



**Data mining/Predictive analytics/Machine learning
software Recommendation Report**

UPDATED: AUGUST 7, 2017

ROSHAN T.



TABLE OF CONTENTS

Overview	2
Software Category I – Novice category	3
XLMiner	3
DataMinerXL	5
Alyuda ForecasterXL	6
Software Category II - Intermediate category	8
RapidMiner Studio	8
AdvancedMiner	10
KNIME	12
Software Category III - Expert category	13
Orange/Python	13
WEKA/Java	14
Rattle GUI/R	15

Overview

Data mining/Predictive analytics/machine learning is about extracting knowledge from data. It is a fast growing trend in the health care industry largely due to the advancement of wearable devices and sensors that can use data to assess a patient's health in real time. It can also help medical experts analyze data to identify trends or red flags that may lead to improved diagnoses and treatments.

Data mining software mines and analyzes historical data patterns to predict future outcomes by extracting information from data sets to determine patterns. In other cases, data mining software can find clusters of information based on logical relationships, or they look at association and sequential patterns to draw conclusions about trends.

Software included in this report are both open source and licensed versions and are categorized into three different categories based on users. Also, this report does not cover any technical details on how to use the software rather it describes what each is capable of.

Software Recommendation Categories

- **Category I – Novice category**

These software can be used by people who previously never would have considered attempting predictive analytics or writing codes but have some knowledge or who can use an *Excel* like environment to build advanced predictive models.

Suggested software under this category:

- XLMiner – Licensed

- DataMinerXL – Licensed

- Alyuda ForecasterXL (Neural network within Excel) - Licensed

- **Category II – Intermediate category**

These software can be use by people who have some knowledge in building predictive models and knows how they work and need more control over building models but still prefer not to write any advanced codes.

Suggested software under this category:

- Rapid Miner – Licensed

- AdvancedMiner – Licensed and Community Edition

- KNIME – Open source

- **Category III – Expert category**

These software can be use by people who have previous knowledge of how to write advanced codes and writing predictive models from scratch.

Suggested software under this category:

- Orange/Python – Open source

- WEKA/Java – Open source

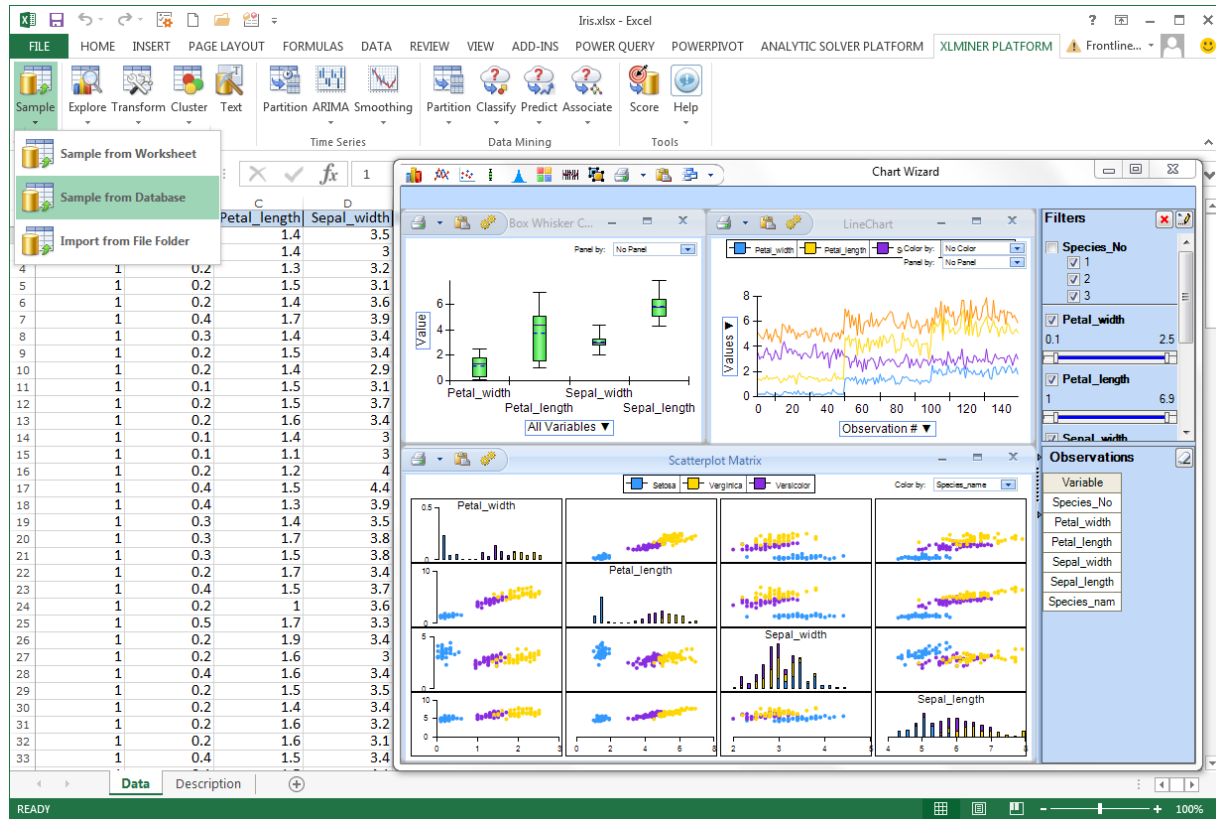
- Rattle GUI/R – Open source

Disclaimer: Please note that this recommendation involves the sole views of myself and it does not represent any entity or institute.

Software Category I – Novice category

1. XLMiner : <https://www.solver.com/xlminer-platform>

XLMiner provides Predictive Analytics, Forecasting, Text mining, Data Visualization for large datasets. Users do not have to learn the software from the scratch, those who are already familiar with Microsoft Excel can make an effective use of the built-in tool. The software allows data to be imported from spreadsheets, Microsoft PowerPivot, Microsoft/IBM/Oracle databases and data can be analyzed how a user wants.



Features included are:

- Use data from many sources -
Sample data from spreadsheets, text files and SQL databases, including Microsoft's PowerPivot in-memory database handling 100 million rows or more.
- Visualize your data –
Use visualization aids from simple bar, line and histogram charts to multiple linked charts, instant axis changes, colors, panels, zooming, brushing and more.
- Easy to use interface –
Use a Ribbon interface, with Wizard-style dialogs to guide you through model and option selections.
- Clean & transform data –
Clean and transform your data with a comprehensive set of data handling utilities including categorizing data and handling missing values.

- Prediction methods –
Use powerful multiple linear regression with variable selection, and data mining methods like k-nearest neighbors, and ensembles of regression trees and neural networks.
- Classification methods –
Use classical discriminant analysis and logistic regression, and data mining methods like k-nearest neighbors, naive Bayes, and ensembles of classification trees and neural networks.
- Forecast time series –
Apply the most popular exponential smoothing and Box-Jenkins (ARIMA) methods with seasonality to forecast time series, such as sales and inventory, from historical data.
- Analyze text for insights –
Automatically transform free-form text into structured data, identifying most frequent terms and extracting key concepts with latent semantic indexing
- Identify key features –
Use feature selection to automatically identify columns or variables with the greatest explanatory power for your desired classification or prediction task.
- Partition for training –
Easily partition your data into training, validation, and test datasets, with no limits on dataset size -- even "on the fly" as you build a predictive model.
- Reduce and cluster data –
Use principal components to reduce columns, and k-means clustering or hierarchical clustering to group data by rows.

Pricing:


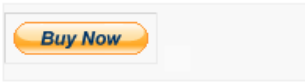
- 1 year single user academic subscription \$500.
- Do not have a site license.
- After purchase support : email or telephone tech support only, M-F 8am-5pm
- For additional information contact sales@solver.com or duane@solver.com

2. **DataMinerXL** : <http://www.dataminerxl.com/>

DataMinerXL is a Microsoft Excel add-in which provides a collection of functions for building predictive models. It supports the creation of predictive models using a wide variety of techniques, including regression (linear and logistic), naive Bayes, decision trees, neural networks, support vector machines (SVM) and will even solve linear, quadratic and linear complementarity problems. Basic statistical functions are also included.

