

PREDICTING STUDENT PERFORMANCE USING RESULT MINING AND KNOWLEDGE FLOW IN WEKA

Dr. A. Kanaka Durga*

IT Department, Stanley College of Engineering and Technology for Women, Hyderabad,
India.

Article Received on 24/11/2017

Article Revised on 15/12/2017

Article Accepted on 05/01/2018

***Corresponding Author**

Dr. A. Kanaka Durga

IT Department, Stanley
College of Engineering and
Technology for Women,
Hyderabad, India.

ABSTRACT

It is natural that the quantity of data collected will continue to expand rapidly because of the increasing ease, availability and popularity of the web. Data Mining has its great application in organizations because it collects large amount of data. By applying data mining techniques people can work on the extraction of hidden, historical and previously unknown large databases. In this paper we used weka tool for the pre-processing, classification and analysis of institutional results of engineering students. Results show analysis of marks and backlogs. Knowledge flow analysis has been carried out on engineering students' results.

KEYWORDS: Classification, clustering, WEKA, data mining, Knowledge flow, engineering students.

1. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools^[1] can answer business questions that traditionally were too time consuming to resolve. They scour

databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Huge amounts of data are being accumulated at current situations. Manual way of extracting information is very difficult in the current scenario as the volume of data to be processed is very huge. Data mining tools provide a better alternative to the manual method. In this paper we are using WEKA Tool for the analysis of engineering students' data.

This paper uses Result Data Mining Techniques (RDM) to provide more accurate result analysis. Data mining is a dynamic technology^[2] to deal and extract the hidden potential data which is to be converted to useful information. It discovers information within the data that queries and reports can't effectively reveal. After gathering data from the university result, data mining technique need to be applied to determine the status of the students in various subjects.

2. RESULT DATA MINING

Predicting under performance students in educational organizations is a very challenging task if it is done manually. Data Mining plays a major role in analyzing the weak areas of students and focus on the key areas where there is a scope for poor performance. Student results in various subjects of engineering have been mined to focus the thrust areas which affect the student performance. Student result mining involves extracting information from various subjects and the gathered information is processed using classification algorithms^[3] of data mining. Tools can be developed using data mining such that performance evaluation becomes easier for the teachers.^[5]

3. METHODOLOGY

The results of engineering students are taken from the university website and preprocessed using WEKA tool.^[4] WEKA is a popular data mining tool developed by the Waikato University, Newzealand. It consists of many algorithms used for filtering, classification, clustering, regression, association analysis etc, which are useful in data mining and many other fields like Natural Language Processing and Machine Learning. The algorithms in WEKA can be applied in the GUI or they can be invoked from the programmer's code.

WEKA GUI consists of four tabs which are explorer, experimenter, knowledge flow and Simple CLI. This paper implements knowledge flow on the selected data set i.e engineering students' results and displays the experimental results.

4. RESULTS

The dataset is taken from the university website and preprocessed using the explorer tab of WEKA tool. The dataset is prepared in notepad. Any text editor can be used for this purpose. The data should be written in ARFF format and the extension of the file consisting the target data is arff. The general structure of an ARFF file is shown in Fig. 1.

```
@relation relationname
@attribute attributename type
@attribute attributename {options}
@data
----data comes here-----
```

Fig. 1: ARFF file general structure.

Any ARFF file consists of two sections: one is the header section which includes the name of the relation and the names and types of the attributes. The attribute types can be numeric, nominal, ratio etc. Numeric types can be used in this dataset(engineering students results) for the marks scored by the student. Nominal attributes take only a set of values which are specified by options in Fig. 1. The attributes result, backlog and appeared in our dataset are nominal.

Once the dataset is ready WEKA tool is used for preprocessing the data and visualized shown in Fig 1. Fig 2 shows the classification of data using Naïve Bayesian Classifier.

Knowledge flow is applied in WEKA tool and is illustrated in two cases i.e. Fig 3 and Fig 4.

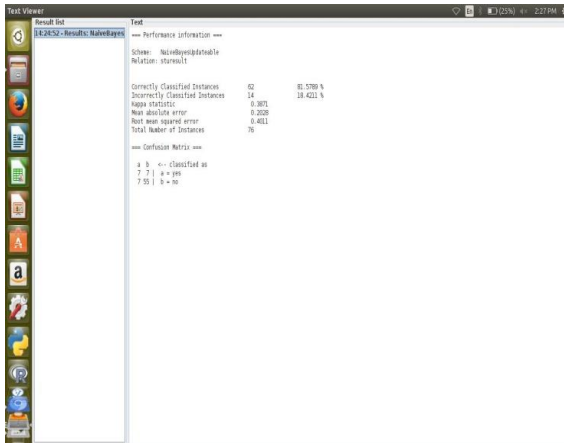


Fig 1: Data Visualization.

