adhyapak
good,yes,doctorate,associate
good,yes,doctorate,associate
good,yes,master,adhyapak
good,yes,master,adhyapak

best,no,doctorate,associate
best,no,doctorate,associate
best,no,master,adhyapak
best,no,master,adhyapak
best,yes,doctorate,associate
best,yes,doctorate,associate
best,yes,master,adhyapak
best,yes,master,adhyapak

```

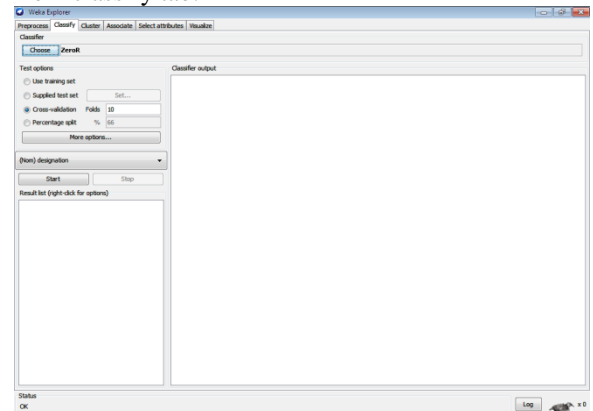
4.2 Open file in weka

We can open file directly from double clicking on it. Also we can open weka, select explorer, we can open file from open file button and then select the specific file.



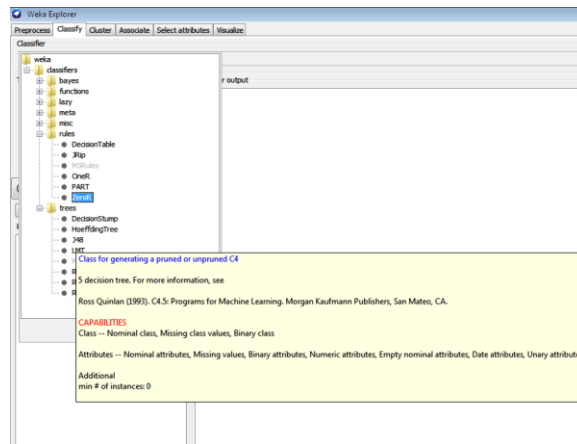
4.3 Select Classify tab

Six tabs available for the various pattern in weka like preprocess, classify, cluster, associate, select attributes, visualize. We can classify the information from classify tab.



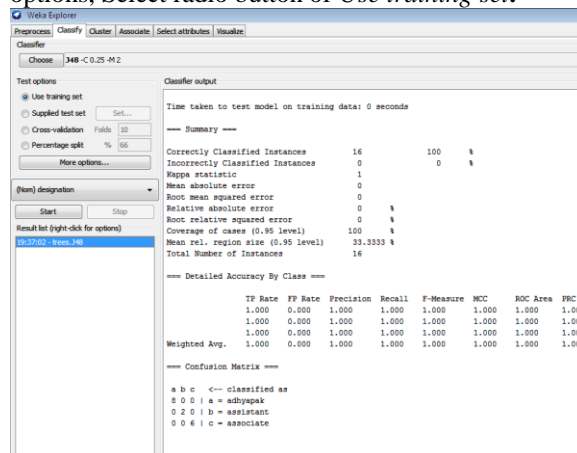
4.4 select j48 option

Select *tree* from drop down list and select j48 option from choose button. It is use for decision tree.



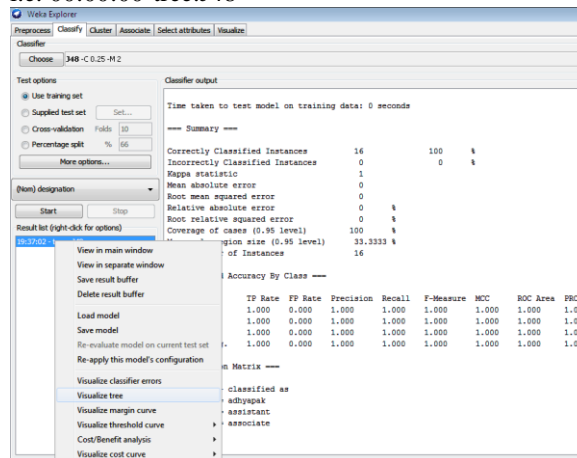
4.5 Test Options

There are many different options available in test options, Select radio button of *Use training set*.



