



Open Source Analytics Trends World-Wide and at Hungarian Enterprises

Ralf Klinkenberg

RapidMiner

formerly Rapid-I

www.RapidMiner.com

S. Encsy Mátyás

United Consult Ltd.

Hungarian RapidMiner Platinum Partner

www.United-Consult.hu

Budapest BI Forum

November 2013



Open Source Analytics Overview

- Current Trends, Challenges, and Solutions
- Integrate Data from Many Diverse Sources
- Integrate Multiple Solutions and Deploy Automatically
- Analytics from Classical Business Intelligence to Predictive Analytics
- Big Data and Big Data Analytics Solutions
- Making Analytics Easy to Use for Business Users
- RapidMiner: Key Facts, Solutions, Customer Successes
- United Consult: Analytics Expertise in Hungary



OVERVIEW: **OPEN SOURCE ANALYTICS TRENDS**

Trends, challenges, and innovative solutions.

Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Classic: structured data:
 - SQL databases: Oracle, IBM DB2, MS SQL Server, MySQL, PostgreSQL, Ingres, Actian, Teradata ...
 - Flat files like Excel, CSV, SAS, SPSS ...



Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Classic: structured data:
 - SQL databases: Oracle, IBM DB2, MS SQL Server, MySQL, PostgreSQL, Ingres, Actian, Teradata ...
 - Flat files like Excel, CSV, SAS, SPSS ...
 - New: unstructured data:
 - Text documents, web pages, social media, web services ...
 - Audio, images, multi-media data ...



Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Classic: structured data:
 - SQL databases: Oracle, IBM DB2, MS SQL Server, MySQL, PostgreSQL, Ingres, Actian, Teradata ...
 - Flat files like Excel, CSV, SAS, SPSS ...
 - New: unstructured data:
 - Text documents, web pages, social media, web services ...
 - Audio, images, multi-media data ...
 - Diverse data sources need to be integrated
 - Business applications and IT systems: CRM, ERP, Web Shops, etc.
 - Analytics and data mining engines: RapidMiner, R, Weka, MatLab/Octave, etc.
 - External Data: Web, Social Media, News, Data Providers, etc.
 - New internal data sources: Sensor Data, Data Streams, etc.



Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Classic: structured data: SQL databases and flat files like Excel, CSV, SAS ...
 - New: unstructured data: text documents, web pages, audio, images ...
 - Diverse data sources need to be integrated
 - Business applications and IT systems: CRM, ERP, Web Shops, etc.
 - Analytics and data mining engines: RapidMiner, R, Weka, MatLab/Octave, etc.
 - External Data: Web, Social Media, News, Data Providers, etc.
 - New internal data sources: Sensor Data, Data Streams, etc.
 - Big Data: very large amounts of data need to be analyzed
 - The amount of data often cannot be stored on a single computer or in a single database anymore => distributed data storage and processing: Hadoop, Hive, Mahout, etc.



Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Classic: structured data: SQL databases and flat files like Excel, CSV, SAS ...
 - New: unstructured data: text documents, web pages, audio, images ...
 - Diverse data sources need to be integrated
 - Business applications and IT systems: CRM, ERP, Web Shops, etc.
 - Analytics and data mining engines: RapidMiner, R, Weka, MatLab/Octave, etc.
 - External Data: Web, Social Media, News, Data Providers, etc.
 - New internal data sources: Sensor Data, Data Streams, etc.
 - Big Data: very large amounts of data need to be analyzed
 - The amount of data often cannot be stored on a single computer or in a single database anymore => distributed data storage and processing: Hadoop, Hive, Mahout, etc.

=> RapidMiner can handle, integrate, and analyze all these data sources and systems



Analytics Trends and Challenges

- More and more data becomes available for analysis
 - Structured and unstructured data
 - Diverse data sources need to be integrated
 - Big Data: very large amounts of data need to be analyzed
- The tasks become more demanding and complex
 - Business users need insights for their decisions
 - More business processes need to be automated
 - Decisions need to be automated or supported in near real-time
 - Reporting the past is not enough, foresight is needed
 - Self-service BI requested => ease of use essential



Predictions and Foresight a Must

- Analytics from Classical Business Intelligence to Predictive Analytics:
 - Classical BI: Report, what has happened in the past



Predictions and Foresight a Must

- Analytics from Classical Business Intelligence to Predictive Analytics:
 - Classical BI: Report, what has happened in the past
 - Predictive Analytics: Predict what will happen in the future:
 - Which customers are most likely to buy or to churn?
 - Which product or service is a customer most likely to buy?
 - Which machine is likely to fail and why and how can I prevent it?



