Heart Disease Data Analysis Using With Orange Data Mining Tool

Sarangam Kodati[1], Dr. R P. Singh[2]
Sri Satya Sai University of Technology and Medical Science, Sehore,
Bhopal, Madhya Pradesh , India

Abstract- *Data Mining is a process of computing models or design in large collection of data. Orange tools to analyze, visualize and extract data using data mining. The Orange tool are compatible to perform all analysis operations, In this paper we have attempted data mining tool for analysis of heart disease.*

Keywords- Data Mining, Orange,Attribute Statistics, Heart Disease Datasets.

I. INTRODUCTION

Orange is a Python-based tool for Data Mining life developed at the Bioinformatics Laboratory concerning the Faculty of Computer then Information Science at the University of toughness Ljubljana. It can keep used either through Python scripting as a Python plug-in, or through visual programming. Its visible programming interface, Orange Canvas, provides a structured view on supported functionalities grouped into 9 categories: data operations, visualization, classification, regression, evaluation, unsupervised learning, association, visualization using Qt, and prototype implementations.

Functionalities are visually represented by distinctive widgets (e.g. read file, discretize, train SVM classifier etc). A brief description over each widget is accessible inside the interface. Programming is performed by placing widgets regarding the canvass and connecting theirs inputs and outputs. The interface is very polished and visually appealing, offering a pleasant user experience. One apparent downside about Orange is that the number on available widgets seems limited when compared according to other tools such as RapidMiner or KNIME, especially because of the lack about integration with Weka. Still, the coverage of standard data mining techniques is quite good, as much can lie seen from Table. Furthermore, there are a number about interesting widgets currently in improvement that can lie found of the "Prototype" category, so that is reasonable according to expect that the feature set will be expanded within the future.

II. DATASET INFORMATION HEART DISEASE

This database includes seventy six attributes, but every published experiments allude to using a subset about 14 on them. In particular, the Cleveland database is the only certain as has been old by way of Machine Learning researchers in imitation of this date. The "goal" field refers in imitation of the emergence of bravery disorder between the patient. It is integer dear out of 0 (no presence) to 4 Experiments together with the Cleveland database hold concentrated concerning absolutely attempting to distinguish emergence (values 1,2,3,4) beyond non appearance (value 0).

The names yet communal security numbers over the sufferers had been lately eliminated beyond the database, changed with appointed values. One file has been "processed", that one containing the Cleveland database. All IV unprocessed archives also inhabit between this directory. Longevity To parley Test Costs (donated by means of Peter Turney), opt confer the folder "Costs".

Table:1. Heart disease database of instance

Database	# of instances
Cleveland	303
Hungarian	294
Switzerland	123
Long Beach VA	200

Table: 2.Heart disease Attribute Information: Only 14 used.

S.No:	Attribute Information:
1	#3 (age)
2	#4 (sex)
3	#9 (cp)
4	#10 (trestbps)
5	#12 (chol)
6	#16 (fbs)
7	#19 (restecg)
8	#32 (thalach)
9	#38 (exang)
10	#40 (oldpeak)
11	#41 (slope)
12	#44 (ca)
13	#51 (thal)
14	#58 (num)

Missing Attribute Values: Several. Distinguished with value - 9.0.

Table:3. Heart disease Class Distribution

Database	0	1	2	3	4	Total
Cleveland	164	55	36	35	13	303
Hungarian	188	37	26	28	15	294
Switzerland	8	48	32	30	5	123
Long Beach VA	51	56	41	42	10	200

III. METHODOLOGY

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 uses for practical Implementation:

IV. DATA SET ANALYSIS

The Data Table widget receives one or more heart disease data sets in its input and presents them as a spreadsheet. Data instances may be sorted by attribute values. The widget also supports manual selection of data instances. Orange uses for practical Implementation:

The name of the heart disease data set (usually the input information file). Data instances are in rows and theirs attribute values within columns. In it example, the data set is sorted by using the attribute "sepal length". Info on heart disease dataset size and number and types of attributes. Values of continuous attributes can remain visualized together with bars; colors may be attributed to different classes. Data instances (rows) can be select and sent to the widget's output channel. Use the Restore Original Order button according to

reorder data instances after attribute-based sorting. Produce a report. While auto send is on, all changes will be automatically communicated in accordance with other widgets. Otherwise, press Send Selected Rows.

The Data Table widget receives one or more data sets of its input and presents to them as like a spreadsheet.

Data instances may additionally be sorted by attribute values. The widget also supports manual selection of data instances.

