Predictive Analytics: The Next Significant Breakthrough in Cloud Computing

Introduction

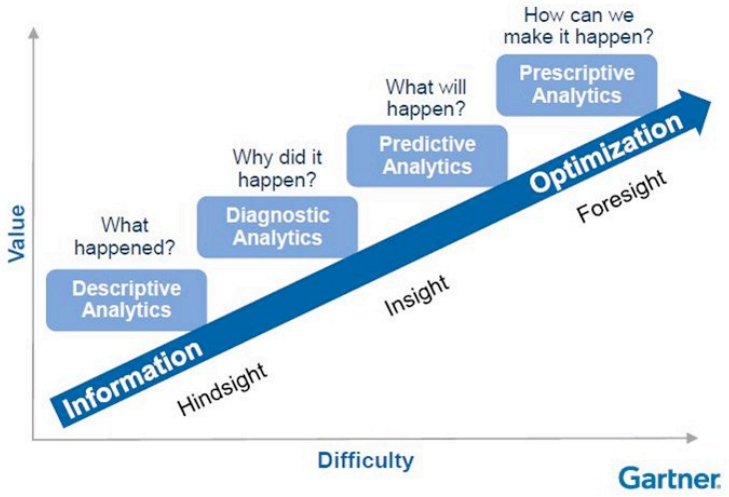
Analytics is widely regarded as the systematic computational analysis of data or statistics. The new economy that is emerging currently is based on data, and much of the innovation occurring throughout the technology landscape is geared towards using this data to generate business value. With advances in technologies such as connected devices incorporating sensors combined with the massive compute and storage capabilities within the cloud, also known as Internet of Things (IoT), data can now be gathered, stored and analyzed from thousands of nodes, at much faster rates than ever before. While businesses and governments have been collecting data for years, what has changed the landscape is the advancement in technology has increased the types of data that is collected and enabled the speed at which this data is gathered. Most companies today use some form of analytics tool to use data to evaluate their key metrics. This data is largely historical in nature and is usually presented in a dashboard or other visual format. The most impactful change in this ecosystem is that businesses can now leverage newer statistical methods and technologies to create predictive models that help analyze their existing data to enable changes in business practices that would help drive more efficient operations, increase sales, and improve customer service, among other important activities – what is termed as “advanced analytics”.

Even though advanced analytics are improving, there remains a continued skepticism within the corporate environment about using data driven sources to make critical business decisions. According to a study conducted by PwC in 2014, while 64% of large companies have made the shift to using some sort of analytics tool, only 32% of executives actually rely on advanced analytics to make critical decisions, while, according to another study by McKinsey & Co., only 28% of senior executives believe the quality of their strategic decisions was generally good. The good news is recent

innovations in technology are making advanced analytics easier to use, more available to the masses, and a hot bed of innovation. As such, we believe we are at the cusp of a sea change in the use of this exciting technology.

Overview of analytics

There are 4 types of analytics: descriptive, diagnostic, predictive, and prescriptive. All of them are about using data to gain insights into business trends, to enable better decisions.



Descriptive analytics describe what is happening in our business today (Business Intelligence, or BI, tools like financial reporting). Diagnostic analytics describe why a particular event happened. Both of these types of tools require a fair amount of human input and analysis, and neither by itself answers questions about future trends. Predictive analytics is the practice of taking the data and patterns we discovered to create models that answer the question “what will happen” in the future. Prescriptive analytics are tools that help drive decisions about what to do in the future by modeling the potential outcomes of each possible decision and helping us choose the best option. Predictive and Prescriptive analytics are often referred to as “Advanced Analytics”, which are tools that use sophisticated quantitative methods to produce insights that traditional BI tools will likely not discover. In our view, predictive analytics is positioned for

explosive growth over the next few years. According to TDWI Research, while 39% of corporate analytics users are currently using Predictive Analytics currently, an incremental 46% are planning to use it in the next 3 years.