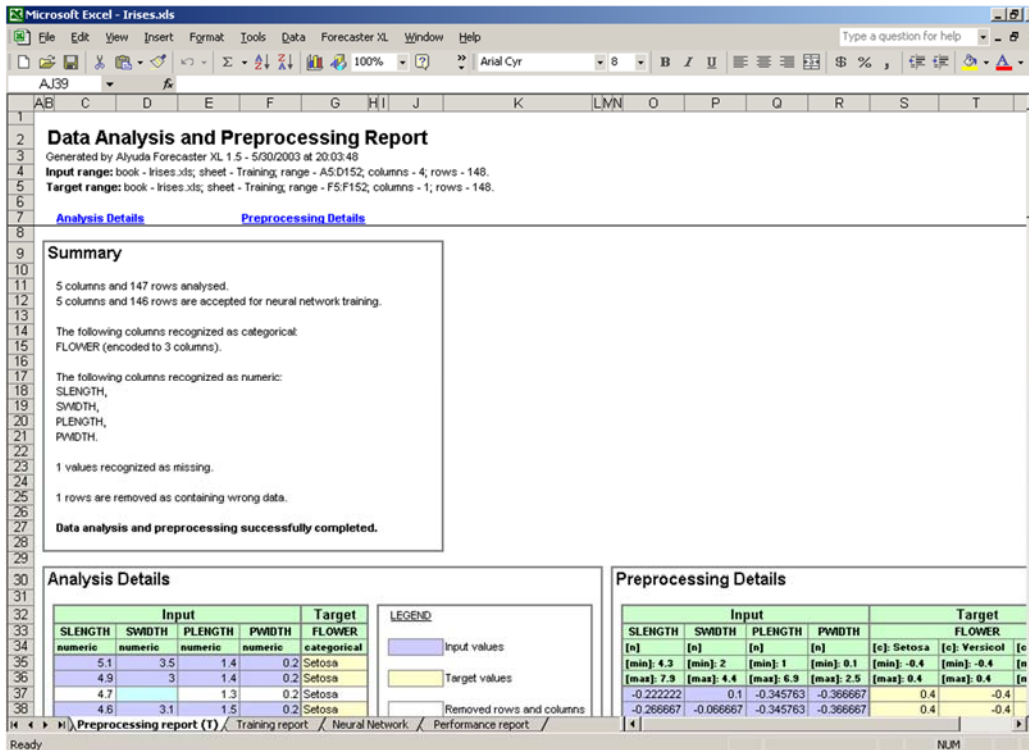
Features included are:

- Perform basic statistical analysis
- Analyze importance of variables using weight of evidence (WOE) transformation
- Perform predictive analytics -
 - Perform principal component analysis and factor analysis, build linear regression models, build partial least square regression models, build logistic regression models, perform time-series analysis, perform linear discriminant analysis and quadratic discriminant analysis, build naive Bayes classifier models, build neural network models, build support vector machine (SVM) models, build decision tree-based models, perform clustering and segmentation analysis
- Estimate the parameters of a stochastic process: normal (Gaussian) process, lognormal process, and shifted lognormal process
- Evaluate model performance given a model
- Solve linear programming, quadratic programming, and linear complementary programming problems
- Pricing: Single user student and professional edition available. Site license is available and will depend on number of users. On-site training may be available. Contact DataMinerXL team at support@dataminerxl.com for more information about On-site training.

Edition	Trial Edition	Student Edition	Professional Edition
License	30 Days Free	1 User License: Unlimited (perpetual) use. You can download the updated version for one year from date of purchase	1 User License: Unlimited (perpetual) use. You can download the updated version for one year from date of purchase
Delivery of License File	No License Needed	Email Delivery of License File	Email Delivery of License File
Maximum Records of Data	1,000 Records	10,000 Records	Unlimited
Functions Included	All Functions	All Functions	All Functions
Price	Free	\$99.00 	\$499.00 

3. **Alyuda ForecasterXL** : <http://www.alyuda.com/forecasting-excel-software-with-neural-network.htm>

By providing embedded support for Microsoft Excel, Alyuda allows users to use Neural Networks for forecasting directly inside the Excel worksheets. Users can instantly apply forecasting capabilities of neural networks to Excel data, while still retaining all of Excel's data manipulation and formatting tools.



Features included are:

- **Ease of Use –**
All features are easily accessed from additional menu items and use only standard Excel interface for data manipulation
- **Forecasting in several clicks –**
To get forecasting simply click "Create Network..." from Forecaster XL menu, select input and target data and click "Train" to let Forecaster XL prepare the network. After this, user needs to just select or enter new data and make one more click to get the forecasting ready.
- **Hidden details of neural network theory –**
Forecaster XL frees user from the need to learn details of neural network theory. It greatly simplifies the process of preparing a neural network needed for forecasting, hiding from user any difficulty in neural network preparation and tuning.
- **Automated network selection -**
Forecaster XL automatically selects the most appropriate architecture for forecasting problems. It runs a reliable constructive algorithm which finds the suitable network architecture automatically and saves users a lot of time.
- **Detailed Reporting –**
For those who want to have a detailed report about the created neural network and its structure as well as to inspect network performance during forecasting and data preprocessing, Forecaster

XL delivers all needed information in separate Excel sheets. When creating a final report you may add training graph and error values of each iteration.