Predictions and Foresight a Must

- Analytics from Classical Business Intelligence to Predictive Analytics:
 - Classical BI: Report, what has happened in the past
 - Predictive Analytics: Predict what will happen in the future:
 - Which customers are most likely to buy or to churn?
 - Which product or service is a customer most likely to buy?
 - Which machine is likely to fail and why and how can I prevent it?
 - From Insight to Action: Decision Support or Automation:
 - Automated Product Recommendation Systems in Online Shops
 - Marketing Campaign Optimization, Cross- & Up-Selling, Churn Prevention
 - Manufacturing Machine Failure Prediction and Prevention



Predictions and Foresight a Must

- Analytics from Classical Business Intelligence to Predictive Analytics:
 - Classical BI: Report, what has happened in the past
 - Predictive Analytics: Predict what will happen in the future:
 - Which customers are most likely to buy or to churn?
 - Which product or service is a customer most likely to buy?
 - Which machine is likely to fail and why and how can I prevent it?
 - From Insight to Action: Decision Support or Automation:
 - Automated Product Recommendation Systems in Online Shops
 - Marketing Campaign Optimization, Cross- & Up-Selling, Churn Prevention
 - Manufacturing Machine Failure Prediction and Prevention
 - Actionable Insights for a Predictive Company: enable a faster, more agile, more effective, and more successful company:
 - => Act, before a problem occurs
 - => Act, before an opportunity is missed
 - => Reduce risks and costs
 - => Increase sales and profits

Big Data, Analytics, and Scalability

- Big Data Challenge and Big Data Analytics Solution
 - Some data sets too large to fit into one database or server
 - => distributed data storage and processing
 - => RapidMiner seamlessly integrates Hadoop, Hive, Radoop, etc.



Big Data, Analytics, and Scalability

- Big Data Challenge and Big Data Analytics Solution
 - Some data sets too large to fit into one database or server
 - => distributed data storage and processing
 - => RapidMiner seamlessly integrates Hadoop, Hive, Radoop, etc.
 - RapidMiner offers flexible options for scaling to large data sets:
 - In-memory analytics
 - In-database analytics
 - Hadoop-based analytics
 - => you can choose the best approach for your application and data



Big Data, Analytics, and Scalability

- Big Data Challenge and Big Data Analytics Solution
 - Some data sets too large to fit into one database or server
 - => distributed data storage and processing
 - => RapidMiner seamlessly integrates Hadoop, Hive, Radoop, etc.
 - RapidMiner offers flexible options for scaling to large data sets:
 - In-memory analytics
 - In-database analytics
 - Hadoop-based analytics
 - => you can choose the best approach for your application and data
 - Seamless integration of RapidMiner, Hadoop, Hive, Radoop, etc.
 - => Big Data Analytics Made Easy



Big Data, Analytics, and Scalability

- Big Data Challenge and Big Data Analytics Solution
 - Some data sets too large to fit into one database or server
 - => distributed data storage and processing
 - => RapidMiner seamlessly integrates Hadoop, Hive, Radoop, etc.
 - RapidMiner offers flexible options for scaling to large data sets:
 - In-memory analytics
 - In-database analytics
 - Hadoop-based analytics
 - => you can choose the best approach for your application and data
 - Seamless integration of RapidMiner, Hadoop, Hive, Radoop, etc.
 - => Big Data Analytics Made Easy
- Big Data: How big is big? When to use which approach?
<http://rapidminer.com/2013/07/31/rapidminer-big-data-how-big-is-big/>





RAPIDMINER: SOLUTIONS AND CUSTOMER STORIES

RapidMiner delivers the power of predictive analytics to business users. No programming required.

RapidMiner



RapidMiner Predictive Analytics: easy to use, deploy, and integrate – no programming required

Company Overview

- Provides software, solutions, and services for predictive analytics, data mining, and text mining
- Flagship product **RapidMiner** is the world-leading open-source system for knowledge discovery, data mining and sentiment analysis
 - Easy to use, blazing fast, and very easy to integrate with any IT infrastructure
 - Support from a thriving community of contributors creating new extensions and applications
 - Processes designed in RapidMiner can be one-click deployed to **RapidMiner Server**, and instantly used for interactive reports, dashboards & web services
- A unique **Marketplace** for independent developers to publish and monetize their innovative extensions

Key Business Metrics

- Founded as Rapid-I in 2006 in Dortmund, Germany; U.S. office opened in 2012
- 30 employees worldwide
- 10,000+ downloads/mo.
- 35,000+ active deployments, with 200,000+ users
- 1000+ trained experts
- 500+ customers in over 50 countries
- 35+ channel partners
- 100+ active developers

RapidMiner Visual Environment

With access to over 1,500 operators, the Java-based visual environment of RapidMiner allows for rapid process development

The screenshot displays the RapidMiner interface with a process flow diagram in the center. The flow starts with a 'Retrieve' operator, followed by a 'Decision Tree' operator, then a 'Weight by Reli...' operator, a 'Clustering' operator, and finally a 'Correlation M...' operator. The 'Decision Tree' operator is selected, and its parameters are visible on the right side of the window. The parameters include: criterion (gain_ratio), minimal size for split (4), minimal leaf size (2), minimal gain (0.1), maximal depth (20), confidence (0.25), and number of preprun... (3). There are also checkboxes for 'no pre pruning' and 'no pruning'. The interface includes a menu bar (File, Edit, Process, Tools, View, Help), a toolbar, and a sidebar with 'Operators' and 'Repositories' sections. The 'Operators' section is expanded to show various categories like Process Control, Utility, Repository Access, Import, Export, Data Transformation, Modeling, Evaluation, Text Processing, Reporting, and Series.