Predictive analytics technology trends and drivers

Techniques used in predictive analytics include data mining, statistics, modeling, machine learning, ensemble techniques, and artificial intelligence, collectively used to analyze current data to identify risks and opportunities in the future. Some of the techniques used in data mining, statistics, and text analytics are used to create potential outcomes and possibilities by linking patterns and relationships between both structured and unstructured data. Structured data includes data that is easily organized into readable and useable formats (e.g. data found in spreadsheets), while unstructured data refers to data that is not organized in a pre-defined manner (e.g. social media content). According to industry sources, only about 20% of the data we currently have access to is structured, while about 80% is unstructured, so there is a lot of work to be done. Many companies are tackling the difficult challenge of extracting meaningful data from unstructured sources. For example, since 2012, IBM has made 25 acquisitions in cloud related technologies, most recently its $1B acquisition of Merge Healthcare, giving its Watson analytics platform the ability to “see” by incorporating data and images obtained from Merge Healthcare’s medical imaging management platform. This is an industry wide trend that should accelerate over the next few years, in our opinion.

Use cases and examples

On the demand front, use cases for predictive analytics are increasing. A few examples of what organizations typically want to predict include customer churn rate, revenue, equipment failure, market volatility, response to a marketing offer, fraud, and cyber security breaches, among others. Some practical applications of predictive analytics include the following:

Application type	Description of application
Sales/Marketing	Marketing campaigns, sales and customer service
Healthcare	Determine patient risks and diagnosis
Insurance	Predicting future claims
Financial services	Fraud monitoring

Internet-of-Things (IoT): While there are numerous use case and examples of how predictive analytics is currently being used, collecting the data is the first step, which arrives from the nine key "settings" identified below where IoT creates value.

Setting		Description	Examples
	Human	Devices attached to or inside the human body	Devices (wearables and ingestibles) to monitor and maintain human health and wellness; disease management, increased fitness, higher productivity
	Home	Buildings where people live	Home controllers and security systems
	Retail environments	Spaces where consumers engage in commerce	Stores, banks, restaurants, arenas—anywhere consumers consider and buy; self-checkout, in-store offers, inventory optimization
	Offices	Spaces where knowledge workers work	Energy management and security in office buildings; improved productivity, including for mobile employees
	Factories	Standardized production environments	Places with repetitive work routines, including hospitals and farms; operating efficiencies, optimizing equipment use and inventory
	Worksites	Custom production environments	Mining, oil and gas, construction; operating efficiencies, predictive maintenance, health and safety
	Vehicles	Systems inside moving vehicles	Vehicles including cars, trucks, ships, aircraft, and trains; condition-based maintenance, usage-based design, pre-sales analytics
	Cities	Urban environments	Public spaces and infrastructure in urban settings; adaptive traffic control, smart meters, environmental monitoring, resource management
	Outside	Between urban environments (and outside other settings)	Outside uses include railroad tracks, autonomous vehicles (outside urban locations), and flight navigation; real-time routing, connected navigation, shipment tracking

Source: McKinsey & Company

IoT is the concept of objects embedded with sensors, software, semiconductors and connectivity, to create a network of "things" that collect and exchange data. This trend has the ability to shift the way the community and companies interact with their

surroundings. The ability to monitor objects electronically can bring data driven decisions possible for resulting in saving time and money for people and businesses. While consumer applications seem to attract a lot of the attention, B2B applications have the ability to create more value. Companies that take advantage of IoT technologies will require leadership within organizations to fully embrace data-driven decision making. Actual use cases and examples are growing in several different areas:

Financial risk and insurance

What's Predicted	**Examples**
Bodily harm from car crashes	Allstate - predictive modeling tripled accuracy of predicting injury liability
Insurance claims	Decreased loss ratio by 1/2 point resulting in savings of $50 million

Health

What's Predicted	**Examples**
Breast cancer	Stanford Hospital - diagnoses breast cancer human doctors could not detect
Effect of drug	Pfizer- Probability patient will respond positive to a new drug

Crime / Fraud detection

What's Predicted	**Examples**
Tax returns	Predict ranking of tax returns suspected of cheating resulting in finding 25 x more evasion
Murder	Maryland - Detect inmates more at risk to be perpetrators

Government

What's Predicted	**Examples**
Voter persuasion	Obama campaign predicted which voters would be influenced by campaign contact
Dropouts	Arizona State - Predict which students are at risk of dropping out

