



RAPIDMINER:

BIG DATA ANALYTICS

RapidMiner delivers the power of Predictive Analytics to Business Users.
No programming required. From Big Data to Actionable Insights.

Ralf Klinkenberg

RapidMiner (formerly Rapid-I)

www.RapidMiner.com

February 2014



RAPIDMINER:

BIG DATA ANALYTICS

**Predictive Big Data Analytics Anwendungen von
RapidMiner in Produktion und Produktionsplanung**

Ralf Klinkenberg

RapidMiner (formerly Rapid-I)

www.RapidMiner.com

February 2014



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 2007 in Dortmund, Germany; U.S. office opened in 2012
- 40 employees worldwide
- 10,000+ downloads/mo.
- 35,000+ active deployments, with 200,000+ active users
- 1000+ trained experts
- 500+ customers in over 50 countries
- 35+ channel partners
- 100+ active developers



Ralf Klinkenberg



Ralf Klinkenberg

General Manager

RapidMiner GmbH

rklinkenberg@rapidminer.com

+49-231-425-786-92

- M. Sc. in Computer Science from Missouri University of Science and Technology (MST), USA: focus on machine learning and artificial intelligence
- M. Sc. in Computer Science from Technical University of Dortmund, Germany: focus on machine learning, data mining, and text mining
- Data mining and text mining research at MST and TU Dortmund
- **2001** Co-Founder of the **RapidMiner** Open Source Data Mining Software Project (formerly YALE) with Dr. Ingo Mierswa and Dr. Simon Fischer
- **2007** Co-Founder of the company Rapid-I GmbH with Dr. Ingo Mierswa
- **2013** Rapid-I GmbH renamed to RapidMiner GmbH
- **2014** Gartner positions RapidMiner in Leaders Quadrant for Advanced Analytics Platforms
- 22 years of experience in machine learning, data mining, text mining, web mining, predictive analytics, and their applications
- Numerous analytics projects in telecom, automotive, aviation, manufacturing, retail, internet, market research, and government