RapidMiner Process Design Environment



RapidMiner Server



Representative RapidMiner Server Dashboard

The RapidMiner Server provides enterprise-wide process development and process to web-service conversion with dynamic dashboards and data visualizations.



RapidMiner Data Visualization

Supported by strong analytics capabilities, the RapidMiner suite of applications let users visualize their data and models in a myriad of dynamic, customizable outputs.





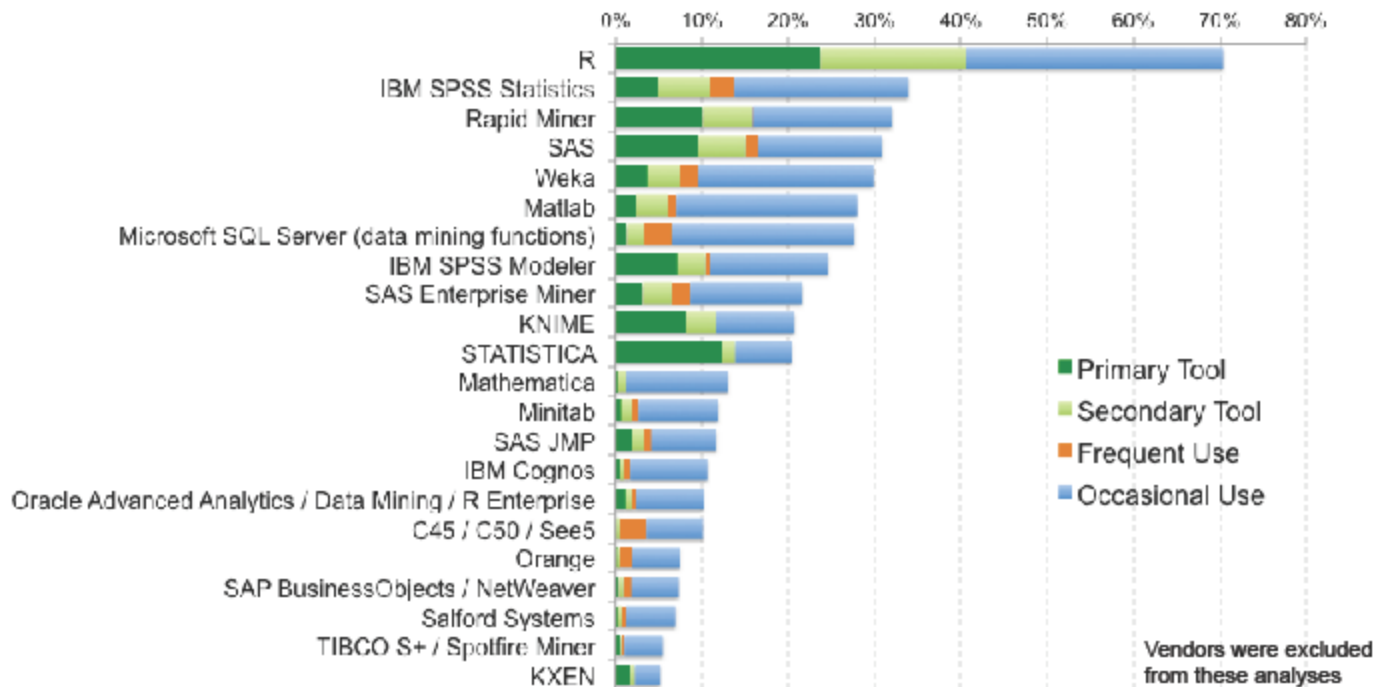
THE #1 PREDICTIVE ANALYTICS SOLUTION

RapidMiner is unmatched in the industry.