Employee / HR

What's Predicted	**Examples**
Quitting	HP-Flight risk score for employees so that managers may intervene in advance
Skills	LinkedIn labels your profile with skills it predicts you will have from written content
Job applications	CareerBuilder- Predicts positions for which each job seeker will apply

Source: *"Predictive Analytics"* by Eric Siegel

The market for Business Intelligence, including predictive analytics, is estimated to be about $13-14B in 2015, according to multiple research firms. Research from TDWI indicates a potential doubling of use of predictive analytics over the next 3 years, driven by 7 key forces:

1. Ease of use: predictive analytics tools are becoming much easier to use. According to the Forrester Research's Forrester Wave study on big data analytics, there are at least 13 large companies developing a variety of cloud and hybrid analytics solutions, with many more companies focusing on industry specific solutions. The tools are becoming easier to use, and, combined with the massive amount of compute and storage horsepower that exists in cloud providers, analytics tools are being made available to more end users.

2. Open Source software: technologies like "R", which is a free statistical modeling environment, and Hadoop, which is a software library that allows for the distributed processing of large data sets across clusters of computers, are enabling a wide community to further innovate and participate in this segment.

3. Innovation from startups: according to data from Mattermark, there are currently at least 100 analytics startups that have raised over $7B, the majority of which (85%) occurring in the past 2 years. The top 10 companies alone (e.g. Palantir) have raised a total of over $3.4B to date. This pace of growth and investment in innovation should ultimately lead to new and useful analytics solutions.

2015	$4,363.70	62%
2014	$1,635.50	23%
2013	$490.00	7%
2012	$252.60	4%
2011	$156.10	2%
Pre-2011	$114.80	2%

Source: Mattermark

4. Integrated analytics tools: advanced analytics tools are increasingly becoming integrated with other tools. To date, using advanced analytics tools have generally

meant most organizations using a separate analytics application to augment their other transaction processing tools. Given these tools operate on a standalone basis, the use of these analytics tools sometimes creates latency and loss of granularity. However, newer technologies such as in-memory computing (IMC) and hybrid transaction/analytical processing (HTAP) are enabling a new generation of integrated, packaged business applications that include analytics, thereby making the analytics capability essentially seamless to the end users. Companies like SAP (with their HANA platform) and Workday, for example, have already incorporated IMC and HTAP, while others are using a hybrid approach. That said, we are early in the adoption cycle of these technologies, and thus expect this to be a multi-year transition.

5. Democratization and Consumerization: While companies are increasingly hiring data scientists to write code and applications to leverage big data, there are more employees with access to and use of these models. There is an emerging class of employees called "Citizen Data Scientists" that are driving the end use of these analytics tools. These are data oriented employees within individual lines of business, such as HR and sales/marketing, who use data to support their business decisions. According to Gartner, the number of citizen data scientists will grow 5 times faster than the number of highly skilled data scientists through 2017. Finally, according to TDWI, over 80% of active users for predictive analytics stated business analysts (not data scientists) are building their predictive models. This explosive growth in data users will continue to increase demand for useful, predictive analytics tools.

6. Social media and mobile: Social media is also increasingly playing a large role in driving demand. More consumers are using social media sites like Facebook and Twitter, applying analytics to mainstream usage and moving it beyond just the major Internet players, such as Google and Yahoo. Additionally, mobile usage is exploding. According to commScore, mobile access to the Internet has quadrupled over the past

four years, with the predominant activity being social networking. More people aged 55+ are now visit social networking sites on their mobile phone than ever. Companies now understand that a fast and simple mobile experience will often result in positive consumer behavior and increased mobile revenue, and are using predictive analytics to improve the consumer experience and ultimately revenue for themselves. The use of mobile as the primary computing device among consumers has significantly outpaced desktop (up 4x vs. up 37% respectively) since 2011, according to commScore. One of the other key drivers is the increased demand and use of podcasts, which have grown to 46M monthly users, according to Edison Research and Triton Digital.

7. Cloud based solutions: The availability of predictive analytics in the cloud is significant in the last few years. Recent announcements from Amazon Web Services and Microsoft on Cloud based products have added to the excitement. A technology that was typically hard to get started with due to complexity, cost, skill set requirements, and processing needs, is now accessible to more end-user organizations that otherwise might not be able, or which would be unwilling, to attempt to use predictive analytics.