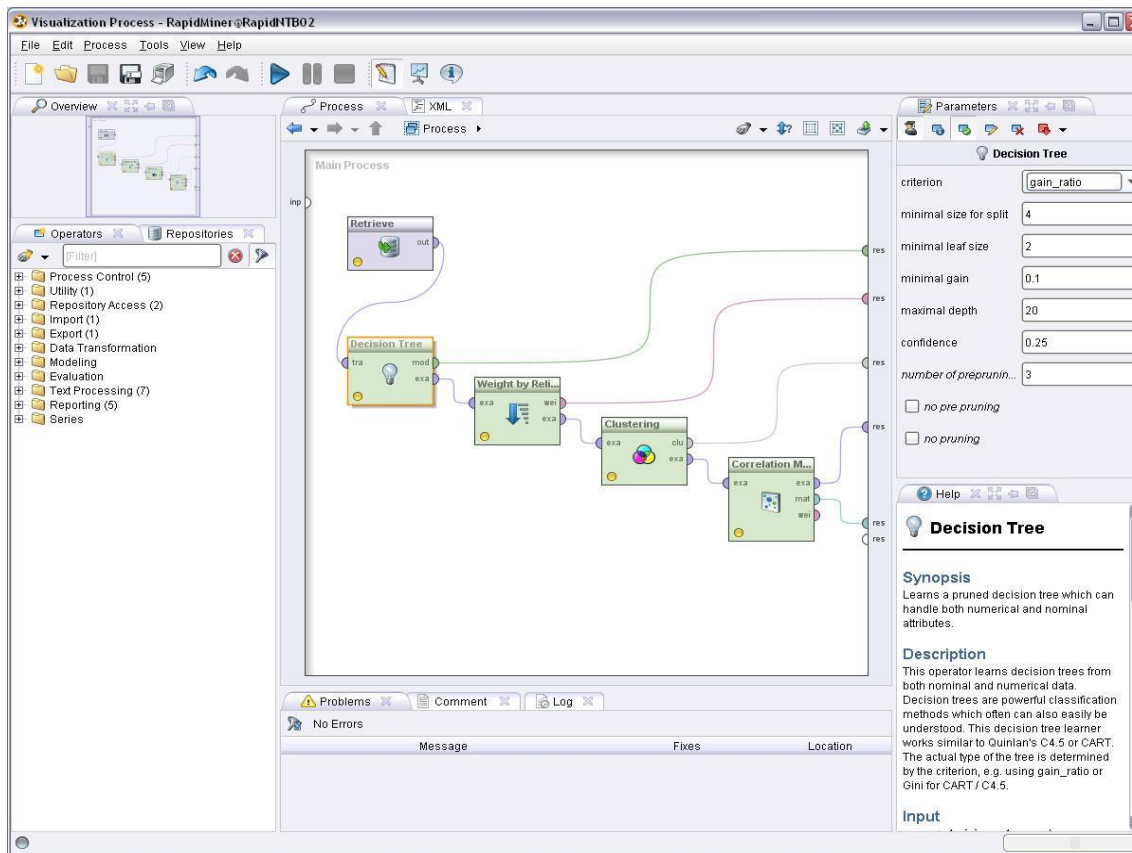


RAPIDMINER: SOLUTIONS AND APPLICATIONS

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

RapidMiner Visual Environment

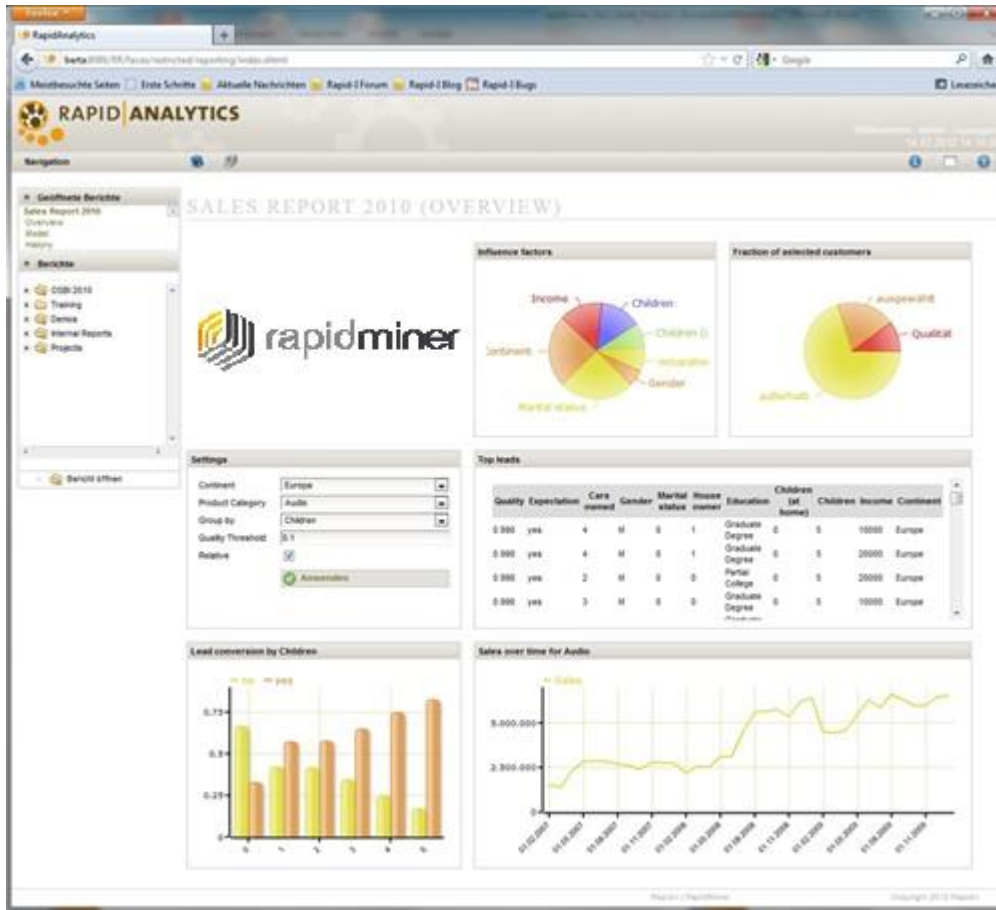
With access to over 1,500 operators, the Java-based visual environment of RapidMiner allows for rapid data mining process development



**RapidMiner
Studio:
Visual 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.



RapidMiner 6: Ease of Use

Application Wizard

Application Data Results

Direct Marketing

1 Select Application



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. The RapidMiner Direct Marketing Wizard allows businesses to invest only in marketing actions with the highest conversion rates and helps reduce costs by improving targeting.