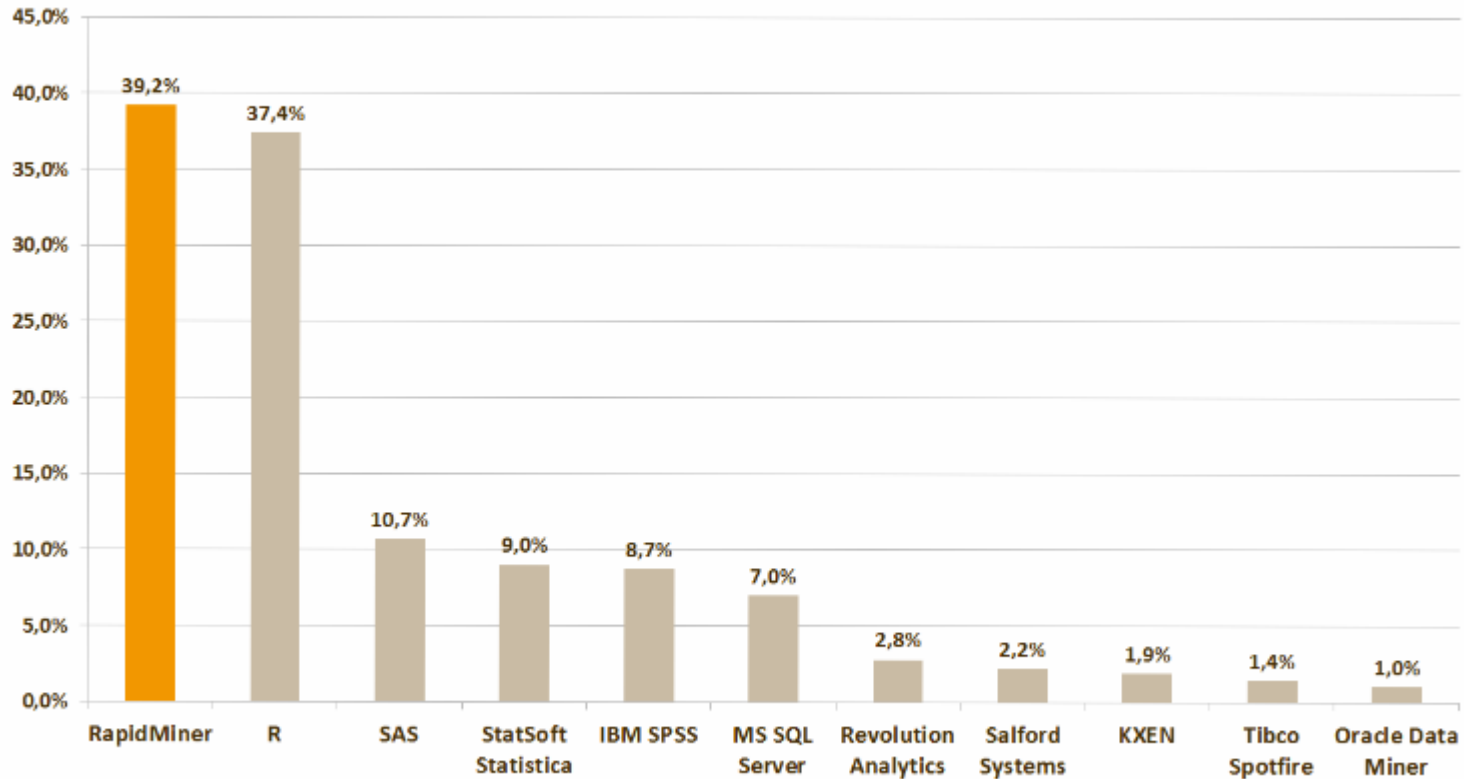
Highly-Ranked Vendor

The Tools We're Using – Ranked by Usage

- The average data miner reports using 5 software tools
- R is the tool used by the most data miners (70%)
- IBM SPSS Statistics, Rapid Miner, and SAS also have many users



RapidMiner Leads the Market



Source: KD Nuggets 2013 - Predictive Analytics / Data Mining Software used in past 12 months for a real project



Here's Why

- **Open Source Community & Marketplace** – Crowd-sourced innovation, quality assurance, market awareness.
- **Fully-integrated Platform** – Integrated, process-based business analytics platform with focus on predictive analytics.
- **No Programming Required** – Easy-to-use, low maintenance costs, standard platform for business analysts.
- **Advanced Analytics at Every Scale** – In-memory, in-database and in-Hadoop analytics offer best option for every size of database.
- **Connectivity** – More than 60 connectors (incl. SAP & Hadoop), allowing easy access to structured and unstructured data.



The Numbers

- 3,000,000 total downloads
- 200,000 RapidMiner active users
- 10,000 monthly product downloads
- 95,000 newsletter subscribers
- 1000 education customers
- 500 customers
- 100 certified analysts
- 35 partners world-wide
- Platinum Partner United Consult with 130 consultants in Hungary



A SELECTION OF **OUR CUSTOMERS**

10,000+ Downloads per Month

SELECT LIST OF RECIPIENT ORGANIZATIONS

Technology

- intel, IBM, Alcatel-Lucent, EMC², MOTOROLA, hp, xerox, DELL, CSC, HUAWEI, CISCO, SIEMENS, chipworks, AMD, CITRIX, Bazaarvoice

Pharma & Healthcare

- CENTENE, MED SOLUTIONS, Pfizer, Medtronic, SANOFI, MERCK, AMGEN

Oil & Gas, Chemicals

- Chevron, ExxonMobil, PRAXAIR, Schlumberger

Government & Defense

- LAWRENCE LIVERMORE NATIONAL LABORATORY, NASA, OAK RIDGE National Laboratory, Sandia National Laboratories, NATIONAL INSTITUTES OF HEALTH, FDA, Pacific Northwest NATIONAL LABORATORY, IRS, NSI, Federal Aviation Administration, USGS, SOCOM, U.S. ARMY, U.S. AIR FORCE, MITRE, Commonwealth of Massachusetts

Manufacturing

- Honeywell, DIEBOLD, CAT

Consumer Products

- P&G, Kimberly-Clark

Aerospace

- NORTHROP GRUMMAN, Raytheon, BOEING

Consulting

- Booz | Allen | Hamilton, SAIC, Deloitte, SLALOM CONSULTING, accenture, The Advisory Board Company, ERNST & YOUNG, McKinsey & Company

Business Services

- LexisNexis, ACXION, FedEx, Pitney Bowes

Software & Analytics

- clarabridge, sas, Ipsos, DML, SAGENCE GROUP, seeker, Microsoft, TERADATA, ORACLE, marc

Financial Services

- citi, GEICO, Bank of America, AON, PayPal, JPMorgan, WELLS FARGO, GREAT AMERICAN INSURANCE GROUP, SmartSoft