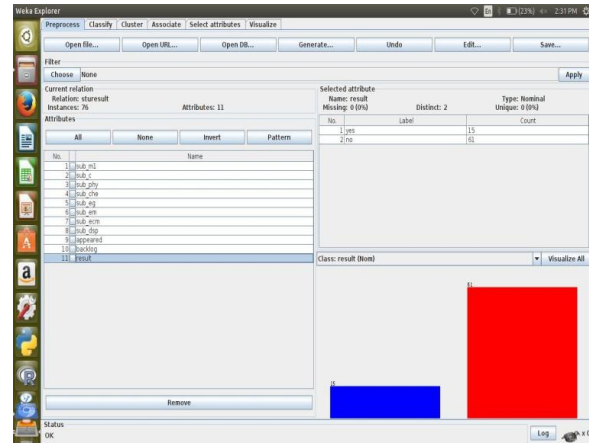


Fig 2: Data Classification.



Fig 3: Knowledge Flow.

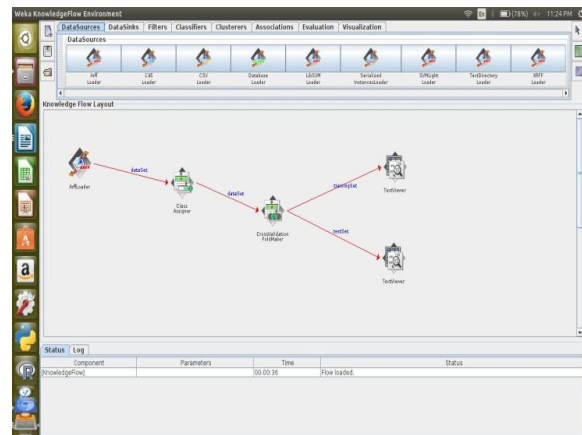


Fig 4: Knowledge Flow.

The knowledge flow in Fig. 3 Contains the following: ARFFLOADER, CLASS ASSIGNER, NAÏVE BAYES UPDATABLE CLASSIFIER, INCREMENTAL CLASSIFIER EVALUATOR, TEXT VIEWER AND STRIP CHART.

The flow of Fig. 3 is explained as follows: The dataset is loaded using the ARFFLOADER and the output of the loader is given as instance to the class assigner which in turn gives instance as input to the NAIVE BAYES UPDATABLE CLASSIFIER. The purpose of a class assigner is to assign a column to be the class for any data set, training set or test set. The incremental output of the NAIVE BAYES UPDATABLE CLASSIFIER is given as input to the INCREMENTAL CLASSIFIER EVALUATOR. The purpose of the Incremental Classifier Evaluator is to evaluate the performance of incrementally trained classifiers. The output of the classifier is given as input to the text viewer and strip chart with respective formats. The final outputs can be viewed in these components of the knowledge flow.

The flow of Fig. 4 is explained as follows: The dataset is loaded using the ARFFLOADER

and the output of the loader is given as dataset to the class assigner which in turn gives dataset as input to cross validation fold maker. The output of the cross validation fold maker is given as input to the two text viewers in one of which training set is viewed and in the other test set is used.

6. CONCLUSION

This paper includes the study of Data Mining tool applied to student result set. By using WEKA tool you can pre-process the data, classify the data for different subjects and do some result analysis the result data. The engineering student university result. We have applied data mining for analyzing the engineering students' results. The development of such analysis tools help in the applications of e-educational systems,^[6] which aid the faculty to predict performance of students.

7. REFERENCES

1. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, 2nd ed., Morgan Kaufmann publishers, San Francisco, 2006.
2. Fayyad, U., & Stolorz, P. Data mining and KDD: promise and challenges. Future generation computer systems, 1997; 13(2): 99-115.
3. Guerra L, McGarry M, Robles V, Bielza C, Larrañaga P, Yuste R. Comparison between supervised and unsupervised classifications of neuronal cell types: A case study. Developmental neurobiology, 2011; 71(1): 71-82.
4. An Introduction to weka data mining tool.
5. Romiro C. and Ventura S., "Educational data mining- A survey from 1995-2005" Expert systems with applications, 2007; 33: 135-146.
6. International Educational Data Society www.educationdatamining.org.