- Pricing –

- Single User \$197

- Unlimited Site -\$997

- 30% discount for educational and academic purposes

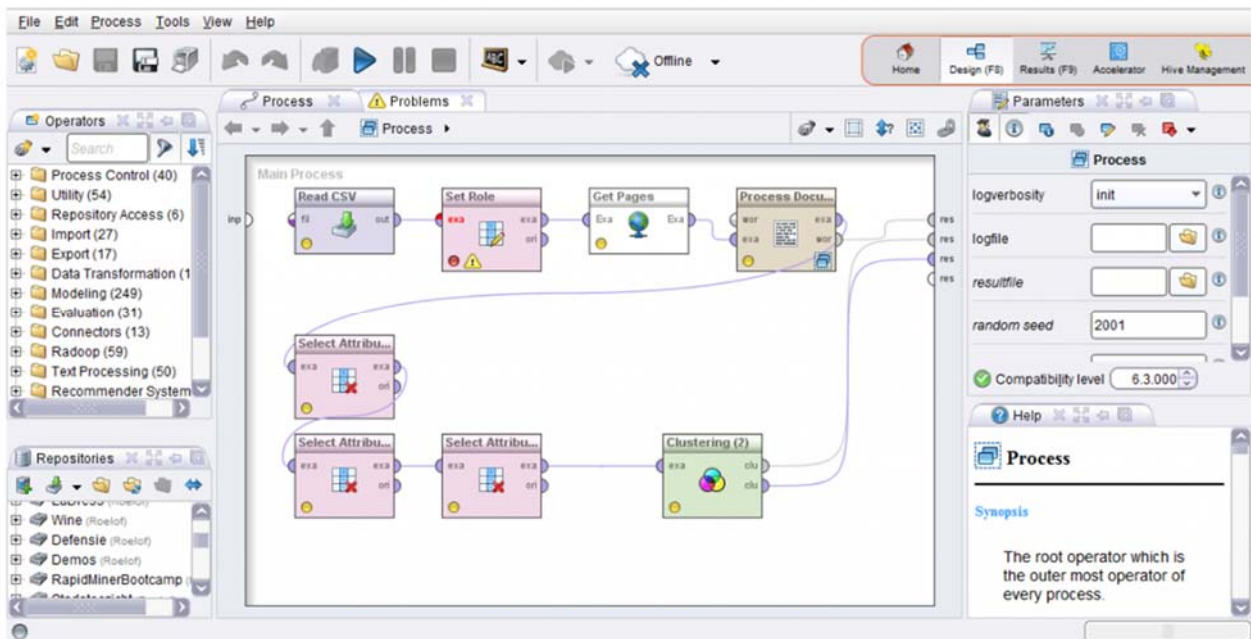
- Do not offer onsite training. After purchase support is free and done by email for ordinary situations

- More purchase info contact salesdept@alyuda.com

Software Category II - Intermediate category

1. RapidMiner Studio : <https://rapidminer.com/products/studio/>

RapidMiner provides an integrated environment for machine learning and is used for business and industrial application as well as for research, rapid prototyping and application development. It supports all steps of the data mining process including result visualization, validation, and optimization. It provides Graphical User Interface to design and execute analytic workflows.



Features included are:

- **Data Access and Management –**
With RapidMiner Studio user can access both traditional structured data and unstructured data like text, images, and media. It can also extract information from these types of data and transform unstructured data into structured data.
- **Data Exploration -**
Immediately understand and create a plan to prepare the data, automatically extract statistics and key information such as descriptive statistics and Graphs.
- **Data Preparation –**
The data preparation capabilities in RapidMiner Studio can handle most real-life data preparation challenges such as Data Partitioning, Sampling, Transformations, Weighting and Selection, Attribute Generation. User can format and create the optimal data set for predictive analytics. RapidMiner Studio can blend structured with unstructured data and then leverage all the data for predictive analysis. Any data preparation process can be saved for reuse.
- **Modeling –**
RapidMiner Studio equipped with set of modeling capabilities and machine learning algorithms for supervised and unsupervised learning such as Neural networks, Similarity Calculation, Rule Induction,

Bayesian Modeling, Regression, Clustering, Support Vector Machines, Decision Trees, Memory-Based Reasoning, Model Ensembles, Market Basket Analysis.

- Validation –

RapidMiner Studio provides the means to accurately and appropriately estimate model performance.

