

## A Comparative Study on Data Mining Tools

Priti S. Patel<sup>1</sup>, Dr. S.G. Desai<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science, Shree Ramkrishna Insti. Of Computer Education and Applied Science, Surat, INDIA, priti\_patel22@yahoo.co.in

<sup>2</sup>Professor, SAL Institute of Engineering, Ahmadabad, INDIA, subhash1948@vahoo.com

### ABSTRACT

A research needs to collect data from various sources and analysis those data with some techniques for predict or decision making process. After collections of various data, main task it to maintain data apply transformation and preprocessing of large data sets for that data mining tools is required. Now a day's various tools for data mining are available either as open-source or commercial software. It includes wide range of software products, from comfortable problem-independent data mining suites, to business-centered data warehouses with integrated data mining capabilities and to early research prototypes for newly developed methods. Researchers needs these kind of tool for analysis their data. In this paper, we are discussed about various available data mining tools and compare their utilities. There are so many tools are there like WEKA, orange, Rapid Miner, Tanagra etc..

**Key words :** data analysis, data preprocessing, open source transformation, software products, etc..

### 1. INTRODUCTION

Data handling and managing of large data sets, there are various data mining tools are available which improve data quality from past to present. Data mining tools are provided for pre-processing data, feeding it into a variety of learning schemes, and analyzing the resulting classifiers and their performance. The development of Information technology has paved way to generate large amount of databases and huge data in various area[1]. The research in databases and information technology has given rise to approach to store and manipulate precious data for future decision making [2].

### 2. CATEGORIES OF DATA MINING TOOLS

There are mainly three different categories of data mining tools. Traditional data mining tools, Application based tools/Commercial based software and web-based data mining tools. Description of each is as follows:

#### 1) Traditional data mining tools

Some mining programs are work as traditional way to collect and analyze data which used by various company for decision making process of large data sets. Majority of these are supported by windows and UNIX versions. However,

some software specialized in a single operating system and sometimes handling with only one database type.

#### 2) Application based tools

An applications which shows the business oriented interface for data performance. In this historical data are represented as a references and check the current trends in order to see the changes in the business. So, application based tools are easy to use and helps in administrative work and provide services for company performance.

#### 3) Web based data mining tools

This kind of tools is called text-mining tool because of its ability to mine various kind of text from any written resources. And also help for scanning and converting data in selected format which is compatible with any tools.[2]

### 3. DATA MINING TOOLS

Some of tools which are available in market are describes as follows like: WEKA, R, Orange, Rapid Miner and Tanagra etc..

#### WEKA

WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. Weka provides 3 ways to use the software: the GUI, a Java API, and a command line interface (CLI) WEKA Explorer preprocessing, classification, clustering, association, attribute selection, and visualization tools. Weka is a collection of machine learning algorithms for data mining tasks and well suited for developing new machine learning schemes. Weka is a java based software capability of working under various operating systems. With the Java-based version, the tool is very sophisticated and used in many different applications including visualization and algorithms for data analysis and predictive modeling. Its free under the GNU General Public License, which is a big plus compared to Rapid Miner, because users can customize it however they please.[7]

#### R

R is open source programming language and environment for statistical computing and graphics. R provides a wide variety of graphics and statistical techniques such as linear and non-linear modeling, classical statistical tests, time series analysis, classification clustering and is highly

extensible. This tool is very useful for data visualize and good data extensible.

### **Orange**

Orange is an open source data mining tool and visualization software with active community and which helps novice and experts for their analysis. This tool is compatible with windows, Mac OS c and GNU/Linux operating systems. Its also very useful for analytical process which have user friendly visual programming or python scripting. This tool have all major data mining algorithms. Specially, this tools have utilities for Bioinformatics Add-On and Text Mining Add-On.