We used two File widgets to read the heart disease data set (provided in Orange distribution), and send them to the Data Table widget. Selected data instances in the first Data Table 1 and Data Table2
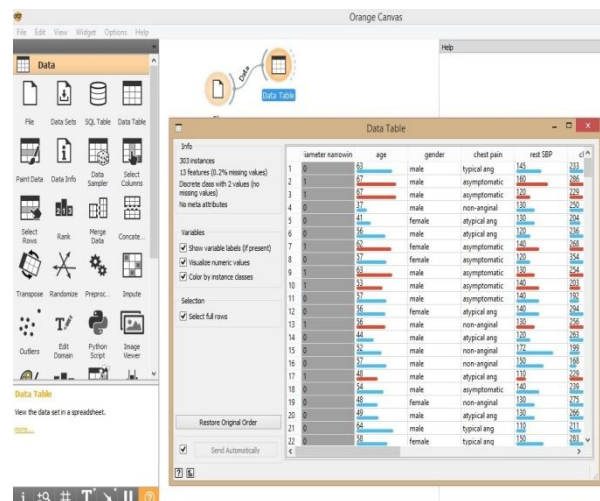

Fig :1. Heart Disease Chest pain data table Orange data mining tool.

V. DATA VISUALIZATION

A. Distributions

The distributions widget displays the value distribution of distinct or continuous attributes. If the data contains a class variable, distributions may stay conditioned concerning the class .Displays value distributions for a single attribute.

Heart disease discrete attributes, the graph displayed through the widget shows how many times (e.g., in what many instances) each attribute value appears into the data. If the data contains a class variable, class distributions because each over the attribute values will stand displayed as well (like of the snapshot below). In discipline in imitation of create this graph, we used the heart disease data set. A list of variables for distributions display.

If bin continuous variables is ticked, the widget desire discretize continuous variables with the aid of assigning them to intervals. The number on intervals is set by precision scale. Alternatively, you can set smoothness because of the distribution curves of continuous variables.

The widget may be requested after display value distributions only for instances concerning certain classification (Group by). Show relative frequencies will scale the records by percentage of the data set.

Show probabilities. Save image saves the graph to your computer in a .svg or .png format. Produce a report.

Heart disease continuous attributes, the attribute values are displayed as a function graph. Class probabilities for continuous attributes are obtained with gaussian kernel density estimation, while the appearance of the curve is set with the Precision bar (smooth or precise). For the purpose of it example, we used the heart disease data set.

Fig :2. heart disease chest pain data visualization Orange data mining tool

B. Scatterplot Cluster

The Scatterplot widget presents a 2-dimensional scatterplot visualization because of each continuous and discrete-valued attributes. The data is displayed as much a collection of points, each having the value concerning the x-axis attribute determining the position concerning the horizontal axis and the value about the y-axis attribute determining the position over the vertical axis. Various properties of the graph, like color, size, and shape of the points, axle titles, maximum point size and jittering do be adjusted regarding the left side concerning the widget. A snapshot below shows the scatterplot regarding the heart

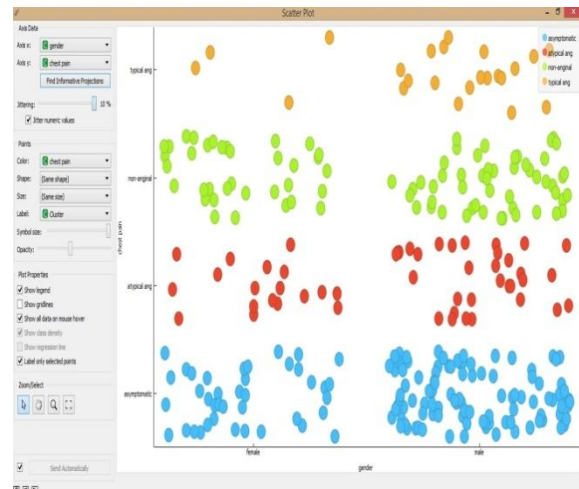disease data accept with the coloring maching of the class attribute.

Fig :3. Heart disease chest pain scatterplot cluster Orange data mining tool.

VI. CONCLUSION

In this study of data analysis heart disease using data mining tool, comparing their parameters to data visualization and Scatterplot Cluster each other 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

[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 recommendation a response to Andreassen and colleagues." Psychological Reports 110.2 (2012): 518- 520.

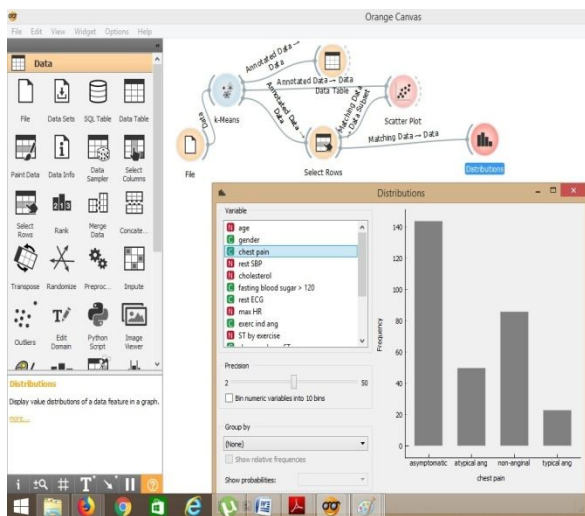[4] Inyan Li and Limsoon Wong. Using Rules to Analyse Bio-medical Data: A Comparison between C4.5 and PCL. WAIM. 2003.