

Predictive Analytics: Looking into the future

Frank Romeike | Andreas Eicher

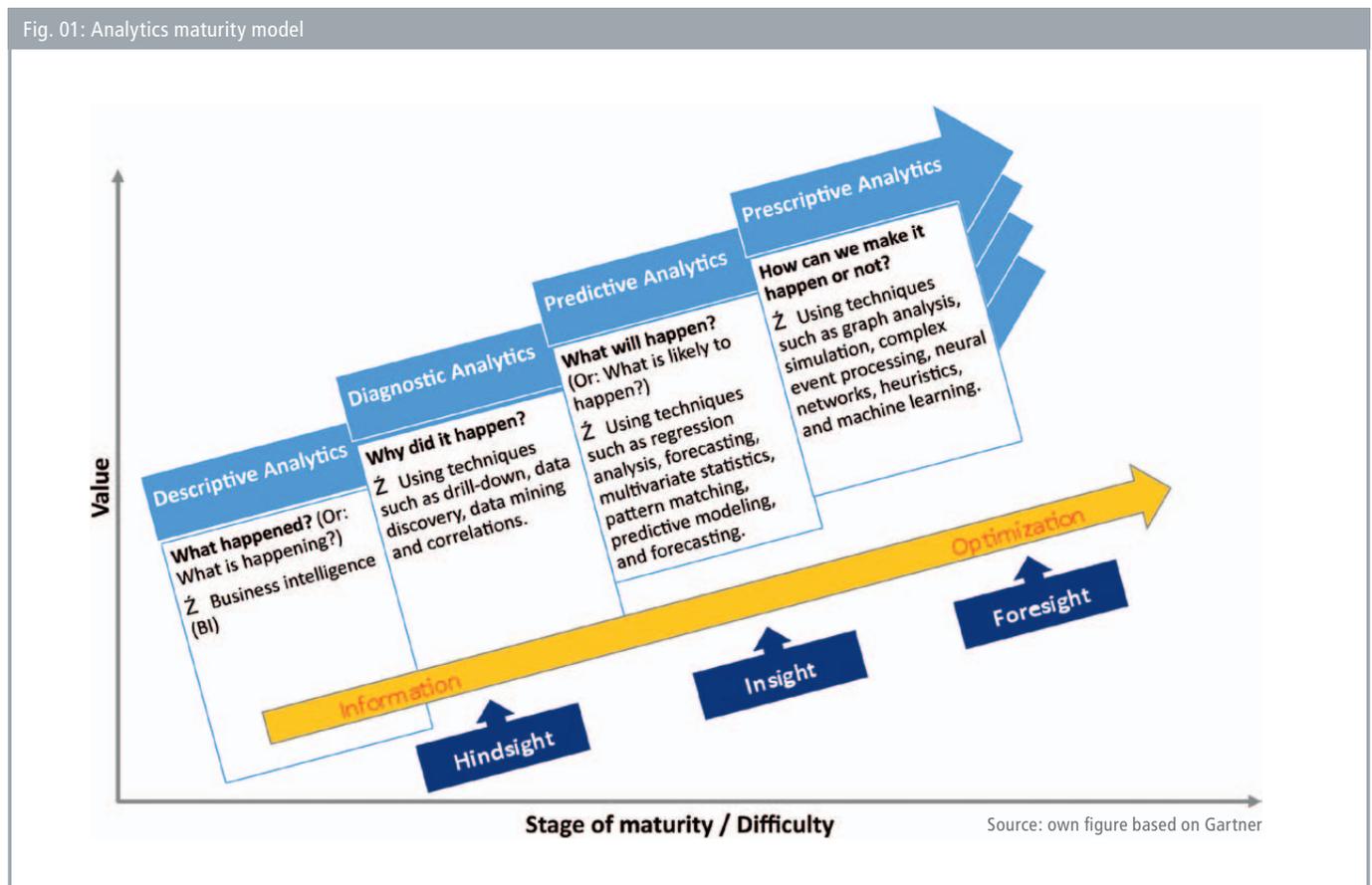
The Oracle of Delphi, a mythical cult centre where human existence came into contact with the world of the gods, was at the centre of Ancient Greek society. Anyone with an important name or rank at that time consulted the Oracle for a prophecy – from the King of Thebes to Croesus to Alexander The Great. It was a legendary place in the days of the Oracle and still has an allure today. Forecasting the future is as old as the human race itself and has lost none of its fascination as time has passed. There is no doubt that the Oracle of Delphi and those in other places would have been well visited in our turbulent times. Politicians, academics and managers would want to look into the future, to ask for advice and take away some recommended actions on their return journey back to their offices. But these days things are a little easier and there's no need to make arduous and time-consuming journeys.

Of ancient oracles and incorrect interpretations

If the Ancient Greeks, Chinese or Egyptians wanted a prediction about what might happen tomorrow, they didn't talk to their risk manager, they consulted their oracle. For example Apollo – the god of divination, moral purity, temperance and the arts, among other things, in Greek and Roman mythology – regularly spoke through his priestess, known as the Pythia. He filled her with his visionary wisdom. The priestess was said to have sat in a cauldron set on a tripod over a chasm in the earth. According to legend, vapours arose from this chasm and transported the Pythia into a trance state (from a scientific perspective this version is improbable, however – archaeologists and geologists have never found a chasm in the rock under the temple nor any explanation for the existence of these vapours).

It is reported that Croesus, the last king of Lydia (born around 591/590 BC, died around 541 BC), was a great believer in transcendental revelation, which he used to obtain answers to questions about the future or to help him make decisions. Croesus left nothing to chance, subjecting the most famous oracular sites of the day (Abai, Delphi, Dodona, Amphiaraios, Ammon) to a test. On exactly the one hundredth day after their departure, the envoys he sent asked what Croesus was doing at that moment. Only the Pythia at Delphi provided the correct answer – Croesus was cooking a tortoise and lamb meat in a cauldron. The test was set because the Lydian king needed help with a strategic question he was facing – what was the likelihood of a victory in the event of war against the Persian Empire. The Pythia prophesied that a great empire would fall if he were to cross the border river, the Halys. The Lydian king is

Fig. 01: Analytics maturity model



said to have interpreted this prophecy in a way that was favourable to him and marched confidently to war. Today, Croesus might well get an analysis of digital information from the world of "Big Data".

Of modern oracles and gold rush fever

The modern oracles of our networked digital age are Big Data and data analytics. Data gatherers such as Google and Amazon survey the world, create personality profiles and comb through huge volumes of data at lightning speed for patterns and correlations, allowing them to make predictions in real time. They provide a targeted look into the crystal ball. States, research institutions and commercial companies hope that they can provide accurate predictions to minimise the risk of their own actions and to better assess the opportunities for future activities. It is also about making structured use of the knowledge in organisations. Overall, around 3.2 billion people use the Internet, producing permanent data through their mobile phones, fitness bands, smart watches, networked navigation units and cars. Online sellers know our secret desires better than we do ourselves. Political attitudes can be accurately ascertained from Twitter messages. Data and algorithms enable potential crimes to be anticipated before they are even planned or committed.

Experts estimate that the worldwide data volume will increase from around 8,500 exabytes today to around 40,000 exabytes. An exabyte is a trillion (10¹⁸) bytes, a billion gigabytes, a million terabytes, a thousand petabytes. In short – the data tsunami is rolling. Such huge data volumes have a lot of potential. The data has been compared to all kind of things, variously described as the new oil, gold or even diamonds of our age. And many companies want to take advantage of this gold rush fever.

Big Data methods: Target groups, risk analyses and management

Analytics methods are becoming increasingly popular – including for banks and insurance companies. One of the main reasons is that they enable relationships to be identified, forecasts derived and then used for decisions, ideally in real time. In addition, they are a key factor in digitalisation of entire business processes. The Fraunhofer Institute believes that Big Data opens up new possibilities in Business Intelligence (BI) and Business Analytics (BA), enabling specific patterns and relationships to be identified and possible trends to be predicted [see Fraunhofer 2015]. A study entitled "Potential and Use of Big Data" by the digital association Bitkom [see Bitkom 2014] draws similar conclusions. The study reveals that 48 percent of companies see the greatest potential of Big Data in providing additional backing for decisions and for trend analyses (39 percent) and setting up forecasting and early warning systems (37 percent). "Deutsche Bank Research" rates Big Data as a production, competitive and therefore growth factor with relevance for the entire economy. And the analysts conclude that: "Modern analysis technologies are becoming established in all areas of life and are changing our everyday existence". [see Deutsche Bank Research 2014].