Predictive Maintenance

Product manufacturing is unlike any other business and sudden malfunctions can stop your business on a dime, resulting in lost revenue and production concerns. The RapidMiner Predictive Maintenance Wizard transforms the unknown into the predictable, reducing both unnecessary maintenance costs and unexpected failures. Start increasing your planned maintenance and stop reacting to the unplanned disasters.



Churn

The cost of retaining an existing customer is far less than acquiring a new one. RapidMiner's Churn Analysis Wizard 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.



Sentiment Analysis

Markets are conversations, and social media is an endless source of conversations where products, services and brands are both praised and criticized in real time. Properly analyzed, these conversations can give you direct insight into your customers' thinking. Plug RapidMiner's Sentiment Analysis Wizard into the social media stream and see what customers are saying about your business.

Application Use Cases
for Business Users

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

RapidMiner 6: 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: 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.

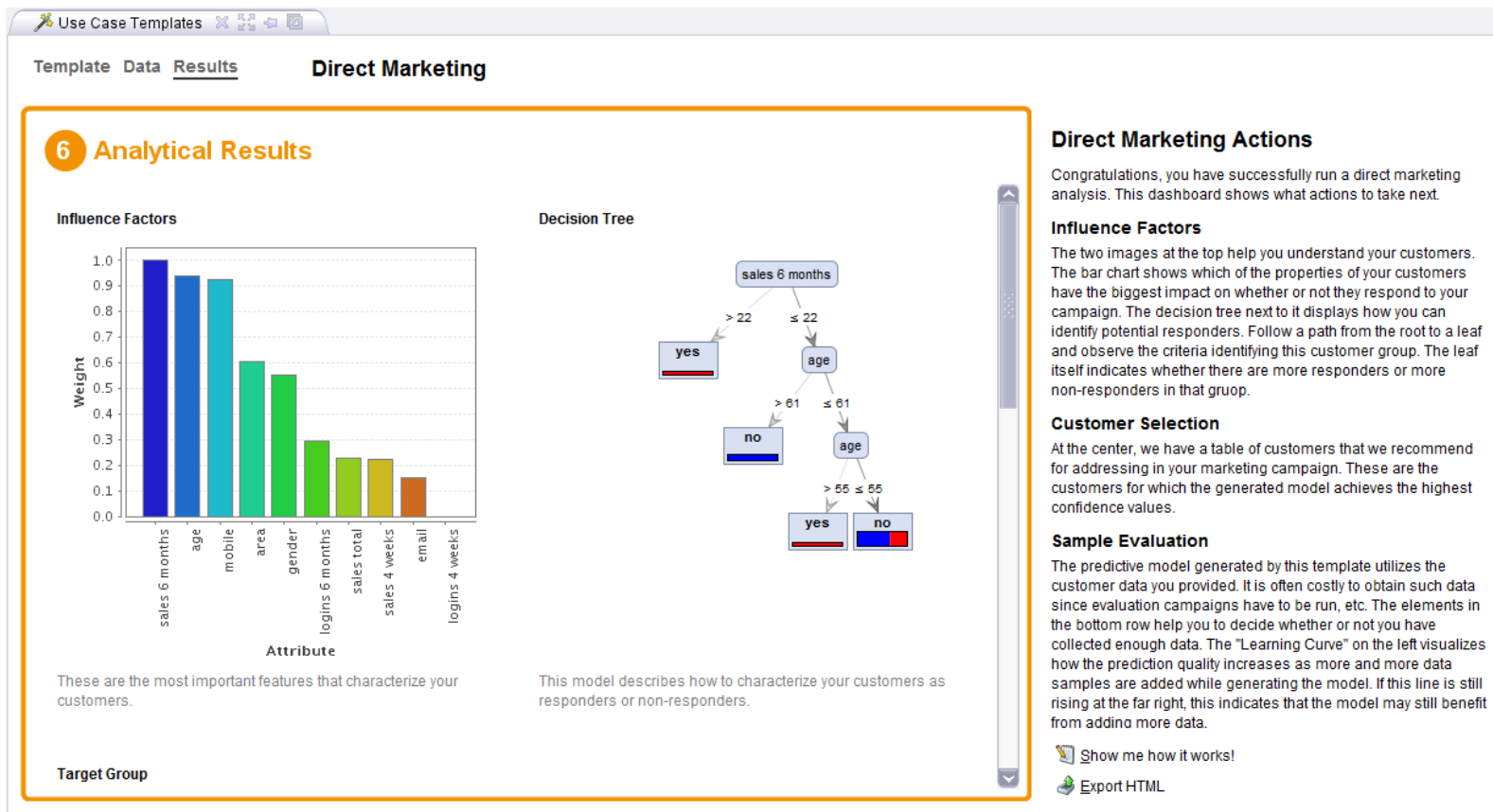
Demo Data

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

[Tell me more...](#)

