MODULE 1- Lecture 5	Data Mining Tools

Data Mining tools

Data Mining is the set of techniques that utilize specific algorithms, statical analysis, artificial intelligence, and database systems to analyze data from different dimensions and perspectives.

Data Mining tools have the objective of discovering patterns/trends/groupings among large sets of data and transforming data into more refined information.

It is a framework, such as Rstudio or Tableau that allows you to perform different types of data mining analysis.

We can perform various algorithms such as clustering or classification on your data set and visualize the results itself. It is a framework that provides us better insights for our data and the phenomenon that data represent. Such a framework is called a data mining tool.

The Market for Data Mining tool is shining: as per the latest report from ReortLinker noted that the market would top **$1 billion** in sales by **2023**, up from **$ 591** million in **2018**

These are the most popular data mining tools:

1. Orange Data Mining:

Orange is a perfect machine learning and data mining software suite. It supports the visualization and is a software-based on components written in Python computing

language and developed at the bioinformatics laboratory at the faculty of computer and information science, Ljubljana University, Slovenia.

As it is a software-based on components, the components of Orange are called "widgets." These widgets range from preprocessing and data visualization to the assessment of algorithms and predictive modeling.

Widgets deliver significant functionalities such as:

○ Displaying data table and allowing to select features

○ Data reading

○ Training predictors and comparison of learning algorithms

○ Data element visualization, etc.

Besides, Orange provides a more interactive and enjoyable atmosphere to dull analytical tools. It is quite exciting to operate.

Why Orange?

Data comes to orange is formatted quickly to the desired pattern, and moving the widgets can be easily transferred where needed. Orange is quite interesting to users. Orange allows its users to make smarter decisions in a short time by rapidly comparing and analyzing the data.It is a good open-source data visualization as well as evaluation that concerns beginners and professionals. Data mining can be performed via visual programming or Python scripting. Many analyses are feasible through its visual programming interface(drag and drop connected with widgets)and many visual tools tend to be supported such as bar charts, scatterplots, trees, dendrograms, and heat maps. A substantial amount of widgets(more than 100) tend to be supported.

The instrument has machine learning components, add-ons for bioinformatics and text mining, and it is packed with features for data analytics. This is also used as a python library.
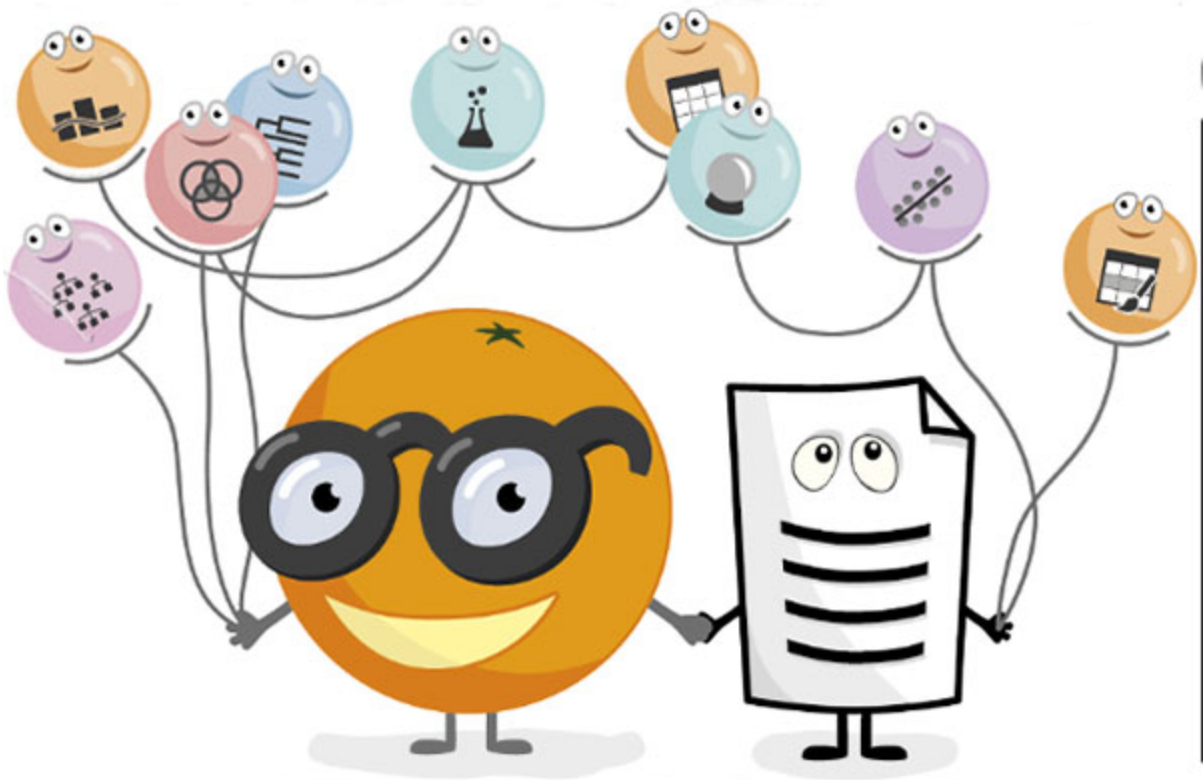
Python scripts can keep running in a terminal window, an integrated environment like PyCharmand PythonWin, pr shells like iPython. Orange comprises of canvas interface onto which the user places widgets and creates a data analysis workflow. The widget proposes fundamental operations, For example, reading the data, showing a data table, selecting features, training predictors, comparing learning algorithms, visualizing data elements, etc. Orange operates on Windows, Mac OS X, and a variety of Linux operating systems. Orange comes with multiple regression and classification algorithms.

