ANALYSIS OF DATA MINING TOOL ORANGE

V.Padmavaty¹, C.Geetha², N.Priya³

¹. Student, Dep of CSE, Bharath University, Chennai, Tamilnadu, India.
². Assistant professor, Dep of CSE, Bharath University, Chennai, Tamilnadu, India.

Abstract

Data mining is the process of sorting through large datasets to identify pattern and establish relationships to solve problems through data analysis. Data mining is the process of discovering pattern in large data sets involving methods at the intersection of machine learning, statistics, and data base systems. There are many tools such as weka, R, orange etc to analyze, visualize and extract data using data mining. All the datamining tools are not compatible to perform all analysis operations. In this paper we have attempted data mining tools for analysis and checking for better tool.

Key words: Data Mining, orange, attribute statistics, Pre-processing.

Introduction

Data analysis is an essential part of data mining and Business Intelligence (BI) and is critical to picking up the understanding that drives business choices. Data analysis is the process of performing three operations which would be major such as cleaning, transforming and data extraction. Comparing with some tools along with features and parameters are decided to use for analysis.

Data Mining Tools

Apache Mahout, Orange, Weka, R, Rapid Miner, Data Melt, Knime

Apache Mahout: Apache mahout focuses mainly on data clustering, classification, and collaborative filtering. Mahout is written in JAVA and includes JAVA libraries to perform mathematical operations like linear algebra and statistics. Mahout is growing continuously as the algorithms implemented inside Apache Mahout are continuously growing. The algorithm of mahout have implemented a level above Hadoop through mapping/reducing templates.

Orange: Orange is a perfect software suite for machine learning and data mining. It is useful for visual programming and explorative data analysis. It is written in python and has multiple components known as widgets. Data mining tool supports macOS, Window and Linux.

Weka: Weka is a machine learning data mining tool in java and best suited for data analysis and predictive modelling. It contains visualization and analysis. Weka supports major data mining tasks including processing, visualization, regression etc. It works on the assumption that data is available in the form of flat file.

R: R is also called open source and a free software environment to perform statistical computing and graphics. It is widely used in academia, research, engineering, and industrial applications.

Rapid Miner: Rapid miner is one of the best predictive analysis system developed by a company with the same name as rapid miner. It is written in java programming language. It provides an integrated environment for deep learning, text mining, machine learning and predictive analysis. It uses client/server model which is used in performing extraction, transformation and data processing operations.

Knime: knime is the best integration platform for data analytics and reporting development by KNIME.com AG. It operates on the concept of the modular data pipeline. KNIME is widely used in
pharmaceutical research. In addition, it performs excellently for customer data analysis, financial data analysis, and business intelligence.

- **Data melt:** Data melt, also known as DMelt is a computation and visualization environment that provides an iterative framework to do data analysis and visualization. Dmelt is written in java and it is a multiplatform utility and an open source data mining software.

### Comparision of data mining with parameters:

<table>
<thead>
<tr>
<th>Offset</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

(Table1 specific value to set in comparision)

In Table 1 and Table 2 shows all data. mining tool has been specified with their parameters either tool supports or ignore. If it gets positive or supporative value to offset assign true means 1 and if gets negative or not supportive value to offset assign false means 0. Offset is using for to balance one to another.

(Table2 comparison of tools with parameters)

<table>
<thead>
<tr>
<th>Features/Parameters of Data Mining tools</th>
<th>Orange</th>
<th>Weka</th>
<th>R</th>
<th>Rapid miner</th>
<th>Knime</th>
<th>Datamelt</th>
<th>Apache mahout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open source</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Data visualization and analysis</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Interaction and data analysis</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Large toolbox</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Scripting interface</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Platform independence</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Covering methods</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Parameters optimized method learning/statistical methods</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>06</td>
<td>05</td>
<td>03</td>
<td>05</td>
<td>04</td>
<td>04</td>
<td>03</td>
</tr>
</tbody>
</table>

**Methodology/Technique:** Orange data mining is a component based visual programming software package for data visualization, machine learning, data mining and data analysis. It also helps in supporting programming language like C,C++ and python that also supports data validation, comparision and prediction. Orange is easy to learn.
and is better and best than other as compare it above

Orange uses for practical Implantation

- **Data Analysis**

(Table 3) Data analysis using orange

Orange has performed practical of data analysis is a long with some csv (Comma Separated Values) to file data many tool and co-relates with sources which provide to the Data Table.

**Data pre-processing:**

Data pre-processing is a data mining technique that involves transforming raw data into understandable data. Pre-processing is a crucial for achieving better quality analysis results. The pre-process widget offers several pre-processing methods that can be combined in a single pre-processing pipeline

(Table 5) Data visualization using orange

**Result**

After analysis practically, orange generates results in numerical or statistical data. It display attribute statistical with mean and median values.
Conclusion

In this study of data analysis using data mining tool, comparing their parameters to each other and find out which tool is better to perform best analysis over time. Therefore Orange tool has performed well and easy to use. Moreover after perform practical implementation Orange has done everything as its feature said. This tool makes analysis work easier.

References


2. Kaliyamurthie, K.P., Balasubramanian, P.C. An approach to multi secure to historical malformed documents using integer ripple transfiguration Sciences 9 2.


29. Sivaraman, K., Khanna, V. "Implementation of an extension for browser to detect vulnerable elements on web pages and avoid clickjacking2016 Journal of Chemical and Pharmaceutical Sciences."