Companies driving predictive analytics solutions

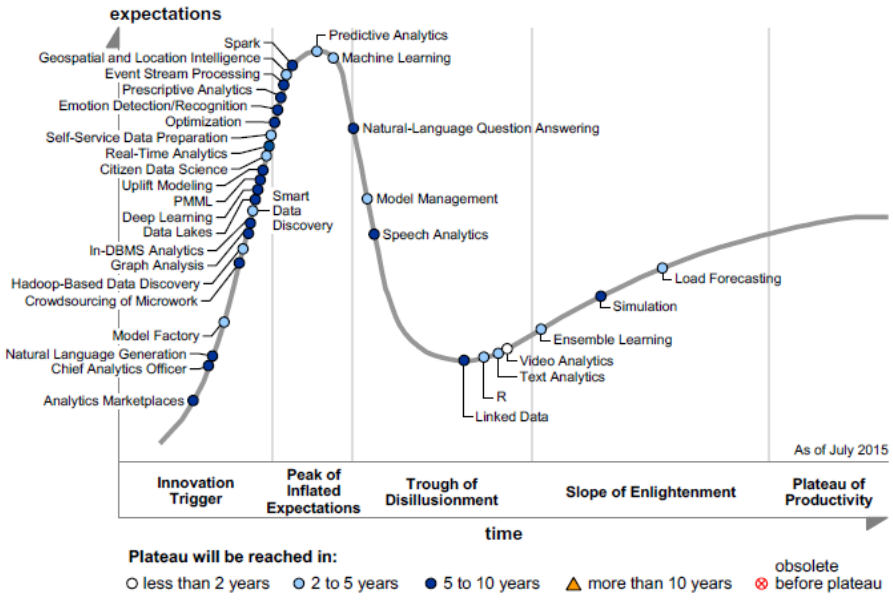
There are two common methods in obtaining predictive analytics capabilities. The first is simply to outsource to an advanced analytics service provider that use either proprietary or commercial analytics tools. The second is to purchase either on-premises or a SaaS (Software-As-A-Service) analytics application. There are several predictive analytics solution providers available for the enterprise to choose from. The analytics solution providers are categorized as either “general purpose” or “specific”. Market research reports will generally focus on the general purpose. Gartner’s 2015 Magic Quadrant for “Advanced Analytics Platforms” has identified 16 leading providers that are used to build these solutions. Innovation in this area is occurring both by large public companies and emerging private enterprises.

Source: Gartner 2015

Time line to mainstream adoption

Gartner's well-known "Hype Cycle" has been used as a graphical representation of the life cycle stages a technology goes through from conception to maturity and widespread adoption. Gartner's Hype Cycle for Advanced Analytics and Data Science is represented in the below diagram. For a detailed discussion and description of the Hype Cycle please refer to pages 12-13.

Figure 1. Hype Cycle for Advanced Analytics and Data Science, 2015



Source: Gartner (July 2015)

According to Gartner, predictive analytics is expected to hit mainstream adoption within the next 2-5 years. The surge in the interest is mainly in the area of predicting demand of products and services. Naturally, companies now want to capitalize on this very valuable resource by utilizing this technology to gain additional insights they can apply to optimize their marketing, and various performance management practices.

Conclusion

The combination of the IoT trend with the emerging economy that is increasingly based on the use of data, we will continue to realize rapid acceleration in the pace of innovation in the data analytics ecosystem. With new tools being introduced by both large companies and startups alike, many of them cloud-based, predictive tools will be the key driver for end user demand. At the same time, end use cases will increase beyond just traditional BI applications to include many new areas such as healthcare. As such, we believe predictive analytics will be the next breakthrough area in cloud computing.

Reference

The hype cycle is five overlapping stages in a technology's life cycle:

1. **Technology Trigger:** This is when the technology is conceptualized. There could be prototypes but normally no functional products or market that have been studied.
2. **Peak of Inflated Expectations:** The technology is implemented, by early adopters and there is likely a lot of publicity about how successful the implementations have been.
3. **Trough of Disillusionment**: Early failures may lead to disappointment in the technology. