



An Introduction to Advanced Analytics



PREDICTIVE
ANALYTICS

DATA
MINING

ADVANCED
ANALYTICS

BUSINESS INTELLIGENCE



An Introduction to Advanced Analytics.

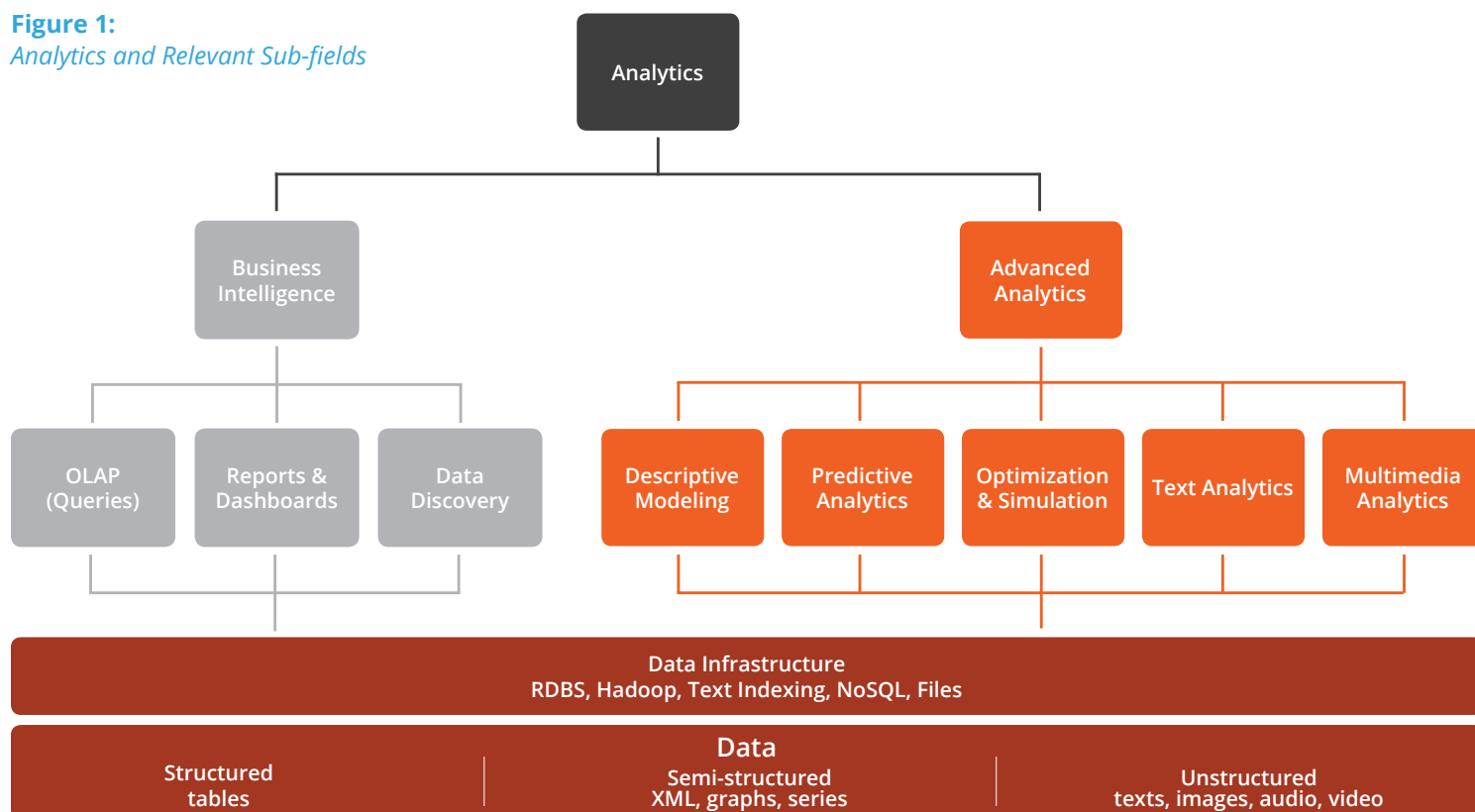
Where Business Intelligence Systems End... and Predictive Analytics Tools Begin

Advanced Analytics is “the analysis of all kinds of data using sophisticated quantitative methods (for example, statistics, descriptive and predictive data mining, simulation and optimization) to produce insights that traditional approaches to business intelligence (BI) — such as query and reporting — are unlikely to discover.”¹

Analytics is an immense field with many subfields, so it can be difficult to sort out all the buzzwords around it. Figure 1 shows how Advanced Analytics, Business Intelligence, and their subfields fit into the overall category of Analytics.

¹ Gartner, *Magic Quadrant for Advanced Analytics Platforms*, Gareth Herschel | Alexander Linden | Lisa Kart, 19 February 2014

Figure 1:
Analytics and Relevant Sub-fields



Analytics

Analytics refers to the skills, technologies, applications and practices for continuous iterative exploration and investigation of data to gain insight and drive business planning. Analytics consists of two major areas: Business Intelligence and Advanced Analytics.

1. Business Intelligence -- traditionally focuses on using a consistent set of metrics to measure past performance and guide business planning. Business Intelligence consists of querying, reporting, OLAP (online analytical processing), and can answer questions including "what happened," "how many," and "how often."
2. Advanced Analytics -- goes beyond Business Intelligence by using sophisticated modeling techniques to predict future events or discover patterns which cannot be detected otherwise. Advanced Analytics can answer questions including "why is this happening," "what if these trends continue," "what will happen next" (prediction), "what is the best that can happen" (optimization).

Business Intelligence

Business Intelligence is a set of architectures and technologies that transforms raw data into meaningful and useful information for business purposes. BI tools, at the most basic level, help business users interpret voluminous data. BI focuses on the storage and retrieval of data from the past, using technologies like data cubes and query engines. Being able to measure past performance is essential in complex business environments. Deriving new opportunities from past data and implementing effective strategies based on Business Intelligence insights can provide a competitive market advantage.

The term Business Intelligence also often refers to the creation and maintenance of data warehouses, often stored in OLAP cubes, in order to support efficient querying and reporting. Besides the querying and visualization of data, traditional BI environments make it possible to implement rule-based alerts to inform decision makers about important events or changes. However, the actions in BI are always defined through human interaction and are performed by humans.

Advanced Analytics

Advanced Analytics deals with the automatic discovery and communication of meaningful patterns in structured as well as in unstructured data. Advanced Analytics goes beyond Business Intelligence and expands the horizons of traditional analytics. With traditional analytics, analysts look at the data and ask "What happened?" With advanced analytics we ask "What will happen?" Business analysts can anticipate actions for their business much like a great hockey player who skates to where the puck is going to be instead of where the puck has been.

Methodologies and technologies from both statistics and computer science have played an important role in the development of advanced analytics, and have contributed to the discipline of Advanced Analytics. The main contributions come from Machine Learning and Data Mining.

Machine Learning

Machine Learning deals with the large-scale and automated creation of statistical models in large data sets, without the need for hypothesis testing, and has been a very active research field for many decades.

Data Mining

Data Mining has enriched machine learning by also covering the necessary steps of data integration and preprocessing in order to create better models. Machine learning focuses on model building. Today, the term data mining is used less often. Among analysts, this term has mainly been replaced by Predictive Analytics, Descriptive Modeling, or other strongly related fields like text analytics or multimedia analytics.

Predictive Analytics

Predictive analytics is the practice of analyzing data to make statistically accurate predictions about future events. Predictive Analytics encompasses a variety of techniques from computer-aided statistics, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown, events. In business environments, predictive models automatically find and exploit patterns found in historical and transactional data in order to extrapolate to future events, and, by that means, predict the most likely future. Models describing those patterns capture relationships among many more factors than human beings can handle. This allows, for example, the identification of previously unknown risks and opportunities.

Generally, the term predictive analytics is used to mean predictive modeling, that is, the “scoring” of new and unseen data with predictive models, as well as forecasting. However, people are increasingly using the term to refer to related analytical disciplines, such as descriptive modeling and decision modeling or optimization. These disciplines also involve rigorous data analysis, and are widely used in business for segmentation and decision-making, but have different purposes and the underlying statistical techniques vary. Predictive analytics is just one part of the advanced analytics market segment.

Going Beyond Tabular Data

When most business people think about data, they envision a classic database, or a spreadsheet – organized rows or columns of usually numbers, that is, “tabular data.” However, more and more of the data being collected today is unstructured. Also related to predictive analytics are fields like:

- Text Analytics, which allows users to derive insights from unstructured data like text collections.
- Multimedia Analytics, which uses similar technologies to transform unstructured data like images, audio, or video into a structured format that can be used as the base for predictive or descriptive analytics.

Analytics models that can process unstructured data provide better predictions.

Summarizing the Differences between Business Intelligence and Advanced Analytics

Business Intelligence uses the data of the past, lets the user find out if what he already believes is true, and requires the user to manually define actions.

Advanced Analytics uses the data of the past, automatically finds hidden patterns too complex for humans to find, and allows the automatic identification and performance of the optimal actions.

Quick Comparison Table

The table below summarizes the differences between Business Intelligence and Advanced Analytics.

	Business Intelligence	Advanced Analytics
Orientation	Rearview	Future
Types of questions	What happened When, who, how many	What will happen? What will happen if we change this one thing? What's next?
Methods	Reporting (KPIs, metrics) Automated Monitoring/Alerting (thresholds) Dashboards Scorecards OLAP (Cubes, Slice & Dice, Drilling) Ad hoc query	Predictive Modeling Data Mining Text Mining Multimedia Mining Descriptive Modeling Statistical / Quantitative Analysis Simulation & Optimization
Big Data	Yes	Yes
Data types	Structured, some unstructured	Structured and Unstructured
Knowledge Generation	Manual	Automatic
Users	Business Users	Data scientists, Business analysts, IT, Business Users
Business Initiatives	Reactive	Proactive

Advanced Analytics vs. Business Intelligence

Sample Questions

Advanced Analytics allows us to ask a whole new kind of question, beyond what we can ask with Business Intelligence. Where Business Intelligence is focused on reporting and querying, Advanced Analytics is about optimizing, correlating, and predicting the next best action or the next most likely action. This table compares the types of questions.

Vertical	Application	Business Intelligence	Advanced Analytics
Financial Services	Fraud detection	Which credit card transactions have been flagged as fraud?	How likely is it that a credit card transaction will be fraudulent?
	Investment holding company: Business development	Which companies went bankrupt last year?	How likely is it that a company will go bankrupt?
	Insurance Underwriting department	Provide me with a list of people who had accidents in the last 6 months.	Is this person likely to have an accident and should we continue to insure them?
Retail and Consumer Products	Retailer	Who bought both beer and pretzels?	How likely am I to purchase a particular product if I have purchased another product? (If I buy beer, how likely am I to buy pretzels?)
	Sales and marketing	What books did you buy from our website last year?	What books might you be most interested in based on your previous patterns of interest?
	CPG Sales and marketing	What is the amount of time between purchases of our product by a single customer?	Predict when individual consumers will exhaust their supply of a certain product.
Telecommunications and Media	Telco Marketing, customer service	Which customers cancelled their service?	Which customers will be most likely to leave your service?
	Media marketing	Which ads were more effective on a particular channel, on a particular night?	Which ads will be more effective on which channel?
	Media Sales and marketing	What movies did your customers rent or purchase?	What movie will your customers be most likely to buy or rent?
Manufacturing	Product design and development	What ingredients can be found in our products?	Which chemical formulations will create a product with specific attributes?

Continued from page 6

Vertical	Application	Business Intelligence	Advanced Analytics
	Product development	What happened last time we introduced a new product?	What will be the impact on profits of introducing a new product line?
	maintenance	What's the last time this machine broke down? What's the maximum production we have ever realized from this machine?	If I changed the maintenance schedule for this machine, how would it impact my production throughput?
Cross industry	Customer service	What reasons did customers give for cancelling their service?	What actions will cause customers to leave your service?
	Economic development	Where are our biggest markets currently?	Where should we focus our expansion efforts?
	Strategic alliances	Which partner was responsible for the most profitable sales?	Which partner has the biggest and best potential?
	Legal department	Collect existing patent infringement cases.	Determine the likelihood of a patent infringement case being filed.
Municipalities, Government	Law enforcement	What crimes occurred in a particular neighborhood last week?	When are certain types of crime most likely to occur?
Shipping services	Logistics	Which packages are on which delivery truck?	How can I optimally assign packages to delivery trucks so as to minimize the time required to deliver all the packages?
Marketing agency	Social media marketing	How many connections does a person have in social media?	Who is the optimal connector in a social network?
Sports/gambling	Odds makers	Who won the most World Cups in the history of the game?	Who will win next year's World Cup, SuperBowl or World Series?
Transportation	Logistics	Which flights were cancelled the most last year.	What cancellation will have the least impact on travelers and our bottom line?

Conclusion

While many companies already use and operationalize business intelligence applications within their business processes to leverage their data assets, the true potential of data is still untouched in many organizations. Advanced analytics, particularly predictive analytics, can help reveal the future and optimize operations.



RapidMiner provides software, solutions, and services in the fields of predictive analytics, data mining, and text mining. We automatically and intelligently analyze data (both structured and unstructured) – including multimedia and text – on a large scale.

Our technology allows companies to use previously unexplored data to inform their decisions, optimize their processes, and take advantage of the highest level of business intelligence. RapidMiner's flagship product is a cutting-edge, open-source data miner. Hundreds of thousands of applications are already in use in more than 50 countries, both as stand-alone applications and engines integrated into customers' own product; this gives both new and existing users a strong competitive edge. RapidMiner boasts a strong and successful list of users, including EADS, GfK, Lufthansa, PayPal, Pepsi, Sanofi, Siemens, Telenor, and Volkswagen.

www.rapidminer.com

RapidMiner USA

RapidMiner, Inc. Headquarters
10 Fawcett St., Suite 502
Cambridge, MA 02138

RapidMiner United Kingdom

RapidMiner Ltd.
Quatro House, Frimley Road
Camberley GU16 7ER
United Kingdom

RapidMiner Germany

RapidMiner GmbH
Stockumer Str. 475
44227 Dortmund
Germany