Analysis of Data Using Data Mining tool Orange

[1] **Maqsud S. Kukasvadiya,** [2] **Dr.Nidhi H. Divecha**

[1] **Research Scholar, KSV, Gandhinagar, India**

[2] **Assistant Professor, SKPIMCS, Gandhinagar, India**

Abstract: Data Mining is a process of computing models or design in large collection of data. There are many tools to analyze, visualize and extract data using data mining. However all the tools are not compatible to perform all analysis operations, In this paper we have attempted data mining tools for analysis however a tool is better than other.

Keywords: Data Mining, orange, attribute statistics, Pre-processing

I.   Introduction

Data Analysis is a process of performing three major operations cleansing, transforming and modeling data. However there are various tools of data mining to perform data visualization, data analysis and data extraction. Comparison of some tools along with parameters and features and decided to use for analysis.

Data Mining Tools:

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

- ➢ Orange : A data mining tool which is useful for visual programming and explorative data analysis. It can be written in Python. Orange has multiple components are known as widgets. This data mining tool supports macOS,Windows and Linux.

- ➢ Weka: Weka is a machine learning data mining tool written in Java,It contains visualization and analysis.Weka has comprehensive collection of preprocessing data and modeling.

- ➢ R : R is also a data mining tool and open source. R has been used in statistical computing. It implements wide variety of statistical and graphical techniques such as linear and nonlinear Data Mining tool behaves as interpreted language.

- ➢ Rapid Miner:A data mining tool, It developed on an open core model. Mainly rapid miner uses client/server model. Rapid Miner has been performed extraction, transformation and data processing operations.

➢ Knime : Knime(Konstanz Information Miner) is a open source data mining tool. Once it was using in pharmaceutical research.

➢ Data Melt : Data Melt is a framework for scientific computation and multiplatform and written in Java. It is open source data mining tool.

Comparison of all data mining tools is with parameters. Some tools get advantage and perform better while others are not well.

Offset	
True	1
False	0

(Table-1 Specific value to set in comparison)

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 supportive value to offset assign true means 1 and if it gets negative or not supportive value to offset assign false means 0.Offset is using for to balance one to another.

| Features/Parameters of Data Mining tools	Data Mining Tools					
	Orange	Weka	R	Rapid Miner	Knime	Data Melt
Open Source	1	1	1	1	1	1
Data Visualization and Analysis	1	1	0	1	0	1
Interaction and Data Analysis	1	1	0	1	1	1
Large Tool box	1	0	0	1	1	0
Scripting Interface	1	1	1	0	1	1
Platform Independence	1	1	1	1	0	0
Covering Methods	0	1	0	1	0	1
Parameters optimized Method Learning/Statistical methods	0	0	0	1	0	1
Total	06	05	03	05	04	04

(Table-2 Comparison of tools with parameters)

II.   Methodology/Techniques

The orange data mining is beneficial to analyze data. It supports programming languages like C, C++ and Python that also supports data validation, comparison and prediction. Orange is Easy to earn. Orange tool is better than other as compare it above.
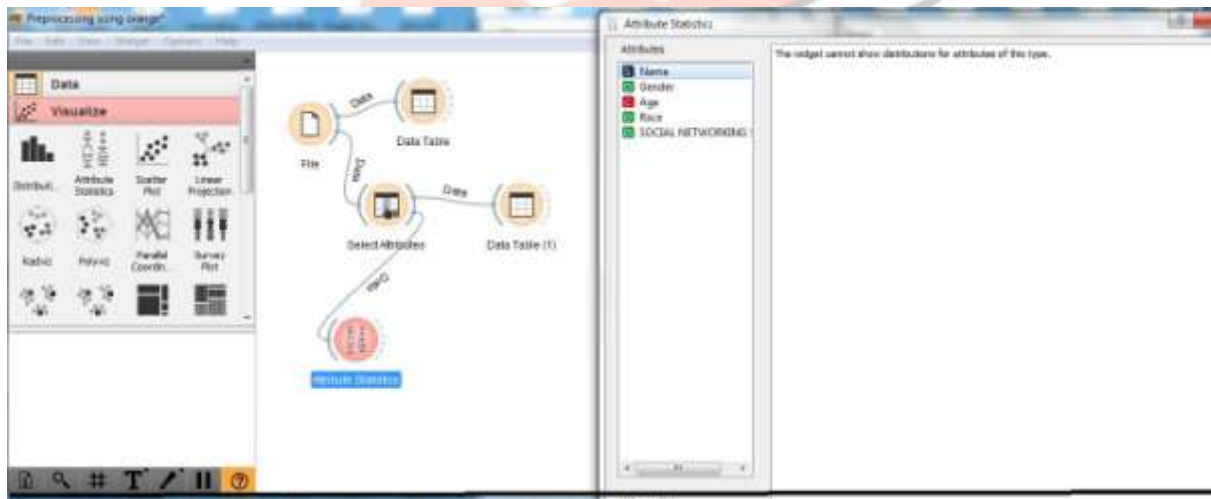
Orange uses for practical Implementation:

> -Data Analysis:
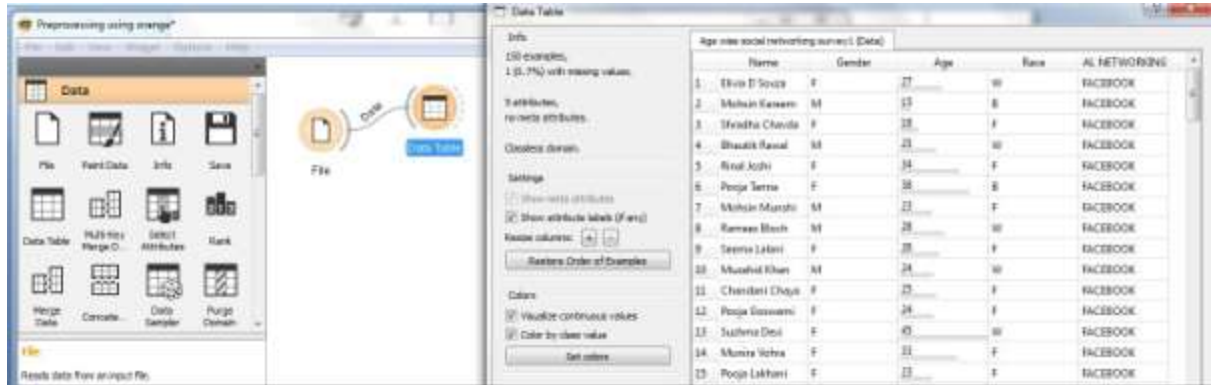
(Figure-3 Data Analysis using Orange)

In Figure-3 Orange has performed practical of data analysis along with some csv (Comma Separated Values) to File Data menu tool. It co-relates with sources which provided to the Data Table.
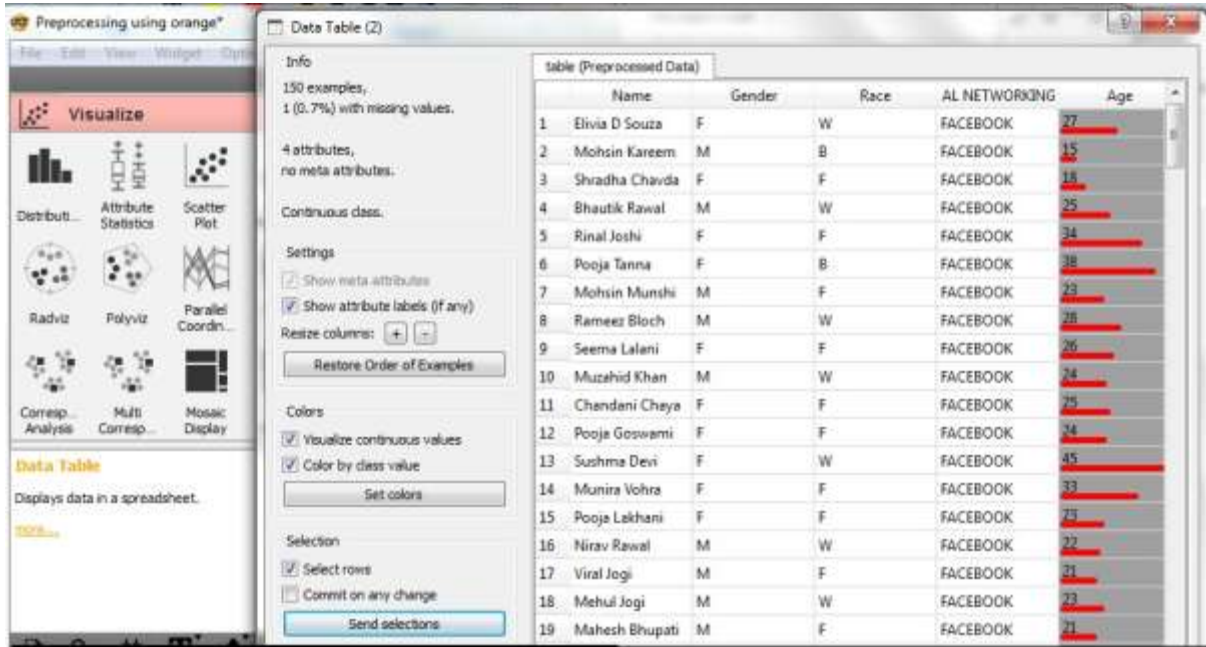
-Data Visualization:

(Figure-4 Data Visualization using Orange)

In Figure-4 Orange has performed practical of data visualization along with some csv (Comma Separated Values) to Visualize Data menu tool. It co-relates with sources which

provided to the Data Table. There are attributes specified in csv file, It consists Name, Gender, Age, Race and Social Networking to represent analyze data.

➢ -Data Pre-Processing:

(Figure-5 Data Visualization using Orange)

III.   Result:

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

Mean:

$$\overline{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \lambda\ + X_n}{n}$$

Median:

$$Median = L_1 + \left(\frac{\frac{N+1}{2} - F_{j-1}}{N_j} \right) \times h$$

Mean value describe the age of male and female where as median display the middle value of data.

IV.   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 data. 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.

V.   References:

1. Zhang, H., Raitoharju, J., Kiranyaz, S., & Gabbouj, M. (2016). Limited random walk algorithm for big graph data clustering. Journal of Big Data, 3(1), 26.

2. Pempek, Tiffany A., Yevdokiya A. Yermolayeva, and Sandra L. Calvert. "College students' social networking experiences on Facebook." Journal of applied developmental psychology 30.3 (2009): 227-238.

3. Griffiths, Mark D. "Facebook addiction: concerns, criticism, and recommendations—a response to Andreassen and colleagues." Psychological Reports 110.2 (2012): 518-520.