Entertainment

- EA, WB, Disney, ACTIVISION, Rhapsody

Academia

- KAPLAN, UIC, Virginia Tech, WPI, UCONN, MIT, HARVARD UNIVERSITY, STANFORD UNIVERSITY, CLEVELAND STATE UNIVERSITY

Select Customer Stories

Sanofi



Who > Global leader in pharmaceutical industry with 100,000 employees in 100 countries

Solution > Prediction of sales volumes; CRM optimization; social media monitoring and sentiment analysis for identifying undesired drug side effects and automated alerts; network analysis and visualization for identification of most relevant target persons or computer network bottlenecks

Telenor Group



Who > Leading Financial Times Global 500 mobile operator with close to 150 million mobile subscriptions in 29 markets

Solution > CRM applications including optimization of direct marketing campaigns, automated generation of product recommendations for cross-selling and up-selling, customer churn prevention, and fraud detection

GfK



Who > One of the world's largest market research firms with 11,000 employees researching how people live, think and consume. Provides clients in 100+ countries with customer insight data, including services for the consumer goods, media, pharmaceutical, and services markets

Solution > Sentiment Analysis of online text sources, including social media and other user generated content

Select Customer Stories (continued)

mobikom
austria



Who > Leading Austrian mobile phone service and internet provider

Solution > Optimize the customer support by automatically categorizing incoming e-mails by content and forwarding them to the appropriate department or person in charge, thereby reducing response time and cost and increasing customer satisfaction

PayPal



Who > World leading online payment services provider

Solution > Customer feedback and voice of the customer analysis, churn prediction and prevention, text mining, automated text categorization, and sentiment analysis to customer support and satisfaction and prevent customer churn

SmartSoft  SmartSoft
Banking Risk Solutions

Who > Provider of solutions for preventing fraud, money laundering, and risks in financial institutions

Solution > Integration of Rapid-I's predictive analytics engine into their solutions for fraud detection and fraud prevention for the financial and telecom sectors

Use Cases

RapidMiner and RapidMiner Server have been implemented and integrated to address key business issues

- Marketing response rates
- Customer segmentation
- Loyalty, retention, and attrition
- Credit scoring
- Fraud models
- Resource planning
- Asset maintenance, failure prediction and prevention
- Life sciences
- Social media monitoring

OUR PREDICTIONS: **VISION & ROADMAP**

Transform Market Penetration to Market Leadership

RapidMiner v5.3

UPGRADED JANUARY, 2013

RapidMiner v5.3 New Features

- More powerful data analysis operators
 - Access to popular data sources including Microsoft Excel 2007 and SAS
 - Easier to use: new documentation for the 200 most important operations; new wizards, e.g. for defining execution schedules on the server or for complex regular expressions
 - More powerful analyses: more aggregation functions and 20 new functions for analysis and data handling
 - Improved robustness: safe-mode
 - In addition, more than 100 small improvements and bug fixes
- Integrated with the Marketplace
 - Users now have access to a central storehouse of product extensions
 - Users and developers now have the option to monetize their extensions: they can now create not only free, open source extensions for RapidMiner but also commercial extensions
 - Supports in-program browsing of the Marketplace, defining favorites among the extensions, as well as online extension purchasing and direct download into the software

RapidMiner v6.0

TO BE RELEASED IN NOVEMBER 2013

RapidMiner v6.0 New Features

- Pre-built templates for self-service predictive analytics
- New Welcome screen makes it easier and faster to get started
- New Result view, with call to action(s)
- New Statistics and Charts views
- Improved Tours and Tutorials
- Other improvements and bug fixes



RapidMiner Accelerating the Growth

- RapidMiner (formely Rapid-I) has been growing and profitable for many years
- US\$ 5M Series A investment secured in October 2013
- Rebranding and new web site: November 4th, 2013
- RapidMiner already is one of the most powerful and flexible predictive analytics solutions in the market
- Now: Focus of on ease of use and self-service predictive analytics: application templates, tutorials, etc.
- Investement for accelerating the development of our innovative solutions and for further growth

Our Vision

- To become the de-facto industry standard for analytics built on an open stack.
- To help business analysts fully harness the power of their data to predict the future.
- To make it easier for organizations to extract value from their data.



RapidMiner 6 Preview: Ease of Use

Use Case Templates

Template Data Results

1 Select Template



Direct Marketing

What was once a postal delivery channel has now expanded to include text messages, emails, online ads and other digital advertising. Businesses must account for spending across these platforms and provide analytics to support ROI metrics. RapidMiner's Direct Marketing template allows businesses and nonprofits organizations a way to track and spend wisely.



Churn

The cost of retaining an existing customer is far less than acquiring a new one. RapidMiner's Churn Analysis template sifts through your customer data to identify which customers are most likely to switch to a competitor and why, so that you can implement a targeted retention campaign in a timely fashion.



Preventive Maintenance

**Application Use Cases
for Business Users**



Sentiment Analysis

**Ease of Use and First
Predictive Analytics
Results within 2-5 min.**



RapidMiner 6 Preview: Few Clicks...

**Step 1:
Select your use case.**

Use Case Templates

Template Data Results Direct Marketing

2 Load Data

Drop your data
HERE
Use Excel, CSV, or repository

OR

Use demo data Open file

Direct Marketing Data

To run a direct marketing campaign you need data about your customers. In your data table, use one row per customer. For some of them you need to know whether or not they responded to previous campaign (or an evaluation sample if a previous campaign does not exist). Name that column **response**, using the values *yes* and *no*. For the remaining customers, you can leave the column blank.

Customer Characteristics

Besides this, you can use as much information as you can get to characterize your customers. Here are some ideas:

- **Demographic data:** *Age* and *gender* are often the most relevant influence factors. From the zip code you can infer other relevant properties (rural/urban, purchasing power, etc.)
- **Web logs:** If you have a Web site, you can extract information on the number of *visits* or *purchases in the past*. Browser type and ISP or email providers (free vs premium) may be characteristic properties.
- **Social media:** Link up with Facebook or Twitter to learn about *interests* and *habits* of your customers.

Any of these should be added as a separate column to your data set. Do not use too much redundancy, but try to focus on what you expect to be the best indicators.

Demo Data

If you want to get an impression of how such a data set should look like, click on "Use demo data".

[Tell me more...](#)

Step 2: Select your data set by drag & drop or via file browser.

RapidMiner 6 Preview: Few Clicks...

Use Case Templates

Template Data Results **Direct Marketing**

Your Data

Clear data


age	gender	area	email	mobile	logins 4 weeks	logins 6 months	sales 4 weeks	sales 6 months	sales total	response
76.0	male	urban	free	never	0.0	0.0	0.0	0.0	67.0	no
63.0	male	urban	premium	always	1.0	7.0	0.0	0.0	0.0	no
43.0	male	urban	fmail	never	0.0	0.0	0.0	0.0	0.0	yes
24.0	male	urban	fmail	never	0.0	0.0	0.0	0.0	0.0	no

Your data set is large. We recommend to run it on a sample of the data to test this template first.

Run template on the full data set.

3 Response Column

Please select a column containing the information whether or not a customer responded to your mailing campaign.



no yes

response Positive class:

4 Run Template

Take me to the results

Direct Marketing Data

To run a direct marketing campaign you need data about your customers. In your data table, use one row per customer. For some of them you need to know whether or not they responded to previous campaign (or an evaluation sample if a previous campaign does not exist). Name that column **response**, using the values **yes** and **no**. For the remaining customers, you can leave the column blank.

Customer Characteristics

Besides this, you can use as much information as you can get to characterize your customers. Here are some ideas:

- **Demographic data:** Age and gender are often the most relevant influence factors. From the zip code you can infer other relevant properties (rural/urban, purchasing power, etc.)
- **Web logs:** If you have a Web site, you can extract information on the number of visits or purchases in the past. Browser type and ISP or email providers (free vs premium) may be characteristic properties.
- **Social media:** Link up with Facebook or Twitter to learn about interests and habits of your customers.

Any of these should be added as a separate column to your data set. Do not use too much redundancy, but try to focus on what you expect to be the best indicators.

Demo Data

If you want to get an impression of how such a data set should look like, click on "Use demo data".

[Tell me more...](#)

Step 3:
View your data and select the target variable, for example the customer response in a direct mailing.



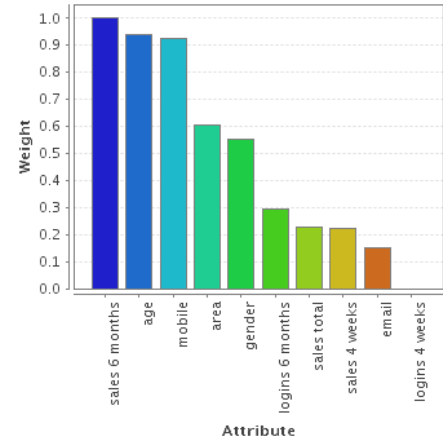
RapidMiner 6 Preview: Quick Results

Use Case Templates

Template Data **Results** Direct Marketing

6 Analytical Results

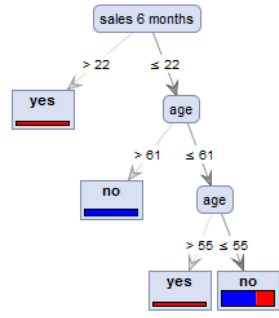
Influence Factors



Attribute	Weight
sales 6 months	1.0
age	0.95
mobile	0.92
area	0.6
gender	0.55
logins 6 months	0.3
sales total	0.25
sales 4 weeks	0.22
email	0.15

These are the most important features that characterize your customers.

Decision Tree



This model describes how to characterize your customers as responders or non-responders.

Direct Marketing Actions

Congratulations, you have successfully run a direct marketing analysis. This dashboard shows what actions to take next.

Influence Factors

The two images at the top help you understand your customers. The bar chart shows which of the properties of your customers have the biggest impact on whether or not they respond to your campaign. The decision tree next to it displays how you can identify potential responders. Follow a path from the root to a leaf and observe the criteria identifying this customer group. The leaf itself indicates whether there are more responders or more non-responders in that group.