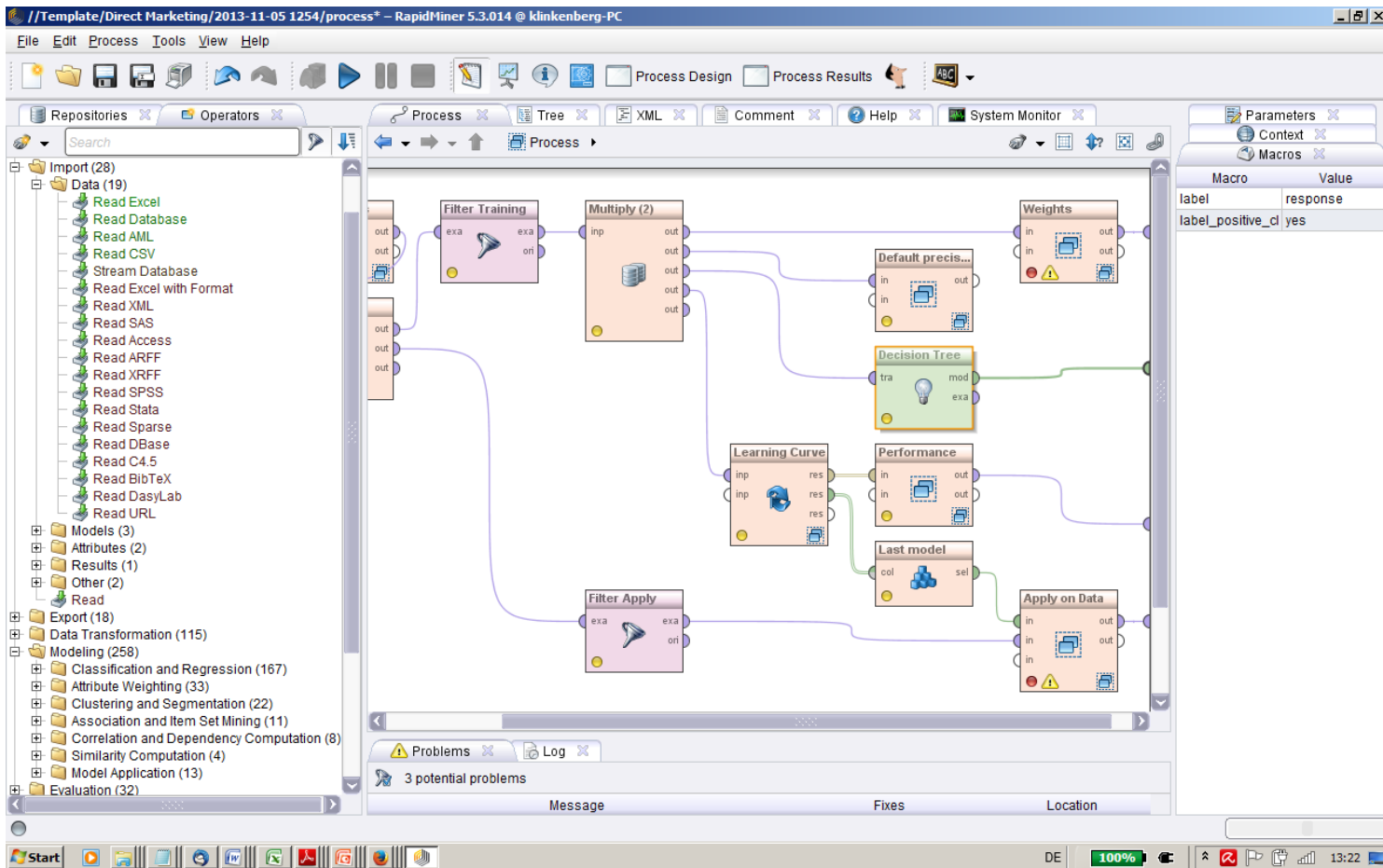
Step 3:
View your data and select the target variable, for example the customer response in a direct mailing.

RapidMiner 6: Quick Results



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

RapidMiner 6: Deployment



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 Integration Options

RapidMiner offers a broad range of integration options:

- RapidMiner Engine: Java API (“RapidMiner inside”)
- RapidMiner Server: Web Services
- RapidMiner Server: Reportlets (“seamless into Cognos & Co.”)
- RapidMiner Server: Process Scheduler & Trigger
- Data Exchange: via Java API, SQL Databases, Files, NoSQL Databases, Hadoop, Hive, Web Services, etc.
- Command line calls (“the old way”, also still possible)
- Bundling



RapidMiner OEM Partnership

Advantages of a RapidMiner OEM Partnership:

- RapidMiner as proven powerful and flexible engine for your solutions: meet all predictive analytics, data mining, text mining, web mining, image mining, audio mining needs
- RapidMiner Java API for seamless integration (“RapidMiner inside”) and easy extension and customization
- RapidMiner web services for easy IT integration
- RapidMiner Marketplace: RapidMiner Extensions by 3rd parties
- World-wide community of RapidMiner experts
- Joint business plan and joint business goals



<http://rapidminer.com/download-rapidminer/>





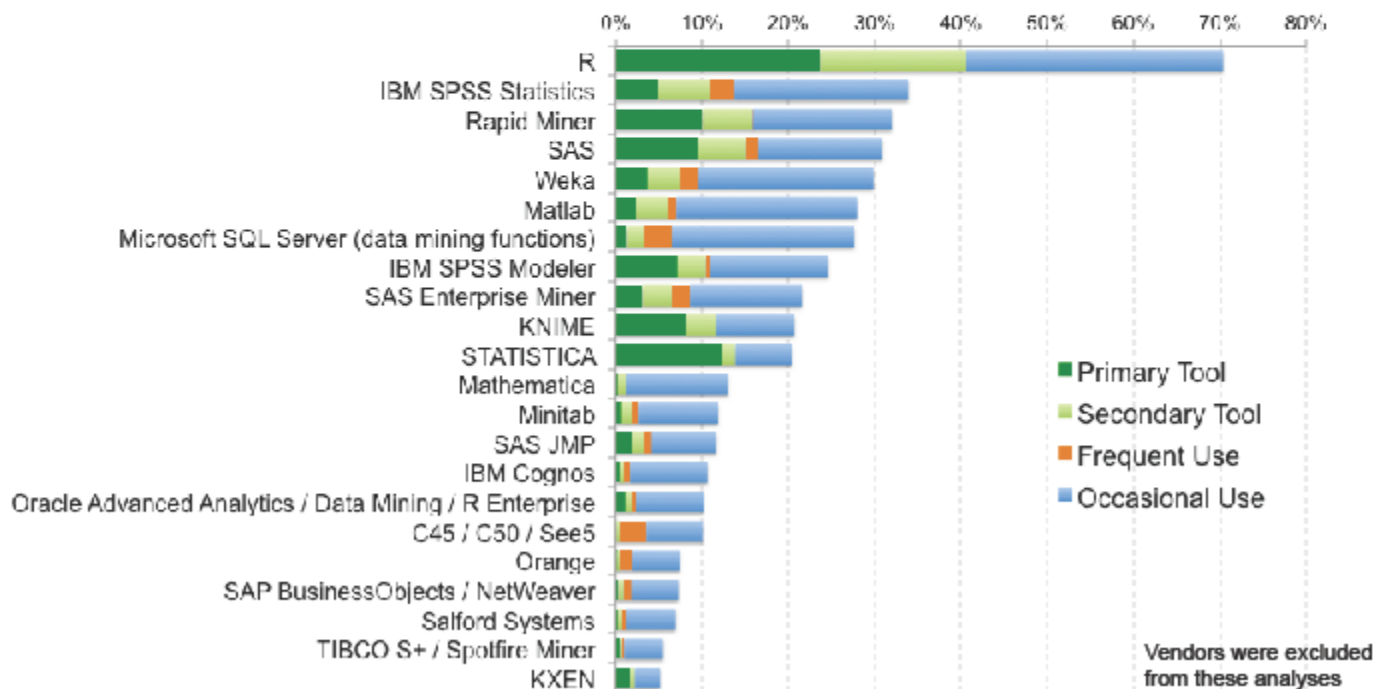
THE #1 PREDICTIVE ANALYTICS SOLUTION

RapidMiner is unmatched in the industry.