- Pricing:

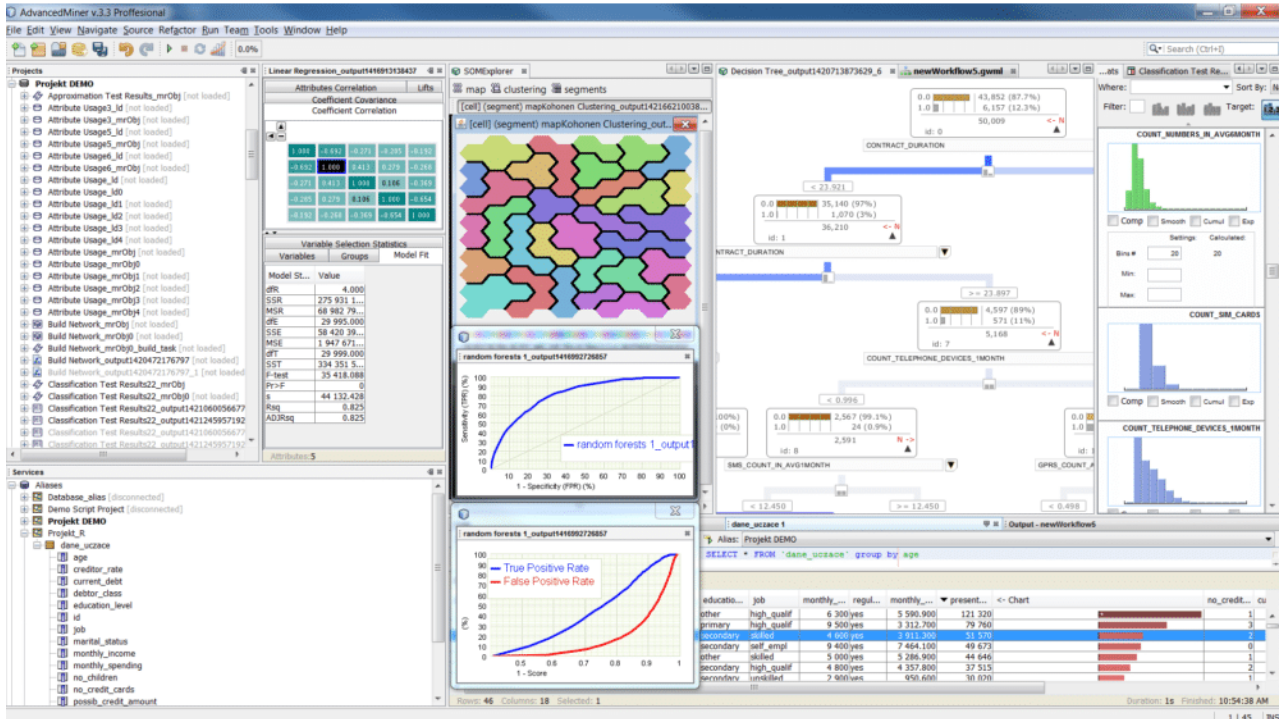
Priced by the number of logical processors and the amount of data used by the model. Offer an 80% discount for funded academic research; if it is institutional, need to discuss the pricing on a per use case basis. More purchase info and after purchase support contact through

<https://rapidminer.com/us/contact-us/> or rfarquhar@rapidminer.com

	FREE	SMALL	MEDIUM	LARGE
	Free	\$ 2,500 <small>Yearly</small>	\$ 5,000 <small>Yearly</small>	\$ 10,000 <small>Yearly</small>
# Data Rows	10,000	100,000	1,000,000	Unlimited
# Logical Processors	1	2	4	Unlimited
Performance Improvements		2x	4x	10x+
Background Process Execution				✓
Customer Support	Community	Enterprise	Enterprise	Enterprise
All RapidMiner Studio Features	✓	✓	✓	✓
	DOWNLOAD	CONTACT SALES	CONTACT SALES	FREE TRIAL

2. **AdvancedMiner** : <http://algolytics.com/>

Advanced Miner is an analytic tool for data processing, analysis and modelling. With graphical interface (work flow) it provides a complete and user friendly environment for data exploration.



Features included are:

- Extracting and saving data from/to different database systems and files
- Performing a wide range of operations on data, such as sampling, joining datasets, dividing into testing/training/validating sets, assigning roles to attributes
- Graphical and interactive data exploration
- Outlier filtering, supplying missing values, PCA, various data transformations, etc.
- Building association models, clustering analyses, variable importance analyses, etc.
- Constructing various analytical models with the use of diverse Data Mining and statistical algorithms (such as classification trees, neuron networks, linear and logistic regression, K-means)
- Model quality evaluation and comparison of Data Mining models (LIFT, ROK, K-S, Confusion Matrix)

- Pricing: Educational License 50% discount from price below.