Decision makers at the ING-Diba bank know all about this. The institution searches through huge volumes of data as part of its target group marketing. In addition, Big Data analysis methods provide fast and detailed options for performing risk analyses. What is more, analysis methods provide valuable ways of identifying previously unknown patterns in existing data records and enable these to be accessed to make more sound decisions in risk management.

Learning from the past to early warning with Predictive Analytics

Big Data experts are convinced that Predictive Analytics is one of the most important Big Data trends, particularly in the area of risk management. The analytics maturity model from Gartner provides a good overview. Gartner differentiates four levels of maturity (see ► Fig. 01):

Descriptive Analytics is all about the question "What has happened?", in other words analysing data from the past to understand potential effects on the present (see Business Intelligence).

Diagnostic Analytics looks at the question of "Why did it happen?", in other words analysis of the cause/effect relationships, interactions and the consequences of events (see Business Analytics).

Predictive Analytics deals with the question of "What will happen?", in other words analysing potential future scenarios and generating early warning information. Based on technologies from data mining, statistical methods and operations research, the probabilities of future events are calculated.

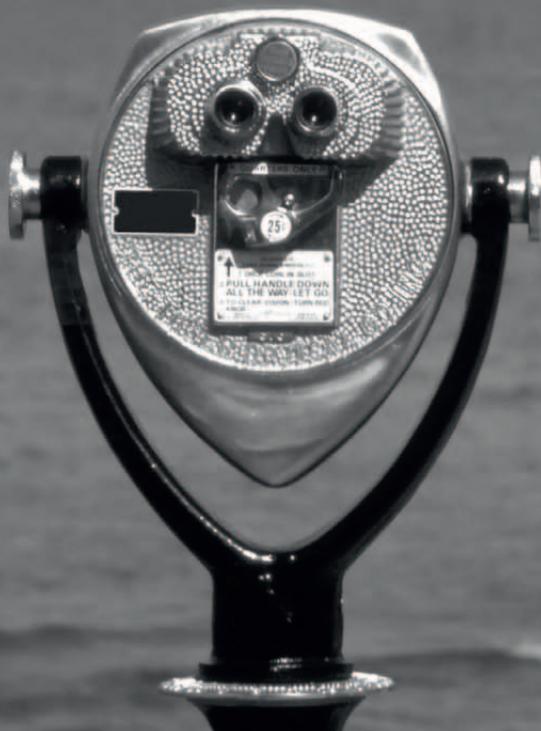
Prescriptive Analytics addresses the question "What do we have to do to make sure a future event will (not) occur?", which essentially – based on the results of Predictive Analytics – means simulating measures based on stochastic scenario analyses and sensitivity analyses [see Romeike 2010 and Romeike 2015].

Analysis methods as an early warning system

A study on "Analytics as a Competitive Factor" [see Gronau/Weber/Fohrholz 2013] summarises the potential benefits for insurers and banks: The added value for insurance companies lies in individual targeting of existing and new customers. For banks, there are a variety of possible uses both for private and business customers. The potential is greatest in the area of risk management, particularly liquidity management. For example, targeted analyses can lead to cost reductions for short-term loans [see Gronau/Weber/Fohrholz 2013]. This also includes the use of analysis methods as an early warning system to detect and counter "weak" signals – for example shifts in the market, changed customer preferences – and also to prevent economic crime. Modern analytics methods also enable information to be compiled about customers' income, wealth, education level, professional career and current behaviour, so that it can be evaluated for scoring and ultimately used in loan decisions. The aims include predicting customers' payment behaviour and ability. It is a highly complex system that has to be managed. Not by machines but by people, as it is ultimately all about people and their needs.

Of increasing requirements ...

In parallel to the data tsunami the requirements to understand the underlying laws and cause/effect relationships are increasing. In his book "Calculating the World" Klaus Mainzer, a German philosopher and scientific theorist, refers to the fact that Newton's understanding of the law of gravity did not come to him just by continuously watching apples falling from trees [see Mainzer 2014, p. 27]. In other words, bits and bytes must be matched by the ability to not only analyse but also interpret the data obtained. Banks and insurance companies, which deal with management of risks and opportunities on a daily basis, depend on drawing the correct conclusions from the available data. The fact that a pattern exists means that it must have come about in the past. This does not necessarily mean that a conclusion based on this pattern will have any validity for the future.



Risk managers and Big Data analysts alike often walk into a trap if they fail to recognise the difference between correlation and causality and, as a result, interpret information incorrectly and draw the wrong conclusions. A mathematically high correlation between two variables does not mean that there is a causal relationship between those variables. The classic example is that a high level of statistical correlation can be identified between the stork population and the birth rate. Theoretically, the relationship between the two variables could be a cause/effect relationship. Variable A can be the cause of B, or B can be the cause of A. However, it is also possible that neither of the two cause anything. Instead, there may be a third variable that has influenced both A and B. In our example, this could be industrialisation, which led to both a fall in the birth rate and a reduction in the stork population.

In large volumes of data involving numerous factors, random correlations can be derived very easily. But this does not mean that a causal relationship exists between the factors. Big Data protagonists respond that in the world of Big Data, correlation replaces causality. Big Data methods prove particularly successful on issues that can no longer be described using simple laws due to their extremely high level of complexity. In this context, Big Data is not really about huge data volumes at all, but about changing the way we think to acquire knowledge. Applying this to the laws of gravity – Isaac Newton was looking for the reason why apples fall. In the world of Big Data, causality is not really important – in an ideal situation it explains what happens but not why. Big Data can thus help in “finding the needle of knowledge in the haystack of data so that the causes can be investigated” [Mayer-Schönberger 2015, p. 17].

There is thus a risk that trends and laws from the past will simply be applied to the future. For companies, this means taking a meas-

ured, intelligent, targeted and cautious approach to the analytic techniques used. Ultimately what matters is not the volume of data and algorithms but creating meaningful links between them.

... and data ownership

Despite all the euphoria in modern data and analysis management, there are also some critical voices. Viktor Mayer-Schönberger, Professor at the Internet Institute at the University of Oxford and author of the book “Big Data” [see Mayer-Schönberger/Cukier 2013], in future the power will lie less with those who analyse data than with those who have access to it. In this context, it is easy to understand why many people are uncomfortable with organisations and companies that appear to be collecting and analysing ever-larger volumes of data.

Deutsche Bank Research asks “whether digital ecosystems, the secret services or other actors on the Internet will not damage themselves with their business practices in the medium to long term (for companies: “undermine the foundations of their business”) because mistrust could increase and the willingness to stay on digital channels undetected (or unprotected) could gradually fall.” The results of the study on “Opportunities from Big Data and the Question of Protecting Privacy” by the Fraunhofer Institute point in a similar direction. Google is the most widely used search engine by some distance. But it is followed by the privacy-friendly search engine “DuckDuckGo”. This behaviour shows that some Internet users already have a certain amount of mistrust towards the Internet and especially the data processing and usage that lies behind it. As a result, this could lead to modified and changed media use and consumer behaviour. This brings a risk of macroeconomic harm for all providers of web-based technologies.

Summary: Hopes and many outstanding questions

Big Data is linked to the hope that in the future we will understand the world better and, for example, will be able to promptly identify and counter the drivers of risk using weak signals. This is based on the expectation that an increase in the quantity of data will also lead to a new level of quality.

As with ancient oracles, in today's world of "Predictive Analytics" and "Big Data" correct interpretation of the results is crucial, as Big Data heralds the end of the causality monopoly. That brings us back to Croesus. Following the oracle's prophecy, he crossed the Halys river on the border and arrived in Cappadocia. The military conflict between the Persian king Cyrus II and Croesus came to an end at the Battle of Pteria – with Croesus on the losing side. What he had not predicted was that by going to war he would ultimately destroy not his enemy's great empire but his own.