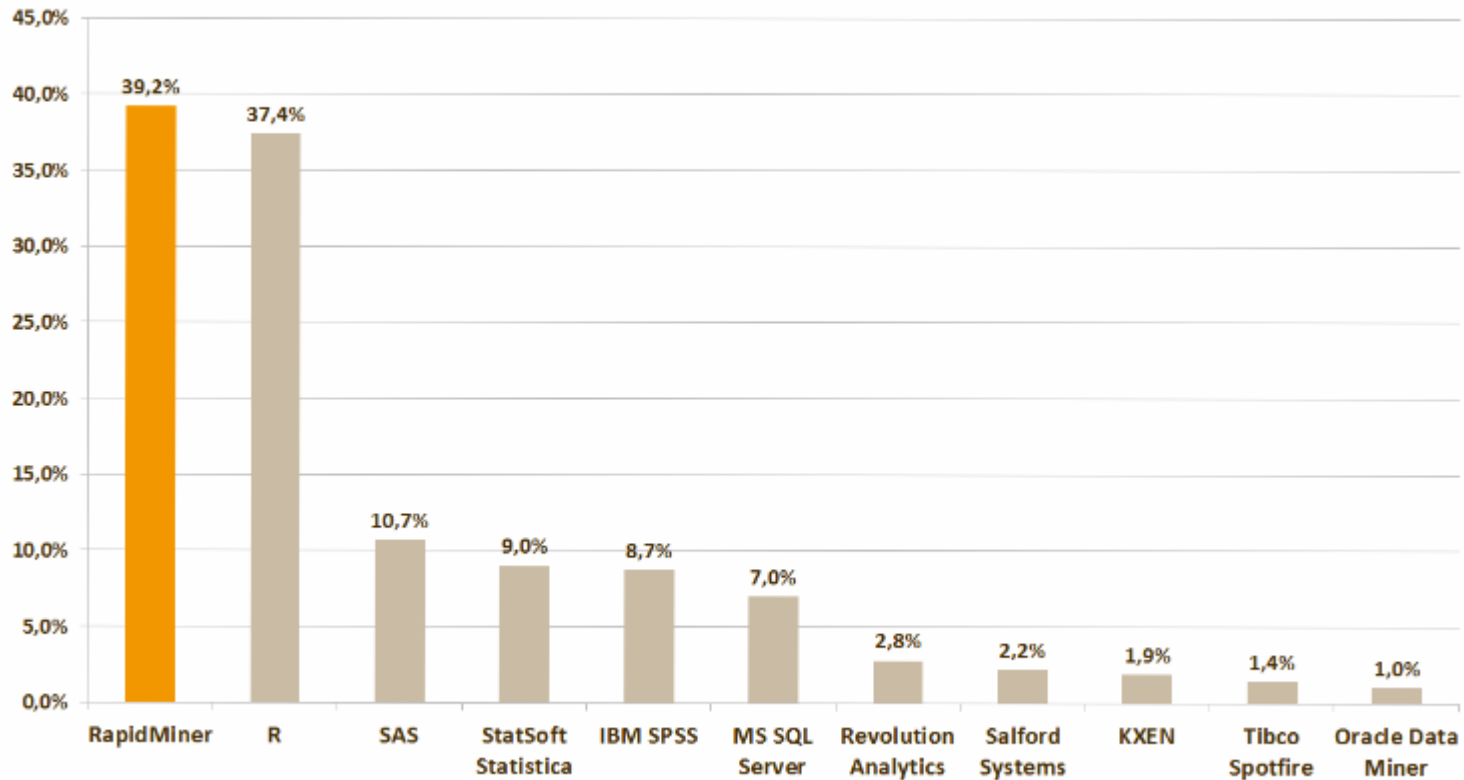
Highly-Ranked Vendor

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

Gartner Magic Quadrant for Advanced Analytics Platforms positions RapidMiner in Leader Quadrant

February 2014

Source:
Gartner "Magic Quadrant for Advanced Analytics Platforms," by Gareth Herschel, Alexander Linden, and Lisa Kart, February 19, 2014.

Free download of the complete Gartner Report:
www.rapidminer.com/gartner2014/



Analysts and Media about RapidMiner

- IT-Director.com: **RapidMiner – A Potential Game Changer:**
<http://www.it-director.com/content.php?cid=14555>
<http://rapidminer.com/news-posts/it-director-com/>
- QuinnSight Research: **RapidMiner Hidden Giant of Analytics?**
<http://quinn sight.com/2013/06/17/is-rapid-i-the-hidden-giant-of-analytics/>
- KDnuggets: Analytics Marketplaces: **The Next Big Think for Big Data:**
<http://rapidminer.com/news-posts/kdnuggets-review/>
- Usama Fayyad, Chief Information Officer, Barclays Bank, previously Yahoo's Chief Data Officer and Head of Data Mining Research at Microsoft Research:
Why RapidMiner? -- The Platform of Choice for anyone analyzing Big Data:
<http://www.kdnuggets.com/2013/12/why-rapidminer-by-usama-fayyad-top-data-scientist-entrepreneur.html>
- Gartner: RapidMiner in Gartner's **Who is Who in Advanced Analytics:**
<https://www.gartner.com/doc/2628123/whos-advanced-analytics>
- Decision Stats: **RapidMiner takes Analytics to the Next Level:**
<http://rapidminer.com/news-posts/decision-stats-rapidminer/>
- KDnuggets: RapidMiner and R are the **most widely used data mining solutions:**
<http://www.kdnuggets.com/2013/06/kdnuggets-annual-software-poll-rapidminer-r-vie-for-first-place.html>
- Rexer Analytics: RapidMiner among **leaders with regards to user satisfaction:**
<http://www.kdnuggets.com/2013/10/rexer-analytics-2013-data-miner-survey-highlights.html>



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



A SELECTION OF **OUR CUSTOMERS & CUSTOMER STORIES**

Select Customers



Select Customer Stories



Who > Global leader in the pharmaceutical industry with nearly 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



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



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)



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



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 in more than 60 languages to improve customer support and satisfaction and prevent customer churn



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

Solution > Integration of RapidMiner predictive analytics engine into their solutions for fraud detection and fraud prevention for the financial and telecom sectors

Select Use Cases

Financial Services	Telecommunications	Manufacturing	Electronics
<ul style="list-style-type: none">• Fraud detection• Marketing campaign optimization• Automated product recommendation• Product affinity scoring• Up- & cross-selling• Churn prevention• Sentiment analysis• Trading• Time series analysis and forecasting• Risk management• Credit risk scoring• Credit default prediction• Customer profitability	<ul style="list-style-type: none">• Marketing campaign optimization• Churn prevention• Resource planning• Report creation• Network analysis• Customer service process automation• Automated e-mail routing• Automated online market research• Social Media Monitoring• Sentiment Analysis	<ul style="list-style-type: none">• Predictive maintenance• Machine failure prediction and prevention• Time series analysis and forecasting• Demand forecasting• Quality assurance analytics• Process parameter optimization• Customer insights	<ul style="list-style-type: none">• Patent text analysis• Anomaly detection• Predictive maintenance• Machine failure prediction and prevention

Value Proposition



To Enterprise Data Scientist

- ◆ Powerful visual environment for integrated data extraction, mining, visualization, reporting and prediction
- ◆ Easy to learn and use
- ◆ No programming required
- ◆ Rapid prototyping
- ◆ Powerful collaboration platform for data analytics
- ◆ Support for very large set of data sources, including Big Data Hadoop clusters
- ◆ Support for popular R language
- ◆ Supported by large and growing open source community, continuously adding value to the platform and ecosystem
- ◆ Ease of deployment and integration with any enterprise IT infrastructure

To CIO

- ◆ Self-documenting applications: never worry about the “hit by a bus” problem again
- ◆ Easy to deploy and maintain
- ◆ Scalable
- ◆ Constantly improving and evolving solution
- ◆ Competitively priced
- ◆ Minimize workforce ramp-up time
- ◆ Maximize workforce productivity and retention
- ◆ Training widely available
- ◆ Widely deployed across multiple industries
- ◆ Large selection of System Integration partners
- ◆ World-class support 24/7

To CFO

- ◆ Flexible licensing options
- ◆ Competitively priced
- ◆ Future-proof

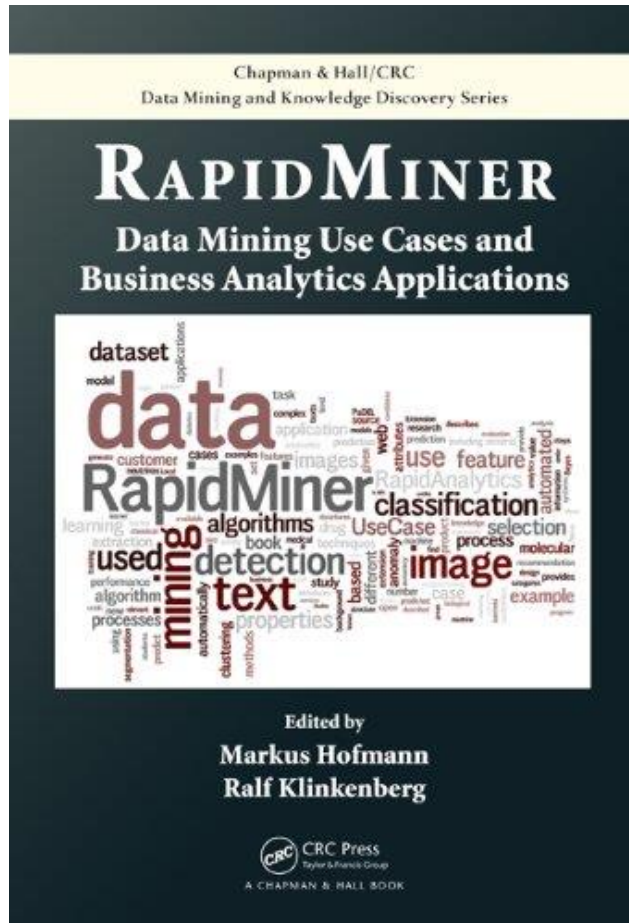


Use Cases

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

RapidMiner Application Use Cases



For more application use 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

<http://www.crcpress.com/product/isbn/9781482205497>

ANWENDUNGEN IN **PRODUKTION & PRODUKTIONS- PLANUNG**

Anwendungen in der Produktion

RapidMiner Anwendung in Fertigung, Produktion und Produktionsplanung:

- Produktdesign & Produktionsplanung: schon beim Produktdesign Prognose von Montageplänen und Montagekosten
- Optimierung von Produktionsparametern zur Qualitätsoptimierung und Kostenminimierung
- Maschinenausfallprognose & -vermeidung
- Präventive bedarfsabhängige Wartung
- Erkennung von Mustern, Ausreißern, Alarmsituationen aus Sensor-Meßwert-Zeitreihen
- Bedarf-, Absatz- und Preisprognose
- Textanalyse & Online/Social Media Monitoring:
Kundenstimmung, Bedarfe, Produkt- & Service-Optimierung

Produktdesign & Montageplanung

ProMondi-Projekt mit Daimler, Miele, Siemens, MTM und diversen Unis,
gefördert vom BMBF und betreut vom KIT:

- Wettbewerbsdruck zwingt zu schnelleren Produktlebenszyklen und Produktvarianten => Zeit- & Kostendruck
- Design-Entscheidung beeinflussen Montage-Aufwände & -Kosten
- Expertise anderer Produkt Designer & Montageplaner nutzen
- => Unterstützung der Designer & Planer mit Data & Text Mining:
- => schon früh beim Produktdesign Prognose von Montageplänen und Montagekosten, um kostengünstige Alternativen früh zu berücksichtigen
- => Zeitersparnis, Kostenersparnis, Know-How-Transfer

Optimierung Produktionsparameter

Optimierung von Produktions- & Fertigungsprozessen:

- Optimierung von Produktionsparametern zur Qualitätsoptimierung und Kostenminimierung
- Erkennung von Zusammenhängen, Mustern, Ausreißern, Alarmsituationen aus Sensor-Meßwert-Zeitreihen
- Maschinelles Lernen von Modellen komplexer Zusammenhänge
- Simulation & Optimierung: Stoffbeigaben, Mengen, Maschineneinstellungen, Reaktionen in kritischen Situationen
- => besseres Verständnis von Zusammenhängen
- => Vermeidung von Ausschußproduktion & Maschinenausfällen
- => Verbesserung der Produktqualität
- => Reduzierung der Kosten

Marktanalyse & Produktoptimierung

Analyse von Kunden-Feedback sowie Online & Social Media Monitoring:

- Sammlung von Kunden-Feedback über E-Mail, Web
- Sammlung von Wartungsberichten, Reperaturberichten
- Online & Social Media Monitoring
- Textanalyse & Stimmungsanalyse
- => Produktprobleme erkennen
- => Bedarfe erkennen
- => Produkt- & Service-Optimierung

OVERVIEW: **BIG DATA 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 ...
 - 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.

=> 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
 - 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.
 - 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/>





GETTING STARTED?

CONTACT US!

Ralf Klinkenberg, RapidMiner GmbH

www.RapidMiner.com

rklinkenberg@RapidMiner.com