Running on Windows Personal Computers or Virtual Desktops

Commercial License in EUR

Number of PC's	FYF (First Year)	Ryf (Renewal)
1	3,800	1,140
2	6,460	1,938
3	9,120	2,736
4	11,780	3,534
5	14,440	4,332
6	17,100	5,130
7	19,760	5,928
8	22,420	6,726
9	25,080	7,524
10	27,740	8,322

Pricing for Algolytics® Advanced Miner Server

Commercial License in EUR

Number of cores	FYF (First Year)	Ryf (Renewal)
1	30,000	15,000
2	33,000	16,500
4	36,000	18,000
6	39,000	19,500
8	42,000	21,000
10	45,000	22,500

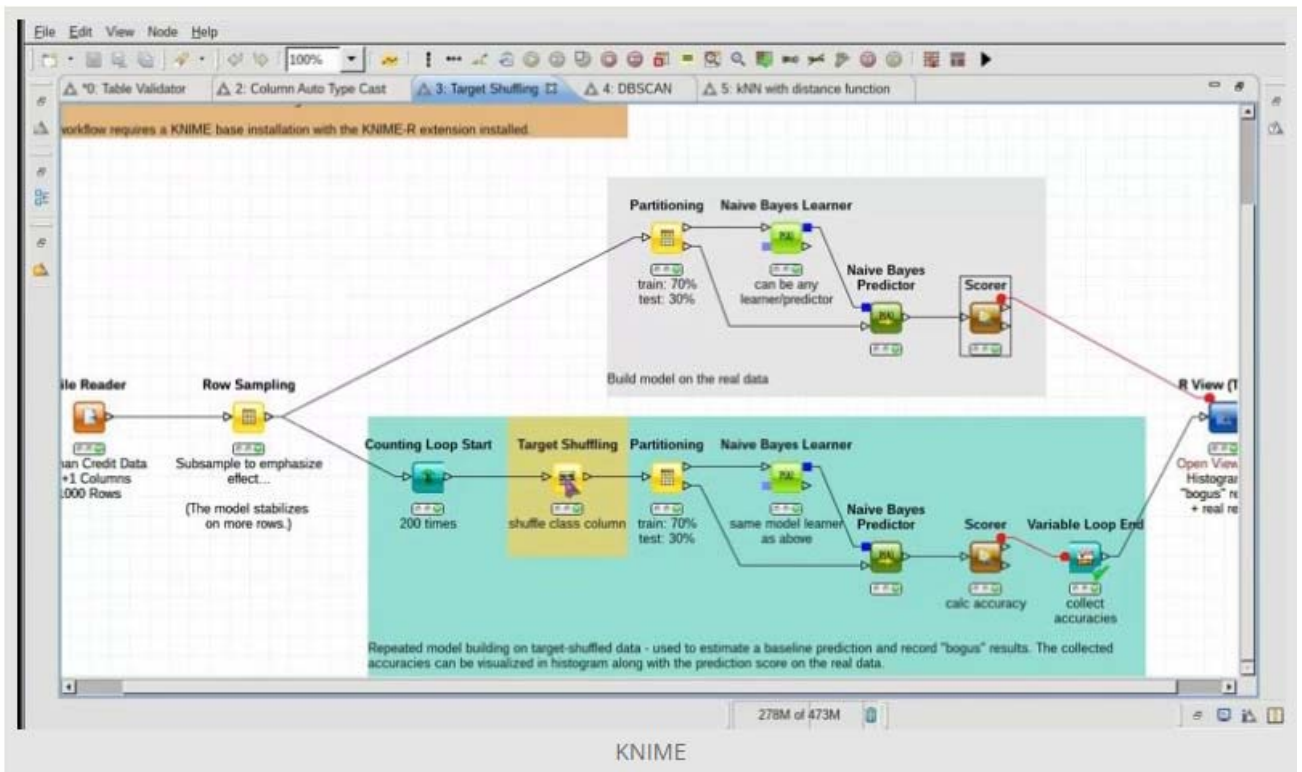
Standard maintenance is included in the price of the license.

Standard maintenance Problem / error class	Predicted time to solve the problem or restore the system to full efficiency
A	Up to 3 business days
B	Up to 7 business days
C	Up to 20 business days

The dates for resolving the problem are not guaranteed. Approximately 90% of the incidents are resolved by the dates indicated in the table. In exceptional situations caused by heavy loading of the Algolytics Service Center, or for applications with a high degree of complexity (for example, requiring a third support line), these terms may be slightly exceeded. For more purchase info contact info@algolytics.com or piotr.orlowski@algolytics.pl

3. **KNIME** : <https://www.knime.org/knime-for-data-scientists>

KNIME, the Konstanz Information Miner is an open source data analytics, reporting and integration platform. It integrates various components for machine learning and data mining through its modular data pipelining concept and provides a graphical user interface that allows assembly of nodes for data processing.