Customer Selection

At the center, we have a table of customers that we recommend for addressing in your marketing campaign. These are the customers for which the generated model achieves the highest confidence values.

Sample Evaluation

The predictive model generated by this template utilizes the customer data you provided. It is often costly to obtain such data since evaluation campaigns have to be run, etc. The elements in the bottom row help you to decide whether or not you have collected enough data. The "Learning Curve" on the left visualizes how the prediction quality increases as more and more data samples are added while generating the model. If this line is still rising at the far right, this indicates that the model may still benefit from adding more data.

Show me how it works!
Export HTML

Target Group

Step 4: Run the analytics process.
Step 5: View analytical results, valuable gain insights.

RapidMiner 6 Preview: Deployment

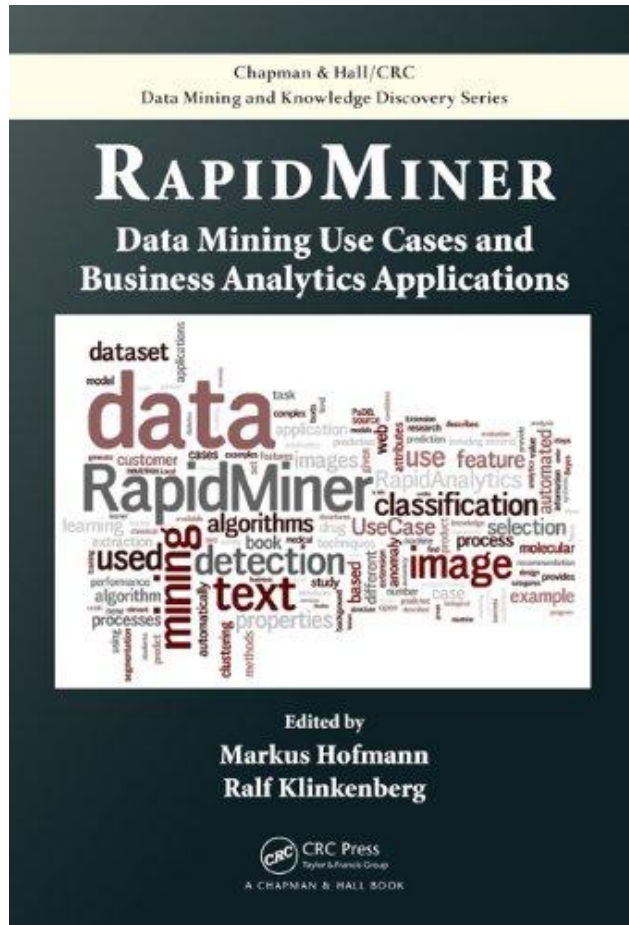
The screenshot displays the RapidMiner 5.3.014 software interface. The main window shows a process design canvas with the following nodes and connections:

- Filter Training** (pink node) receives input from the left and connects to **Multiply (2)** (orange node).
- Multiply (2)** connects to **Default precis...** (orange node).
- Default precis...** connects to **Weights** (orange node).
- Weights** connects to **Decision Tree** (green node).
- Decision Tree** connects to **Performance** (orange node).
- Performance** connects to **Last model** (orange node).
- Last model** connects to **Apply on Data** (orange node).
- Filter Training** also connects to **Filter Apply** (pink node).
- Filter Apply** connects to **Apply on Data**.

The left sidebar shows a tree view of repositories, including 'Import (28)', 'Export (18)', 'Data Transformation (115)', 'Modeling (258)', and 'Evaluation (32)'. The top menu bar includes 'File', 'Edit', 'Process', 'Tools', 'View', and 'Help'. The bottom status bar shows '3 potential problems' and a 'Message' section.

Step 6: Deployment: View the underlying RapidMiner process details, customize if necessary, apply it, integrate it, automate it, deploy it, ... it is that simple.

RapidMiner Application Use Cases



For more application used cases:

New book published end of October 2013:

RapidMiner: Data Mining Use Cases and Business Analytics Applications

in the Chapman & Hall / CRC Data Mining and Knowledge Discovery Series, editors Markus Hofmann and Ralf Klinkenberg, authors from the RapidMiner development team and the RapidMiner community including developers, partners, customers, researchers, business users, covering many diverse use cases from research and industry

Our Vision

- To become the de-facto industry standard for analytics built on an open stack.
- To help business analysts fully harness the power of their data to predict the future.
- To make it easier for organizations to extract value from their data.





OPEN SOURCE ANALYTICS IN
HUNGARY:

UNITED CONSULT

Certified RapidMiner Platinum Partner with highly skilled experts and many satisfied customers.