Today, as then, if the cause/effect relationships are not understood, the patterns and correlations obtained from Big Data remain largely random. We need to guard against immediately identifying a logical cause/effect relationship in every statistical correlation. Based on Kant's "Critique of Judgement", both determinative and reflective judgement exist. Determinative judgement subsumes particulars under a given law or rule, while reflective judgement aims to find universals for given particulars. Applied to the world of "Big Data" and "Predictive Analytics", this means that we have to link the flood of data to theories and laws.

As people and a society, we must address the question of how much security and predictability we want on the one hand and how much freedom and risk on the other. In other words, between "standing on the brake" and constantly using the overtaking lane there are a lot of analysis speeds. We need to investigate these. One of the core questions in this area is do we want to surrender ourselves to a data dictatorship and live in a world where Big Data knows more about our past, present and future than we can remember ourselves [see Mayer-Schönberger 2015, p. 18]? In this context, misuse of Big Data correlations and concentration in data monopolies can lead to negative consequences for individuals and the whole of society. This dark side of Big Data needs to result in transparent and binding rules and in a broad-based discussion of the opportunities and limits of the brave new world of data.

Finally, we should note that algorithmic treatment of people violates human dignity. The German constitutional court ruled on this back in 1969: "It would be incompatible with human dignity (...) to register and catalogue (...) the entire personality of a person, even if in the anonymity of a statistical investigation, and to thus treat that person as an object, every aspect of which is available for inventory." [German Constitutional Court (BVerfG) 1969]. So, where do we want to go in the world of data? This question requires an urgent answer – one that is comprehensive and in the common interest of society.

Literature

Bitkom [2015]: *Potenziale und Einsatz von Big Data [Potential and Use of Big Data]*, Berlin 2015.
 BVerfGE [1969]: *Decisions of the German Constitutional Court of 16 July 1969, 27, 1 (sample census)*.
 Deutsche Bank Research [2014]: *Big Data, Die ungezähmte Macht [Big Data, The Untamed Power]*, Frankfurt am Main, 2014.
 Fraunhofer Institute for Secure Information Technology SIT [2015]: *Chancen durch Big Data und die Frage des Privatsphärenschutzes [Opportunities from Big Data and the*

Question of Protecting Privacy], Stuttgart 2015.

Gronau, N./Weber, N./Fohrholz, C. [2013]: *Wettbewerbsfaktor Analytics – Reifegrad ermitteln, Wirtschaftlichkeitspotenziale entdecken [Analytics as a Competitive Factor - Determining Maturity, Identifying Potential Efficiencies]*, Berlin 2013.

Levitt, S. D./Dubner, S. J. [2009]: *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, New York 2009.

Mainzer, K. [2014]: *Die Berechnung der Welt – Von der Weltformel zu Big Data [Calculating the World - From the Global Formula to Big Data]*, München 2014.

Mayer-Schönberger, V./Cukier, K. [2013]: *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, London 2013.

Mayer-Schönberger, V. [2015]: *Was ist Big Data? Zur Beschleunigung des menschlichen Erkenntnisprozesses [What is Big Data? On the Acceleration of the Human Knowledge Process]*, in: *Aus Politik und Zeitgeschichte (APuZ)*, 11–12/2015, p. 14-19.

Romeike, Frank [2010]: *Risikoadjustierte Unternehmensplanung - Optimierung risikobehafteter Entscheidungen basierend auf stochastischen Szenariomethoden [Risk-Adjusted Corporate Planning - Optimization of Risky Decisions Based on Stochastic Scenario Methods]*, in: *Risk, Compliance & Audit (RC&A)*, 06/2010, S. 13-19.

Romeike, F. (2015): *Szenarioanalyse: Lernen aus der Zukunft [Scenario Analysis: Learning from the Future]*, in: *FIRM Yearbook 2015, Frankfurt/Main 2015, S. 118-120*.



Authors

Frank Romeike

Managing Partner of RiskNET GmbH, board member of the Association for Risk Management and Regulation, and editor in chief of RISIKO MANAGER magazine



Andreas Eicher

Editor in chief of the Competence Portal RiskNET – The Risk Management Network, editor of the Frankfurt Institute for Risk Management and Regulation (FIRM) and journalist on the subject of risk management, compliance and digitization