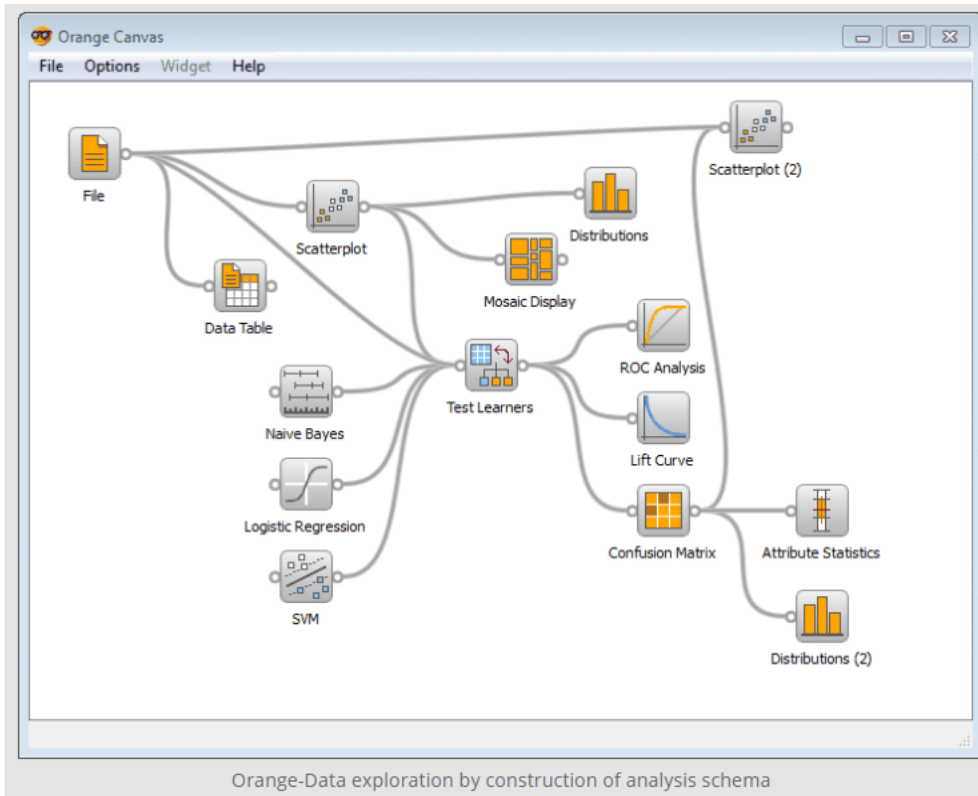
Features included are:

- Allow users to visually create data flow, selectively execute some or all analysis steps and inspect the results and models.
- Analytics –
Preprocessing data like feature elimination, Dimension Reduction and feature selection. Clustering, Classification and predictive modelling, Regression, Time series, Optimization, Deep learning, Meta Learning.
- Reporting –
Using the free report designer extension KNIME workflows can be used as data sets to create report templates that can be exported to documents formats like doc, ppt, xls, pdf and others.
- Pricing - Free

Software Category III - Expert category

1. Orange/Python : <https://orange.biolab.si/>

Orange is an open source data visualization and analysis tool. Data mining is done through visual programming or python scripting. The tool has a component for machine learning and add-ons for text mining. Orange uses common Python open-source libraries for scientific computing such as numpy, scipy and scikit-learn.

