



Zohreh Raghebi

- ▶ A software platform provides an integrated environment
 - ▶ Machine learning
 - ▶ Data mining
 - ▶ Text mining
 - ▶ Predictive analytics
 - ▶ Business analytics
- ▶ Supports all steps of the data mining and machine learning procedures
 - ▶ Data loading and transformation (Extract, transform, load (ETL))
 - ▶ Data preprocessing and visualization
 - ▶ Predictive analytics
 - ▶ Statistical modeling
 - ▶ Evaluation, and deployment



RapidMiner

- ▶ RapidMiner uses a client/server model
 - ▶ Offered as Software as a Service or on cloud infrastructures
- ▶ Speed delivery and reduce errors by nearly eliminating the need to write code
- ▶ RapidMiner is written in the Java programming language
- ▶ It can be extended using **R** and **Python** scripts



- ▶ In 2016, **Gartner Research** placed RapidMiner in the leader quadrant of its Magic Quadrant for Advanced Analytics
- ▶ The report highlighted RapidMiner's "Wisdom of Crowds" guidance for recommended next steps in a predictive analytics process
 - ▶ In **addition to overall ease of use**
- ▶ In the **2014 and 2013** annual software poll **KDnuggets** ranked RapidMiner the most popular data analytics software
 - ▶ With the poll's respondents citing the software package as the tool they use

Product Details



Product Details

- Open source.
- Data visualization and analysis
- Machine Learning
- Data Mining, Text Mining.
- Business Intelligence.
- Works on java runtime.
- Available on all major operating systems and platforms

Features

- A visual - code-free - environment, so no programming needed
- Design of analysis processes
- Predictive analytics (with pre-made templates)
- Data loading
- Data transformation
- Data Modelling
- Data visualization (with lots of visualizations)
- Allows you to work with different types and sizes of data sources
- Platform Independence.
- Acts as a powerful scripting language engine along with a graphical user
- Modular operator concept.
- Multi-layered data view.

Success Stories

- CISCO
- PAYPAL
- EBAY
- MIELE
- VOLKSWAGEN





Procedure	R-Programming	RapidMiner	Weka
Partitioning of dataset into training and testing sets.	Pass (but limited partitioning methods)	Pass (but limited partitioning methods)	Pass (but limited partitioning methods)
Descriptor scaling	Pass	Pass	Fail (cannot save parameters for scaling to apply to future datasets)
Descriptor selection	Fail (no wrapper methods)	Pass	Pass (but is not part of KnowledgeFlow)
Parameter optimization of machine learning/statistical methods	Fail (not automatic)	Pass	Fail (not automatic)
Model validation using cross-validation and/or independent validation set	Pass (but limited error measurement methods)	Pass	Pass (but cannot save model so have to rebuild model for every future dataset)



Tutorials: What is what in RapidMiner Studio

- ▶ <http://docs.rapidminer.com/studio/getting-started/>

