

Data Mining

Munawar, PhD

13. Data Mining Tools



Oracle Data Mining

1. Oracle Data Mining

- It is Proprietary License Software.
- It provides us data mining algorithms for data classification, prediction, regression and specialized analytics that enables analysts to analyse insights, make better predictions, target best customers, identify cross-selling opportunities & detect fraud.
- The algorithms designed inside ODM leverage has the strengths of Oracle database.
- The data mining feature of SQL can dig data out of database tables, views, and schemas.
- The GUI of Oracle data miner is an extended version of Oracle SQL Developer.
- It provides a facility of direct 'drag & drop' of data inside the database to users thus giving better insight.

IBM SPSS Modeler

2. IBM SPSS Modeler

- It is Proprietary License Software.
- It was originally produced by SPSS Inc. and later acquired by IBM.
- It is a software suite owned by [IBM](#) that is used for data mining & text data analytics.
- It has a rich visual interface that allows users to work with data mining algorithms without the need of programming.
- It eliminates the unnecessary complexities faced during data transformations and to make easy to use predictive models.
- IBM SPSS comes in two editions, based on the features
- IBM SPSS Modeler Professional
- IBM SPSS Modeler Premium- contains additional features of text analytics, entity analytics etc.



3. Weka

- It is open source and free software.
- It is best suited for data analysis and predictive modelling.
- It contains algorithms and visualization tools that support data mining tasks and machine learning.
- Weka has a GUI that gives easy access to all its features.
- It is written in JAVA language.
- Weka supports major data mining tasks including data mining, processing, visualization, regression etc.
- Weka can also provide access to SQL Databases through database connectivity and can further process the data returned by the query.

IBM Cognos

4. IBM Cognos

- It is Proprietary License Software.
- IBM Cognos BI is an intelligence suite owned by IBM for reporting and data analysis, score carding etc.
- It consists of sub-components that meet specific organizational requirements Cognos Connection, Query Studio, Report Studio, Analysis Studio, Event studio & Workspace Advance.
 - **Cognos Connection:** It is a web portal to gather and summarize data in scoreboard/reports.
 - **Query Studio:** It contains queries to format data & create diagrams.
 - **Report Studio:** It is used to generate management reports.
 - **Analysis Studio:** It is used to process large data volumes, understand & identify trends.
 - **Event Studio:** Notification module to keep in sync with events.
 - **Workspace Advanced:** User-friendly interface to create personalized & user-friendly documents.

Rapid Miner

5. Rapid Miner Software

- It is an Open source software.
- It is one of the best predictive analysis systems developed by the company with the same name as the Rapid Miner.
- Written in JAVA programming language.
- It provides an integrated environment for deep learning, text mining, machine learning & predictive analysis.
- It can be used for over various applications such as business applications, commercial applications, training, education, research, application development, machine learning.
- Rapid Miner offers the server as both on premise & in public/private cloud infrastructures.
- It has a client/server model as its base.
- Rapid Miner comes with template-based frameworks that enable speedy delivery with reduced number of errors.
- Rapid Miner has three modules,
 1. Rapid Miner Studio- This module is for workflow design, prototyping, validation etc.
 2. Rapid Miner Server- To operate predictive data models created in studio
 3. Rapid Miner Radoop- Executes processes directly in Hadoop cluster to simplify predictive analysis.



6. Orange Software

- It is an Open Source Software
- It is a suite for machine learning & data mining tasks.
- It best aids the data visualization and is a component-based software.
- It is written in Python language.
- Data coming to Orange software gets quickly formatted to the desired pattern and it can be easily moved where needed by simply moving/flipping the widgets.
- It is a component-based software, the components of orange are called 'widgets'. These widgets range from data visualization & pre-processing to an evaluation of algorithms and predictive modelling.
- Widgets provide major functionalities like
 - Showing data table and allowing to select features
 - Reading the data
 - Training predictors and to compare learning algorithms
 - Visualizing data elements etc.



7.KNIME

- It is an Open Source Software.
- It is written in JAVA.
- KNIME is integration platform for data mining and data analytics.
- It works on the concept of the modular data pipeline.
- It consists of various machine learning and data mining components embedded together.
- It is used in pharmaceutical research. Also, it performs excellently for customer data analysis, financial data analysis, and business intelligence.
- It has features of quick deployment and scaling efficiency.
- It is easy to learn software.

SISENSE

8.Sisense

- It is not open source i.e., it is a Licensed software
- It is developed by the company of same name 'Sisense'.
- It is useful and best suited BI software when it comes to reporting purposes within the organization.
- It has a brilliant capability to handle and process data for the small scale/large scale organizations.
- It allows combining data from various sources to build a common repository and further, refines data to generate rich reports that get shared across departments for reporting.
- Sisense generates reports which are highly visual. It is specially designed for users that are non-technical. It allows drag & drop facility as well as widgets.
- Different widgets can be selected to generate the reports in form of pie charts, line charts, bar graphs etc. based on the purpose of an organization. Reports can be further drilled down by simply clicking to check details and comprehensive data.



9.R Tool

- R is an open source and free software environment.
- This is written in C and FORTRAN Language.
- It is used to perform statistical computing, data mining & graphics.
- It also supports graphical analysis, both linear and nonlinear modelling, classification, clustering and time-based data analysis.
- It is widely used in academia, research, engineering & industrial applications.

MAHOUT

10. Apache Mahout

- It is an Open source software.
- It is developed by [Apache Foundation](#) that serves the primary purpose of creating machine learning algorithms.
- It focuses mainly on data clustering, classification, and collaborative filtering which is primary task in data mining.
- It has been written in JAVA.
- It includes JAVA libraries to perform mathematical operations like linear algebra and statistics.
- The algorithms have implemented a level above Hadoop through mapping/reducing templates.
- To key up, Mahout has following major features
 - Extensible programming environment
 - Pre-made algorithms
 - Math experimentation environment
 - GPU computes for performance improvement

SAS DATA MINING

11.SAS Data Mining

- It is having a Proprietary License.
- This is a product of SAS Institute developed for analytics & data management.
- SAS can mine data, alter it, manage data from different sources and perform statistical analysis.
- It provides a graphical UI for non-technical users.
- SAS data miner enables users to analyse big data and derives accurate insight to make timely decisions.
- SAS has a distributed memory processing architecture which is highly scalable. It is well suited for data mining, text mining & optimization.

TERADATA

12.Teradata

- It is a Licensed Software.
- It is commonly known as a Teradata database.
- It is an enterprise data warehouse that contains data management tools along with data mining software.
- It can be used for business analytics.
- Teradata is used to have an insight of company data like sales, product placement, customer preferences etc.
- Teradata works on 'share nothing' architecture as it has its server nodes have their own memory & processing ability.



THANK YOU

Munawar, PhD – moenawar@gmail.com – www.moenawar.web.id