### **Rapid Miner(Yale)**

Rapid Miner Studio combines technology and applicability to serve a user-friendly integration of the latest as well as established data mining techniques. Defining analysis processes with Rapid Miner Studio is done by drag and drop of operators, setting parameters and combining operators. It provides all tasks of professional data analysis, from data partitioning, to market-based analysis, to attribute generation, it includes all the tools you need to make your data work for you. But also methods of text mining, web mining, the automatic sentiment analysis from Internet discussion forums (sentiment analysis, opinion mining) as well as the time series analysis and -prediction are available.[4]

### **Tanagra**

Tanagra is open source data analysis software for academic and research purposes which proposes several data mining methods from exploratory data analysis, statistical learning, machine learning and database area [3]. The main purpose of Tanagra is to provide platform for researchers and students to use data mining software in easy way by conforming to the present norms of the software development and allowing to analyze either real or synthetic data. The second purpose is to propose an architecture allowing the users to add to add their own data mining methods it helps to compare their performances. It acts more as an experimental platform in order to do the essential work, dispensing them to deal with the unpleasant part of the data management. Last purpose is to give the direction to novice developers in diffusing a possible methodology for building this kind of software. It can be considered as a pedagogical tool for learning programming techniques since it permits to access the source code, to look pattern of the software how it is built, the problems to avoid, key steps of the project, tools used and code libraries used for the project.

### **KNIME (Konstanz Information Miner)**

It is a user -friendly and comprehensive open-source data integration, processing, analysis, and exploration platform. From day one, KNIME has been developed using rigorous software engineering practices and is currently being used

actively by over 6,000 professionals all over the world, in both industry and academia. KNIME is a modular data exploration platform that enables the user to visually create data flows (of ten referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models [5].KNIME is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models. Written in Java and based on Eclipse, KNIME is easy to extend and to add plugins. Additional functionalities can be added on the go. Plenty of data integration modules are already included in the core version.

## **4. COMPARISION OF DATA MINING TOOLS**

All Data mining tools have their own pros and cons with respect to efficiency and accuracy. Most of the researchers are works on R,WEKA and Rapid miner for their research work. So, here in this paper mainly comparison of these tools. In which R data mining tool have characteristics of the tool with packages and also supports complex operations which implement in new programming language but sometimes it makes a disadvantage for R data mining tool. Issues of R tools are Less interactive GUI, Less specialized towards data mining, Language is pretty different from current mainstream languages like C, C#, C++, Java, PHP and VB, There is a learning curve, unless you are familiar with array languages.

Now about WEKA data mining tools which is freely open source tool and access data using preprocessing and various data mining techniques mainly clustering, classification and association rules. It is also useful to maintain those friendliness about implement the new method or technique which is totally handle by developer. Rapid Miner provides much more analysis steps (operators) than other tool and much more possibilities to combine different results or solutions. It have small modules of Rapid Miner can be combined. It is good for the Rapid Miner developers to come up with such a clear and modular concept for data analysis processes. The algorithms executes in rapid miner were also optimized for speed is very high with less memory as compare to other tool. Visualization is much better in Rapid Miner. Preprocessing ,data extraction / transformation are available directly within Rapid Miner which are important aspects of data analysis mining tool. This integrates all phases of analysis into one process / tool and my work became really smooth.

## 5. CONCLUSION & FUTURE WORK

So far we discussed the various tools which are used in marketplace now a days. But somehow the usage of specific tools are limited and for these reason here we discussed here about mainly three data mining tools only. In future we will use tools for specific any techniques and try to find out the efficiency and accuracy of result which we will examine on various tools.

## REFERENCES:

1. S. Hameetha Begum,," **Data Mining tools and trends- An overview**",International Journal of Emerging Research in Management & Technology ,ISSN: 2278-9359.
2. Mrs. bharti M. Ramgeri, " **Data Mining Techniques and application**", Indian Journal of Computer Science and Engineering,Vol. 1 No. 4,pp 301-305..
3. Y. Ramamohan, K. Vasantharao, c. Kalyana chakravarti, and A.S. K.Ratnam, " **A Study of Data Mining Tools in Knowledge Discovery Process**", International Journal of Soft computing and Engineering (IJSCE), Vol.2, Issue -3,July 2012,pp 191-194..
4. Abdullah H. Wahbeh, Qasem A. Al-Radaideh, Mohammed N. Al -Kabi, and Emad M. Al -Shawakfa,"**A Comparison Study between Data Mining Tools Over some Classification Methods**" ,(IJACSA) International Journal of Advanced Computer Science and Applications ,Special Issue on Artificial Intelligence
5. Kalpana Rangra ,Dr. K. L. Bansa, "**Comparative Study of Data Mining Tools**",International Journal of Advanced Research in Computer Science and Software Engineering
6. Rapid Miner Studio 6,User manual book.
7. Wake manual tutorial.