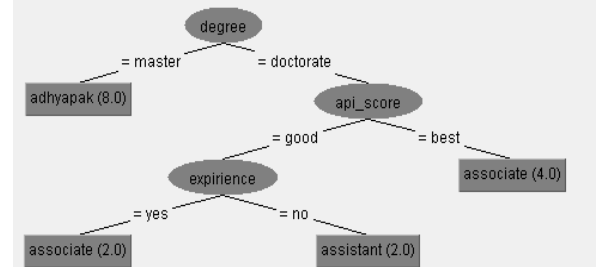
4.6 Start

The above code will automatically generate. It will generate it into the result list by pressing start button. i.e. 00:00:00-tree.J48



4.7 Generate Decision Tree

Right click on result list and select option *visualize tree*.



Thus above seven step procedure will generate the decision tree as per our requirement. It is analysis of data, design of model and evaluates the result.

5. PROCESS AND CONCLUSION

In this experiment, evaluation's score of teacher is very vital factor. We developed one weka (.arff) file. Load it into weka then we select specific pattern also select some related options as per early discussed and generate the result.

It is about the system of recruitment in education system. The designation and promotion is decided based on the API (*Academic Performance Indicators*) score of the instructor. The teacher can increase his/her API with many of research activities and also academic activities.

The below table describes the API point for different activities which is decided by UGC. Teachers can improve his/her API from participating in the below different activities.

No	Activity	Point For Each
1	Research Papers published	
	Refereed Journals	15
	Non-refereed but recognized having ISBN/ISSN numbers.	10
2	Research Publications (books, chapters in books, other than refereed journal articles)	50/25/1 5/10
3	Sponsored Projects carried out/ ongoing	
	Major Projects amount mobilized with grants above 30.0 lakhs	20
	Major Projects amount mobilized with grants above 5.0 lakhs up to 30.00 lakhs	15

	Minor Projects (Amount mobilized with grants above Rs. 50,000 up to Rs. 5 lakh)	10
4	RESEARCH GUIDANCE	
	M.Phil.	3
	Ph.D	10
5	TRAINING COURSES AND CONFERENCE /SEMINAR/WORKSHOP PAPERS	
	Not less than two weeks Duration	20
	One week duration	10
6	Papers in Conferences/ Seminars/ workshops etc.	
	International conference	10
	National	7.5
	Regional/State level	5

The staff recruitment is based on two important thing that is degree and score (API score). So, calculation of API score must be done by follow above table's condition. You can also check the API from the [www.iipsindia.org/pdf/Academic_Performance_Indicator\(API\).pdf](http://www.iipsindia.org/pdf/Academic_Performance_Indicator(API).pdf)

Then comparing the API score with his/her degree, the director may recruit the staff or give promotion to the staff.

New rules by using data mining and J48 tree as a decision tree in this paper are results that education directors could use these rules in future decisions to decide qualification for recruitment of new teachers. For example is discovered these rules as you see in below figure:

If
[degree=doctorate] and [score=best]
Designation: Associate professor
Else if
[degree=doctorate] and [score=good] and [exp=yes]
Designation: Associate professor
Else if
[degree=doctorate] and [score=good] and [exp=no]
Designation: Assistant professor
Else
Designation: Adhyapak Sahayak

Efficiency of this rules depending variety of datasets and statistical examples can vary. But data mining tools such as WEKA as is showed in this paper can conclude range results that help education directors in universities. These results will be used by directors in decision-making.

REFERENCES

- [1] Margaret H. Dunham, "Data Mining Introductory and Advanced Topics
- [2] <http://www.cs.waikato.ac.nz/ml/WEKA/> - WEKA (2007)
- [3] http://en.wikipedia.org/wiki/Data_minig
- [4] <http://en.wikipedia.org/wiki/Weka>
- [5] Sunita B Aher, Mr. LOBO L.M.R.J. Data Mining in Educational System using WEKA, IJCA ,2011
- [6] Heikki, Mannila, "Data mining: machine learning, statistics, and databases", Statistics and Scientific Data Management.
- [7] [www.iipsindia.org/pdf/Academic_Performance_Indicator\(API\).pdf](http://www.iipsindia.org/pdf/Academic_Performance_Indicator(API).pdf)
- [8] Introduction to Data Mining with Case Studies by Gupta G. K. (Author)
- [9] Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) by Jiawei Han (Author), Micheline Kamber (Author)
- [10] <http://www.cs.waikato.ac.nz/ml/weka/citing.html>
- [11] 11. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) by Ian H. Witten (Author), Eibe Frank

AUTHORS



Ashish P. Joshi received his MCA degree from Saurashtra University in 2006 and B.Com. degree from Saurashtra University in 2003. His research interest includes Data Mining.



Ronak K. Panchal received his M.Sc. (Information & Communication Technology) in year 2008, PGDRM (Post Graduation Diploma in Research Methodology) in year 2013 and BSc (Computer Science) in year 2006 from Veer Narmard South Gujarat University surat. His research interest includes Semantic Web & Data Mining.