

- 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

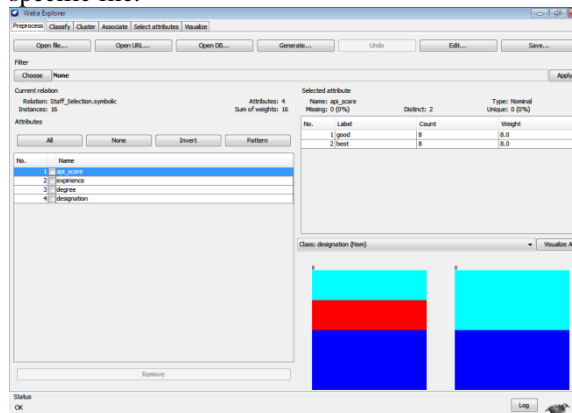
The first step is required to collect the data and prepared it as per specific format. The most preferable format for the weks is .arff (Attribute Relation File Format) format which can be written like below.

```

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
    
```

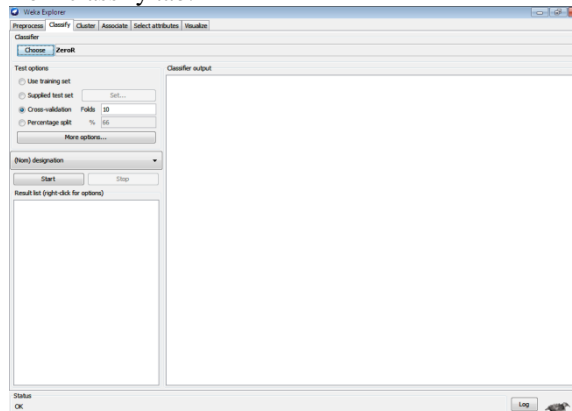
4.2 Open file in weka

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.

	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] 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.