10 years double-digit growth
130 consultants



UNITED CONSULT



Testing and Quality Assurance

Business Intelligence

Software Development

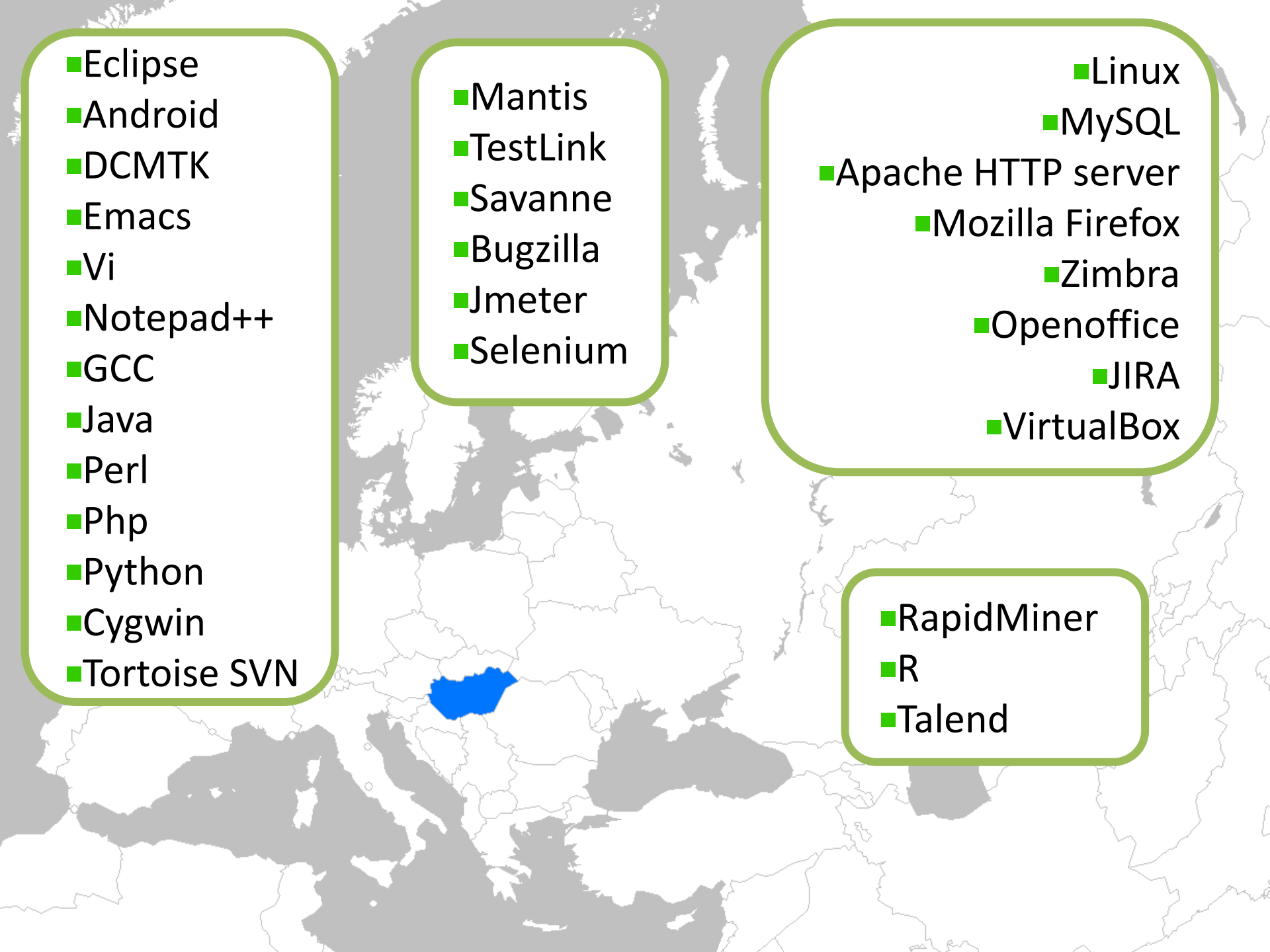
IT Operation and Maintenance



S. ENCSY Mátyás

20

- Business analyst
- Data Scientist
- Architect
- Data modeler
- ETL designer
- Data miner
- Back-end developer
- Front-end developer
- Report developer
- Tester
- Project manager

- 
- Eclipse
 - Android
 - DCMTK
 - Emacs
 - Vi
 - Notepad++
 - GCC
 - Java
 - Perl
 - Php
 - Python
 - Cygwin
 - Tortoise SVN

- Mantis
- TestLink
- Savanne
- Bugzilla
- Jmeter
- Selenium

- Linux
- MySQL
- Apache HTTP server
- Mozilla Firefox
- Zimbra
- Openoffice
- JIRA
- VirtualBox

- RapidMiner
- R
- Talend

Why?

- Cheap initial cost
- Not mature processes

Used as a specialized tool

- Development
- Testing
- Data Mining





UNITED CONSULT

Chosen because

- best in class
- fast growing
- enterprise ready

Platinum partnership

- 1 Certified Rapid-I Master
- 4 Certified Rapid-I Analysts



Other activities

- Official courses
- Education (ELTE, BME, Pannon, etc)
- Hungarian translation



UNITED CONSULT



THE NEXT
DATA
THING

Contact Us



Ralf Klinkenberg

RapidMiner

- Boston, MA, USA
- Sunnyvale, CA, USA
- Dortmund, Germany
- 35+ partners world-wide

www.RapidMiner.com

contact@rapid-i.com

klinkenberg@rapid-i.com

S. Encsy Mátyás

United Consult Ltd.

- RapidMiner Platinum Partner
- Headquarter in Budapest
- 130 Consultants in Hungary
- Local support and expertise

www.United-Consult.hu

Matyas.Encsy@united-consult.hu