Orange can read documents in native and other data formats. Orange is dedicated to machine learning techniques for classification or supervised data mining. There are two types of objects used in classification: learner and classifiers. Learners consider class-leveled data and return a classifier. Regression methods are very similar to classification in Orange, and both are designed for supervised data mining and require class-level data. The learning of ensembles combines the predictions of individual models for precision gain. The model can either come from different training data or use different learners on the same sets of data.

Learners can also be diversified by altering their parameter sets. In orange, ensembles are simply wrappers around learners. They act like any other learner. Based on the data, they return models that can predict the results of any data instance.

2. SAS Data Mining:

SAS stands for Statistical Analysis System. It is a product of the SAS Institute created for analytics and data management. SAS can mine data, change it, manage information from various sources, and analyze statistics. It offers a graphical UI for non-technical users.

SAS data miner allows users to analyze big data and provide accurate insight for timely decision-making purposes. SAS has distributed memory processing architecture that is highly scalable. It is suitable for data mining, optimization, and text mining purposes.

3. DataMelt Data Mining:

DataMelt Data Mining
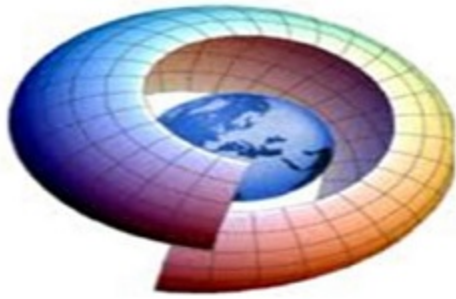
DataMelt is a computation and visualization environment which offers an interactive structure for data analysis and visualization. It is primarily designed for students, engineers, and scientists. It is also known as DMelt.

DMelt is a multi-platform utility written in JAVA. It can run on any operating system which is compatible with JVM (Java Virtual Machine). It consists of Science and mathematics libraries.

○ **Scientific libraries:**

Scientific libraries are used for drawing the 2D/3D plots.

○ **Mathematical libraries:**

Mathematical libraries are used for random number generation, algorithms, curve fitting, etc.

DMelt can be used for the analysis of the large volume of data, data mining, and statistical analysis. It is extensively used in natural sciences, financial markets, and engineering.

4. Rattle:

Rattle Data Mining

Ratte is a data mining tool based on GUI. It uses the R stats programming language. Rattle exposes the statical power of R by offering significant data mining features. While rattle has a comprehensive and well-developed user interface, It has an integrated log code tab that produces duplicate code for any GUI operation.

The data set produced by Rattle can be viewed and edited. Rattle gives the other facility to review the code, use it for many purposes, and extend the code without any restriction.

5. Rapid Miner:

Rapid Miner is one of the most popular predictive analysis systems created by the company with the same name as the Rapid Miner. It is written in JAVA programming language. It offers an integrated environment for text mining, deep learning, machine learning, and predictive analysis.

The instrument can be used for a wide range of applications, including company applications, commercial applications, research, education, training, application development, machine learning.

Rapid Miner provides the server on-site as well as in public or private cloud infrastructure. It has a client/server model as its base. A rapid miner comes with template-based frameworks that enable fast delivery with few errors(which are commonly expected in the manual coding writing process)