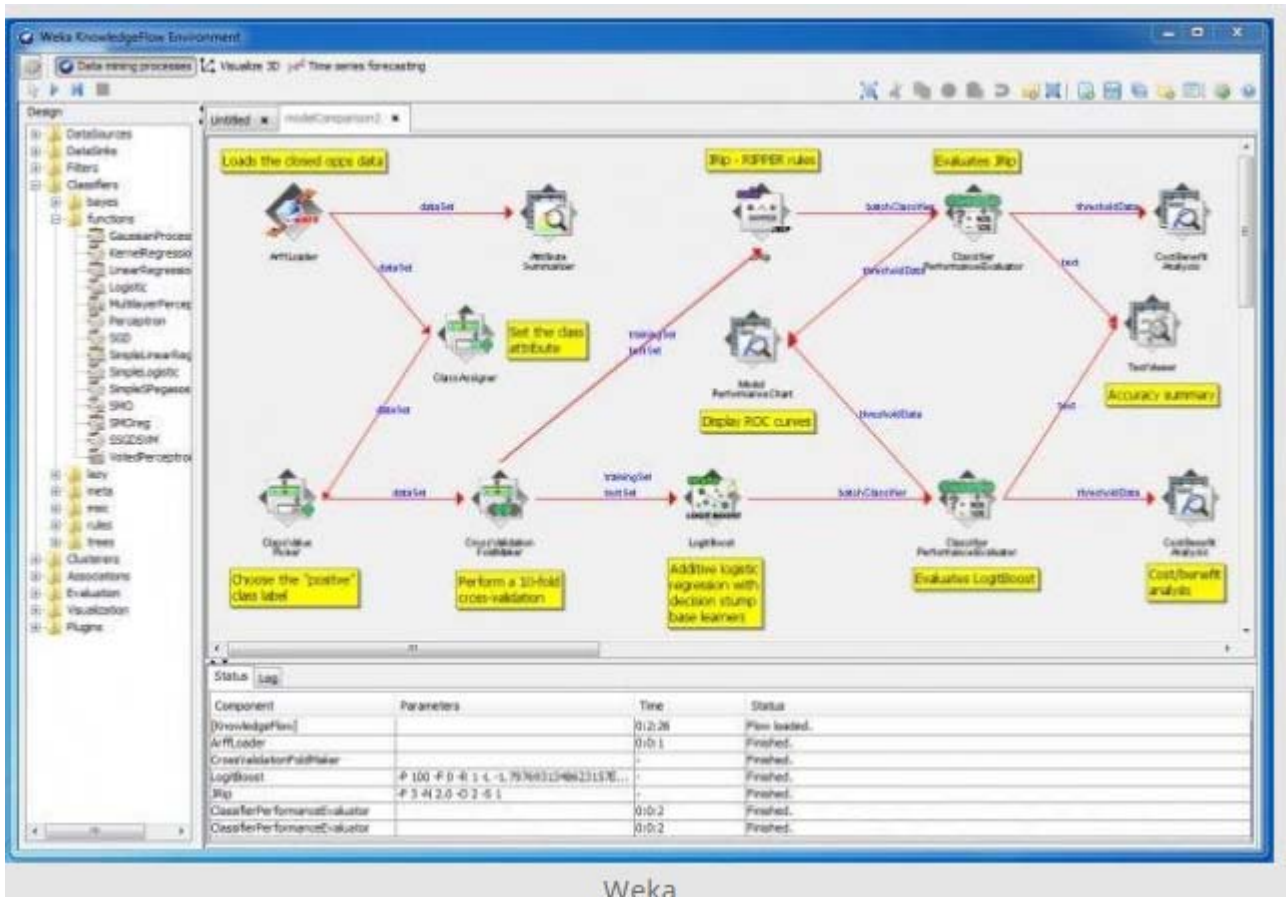


Features included are:

- Interactive Data Visualization –
Data visualizations in Orange help to uncover hidden data patterns, provide intuition behind data analysis procedures. Visualization tools include scatter plot, box plot and histogram, and model-specific visualizations like Dendrogram and tree visualizations, just to mention a few. Many other visualizations are available as add-ons and include visualizations of networks and more.
- Visual Programming –
In orange data analysis process can be designed through visual programming. It remembers the choices, suggest most frequently used combinations. By combining the various widgets the design of data analytics framework can be done.
- Creating Own widgets –
Own widgets can be developed and the scripting interface can be extended to create self-contained add-ons which integrates with the rest of Orange allows component and code reuse.
- Pricing - Free

2. **WEKA/Java** : <http://www.cs.waikato.ac.nz/ml/weka/index.html>

WEKA is a machine learning software application written in the Java programming language. It is a collection of machine learning algorithms for data mining tasks which can either be applied directly to a data set or called from your own Java code.



Features included are:

- Contains tools for data pre-processing, classification, regression, clustering, association rules and visualization.
- Well-suited for developing new machine learning schemes.
- Provides access to SQL databases using Java Database Connectivity and can process the results returned by a database query.
- Pricing - Free

3. **Rattle GUI/R** : <https://rattle.togaware.com/>

Rattle GUI is a free open source software providing a graphical user interface for data mining using R statistical programming language. It provides considerable data mining functionality by exposing the power of R statistical software through graphical user interface.



Features included are:

- It presents statistical and visual summaries of data, transforms data so that it can be readily modelled, builds both unsupervised and supervised machine learning models from the data, presents the performance of models graphically, and scores new datasets for deployment into production.
- All of user interactions through the graphical user interface are captured as an R script that can be readily executed in R independently of the Rattle interface.
- Use it as a tool to learn and develop skills in R and then to build initial models in Rattle to then be fine-tuned in R which provides a considerably more powerful option.
- Pricing - Free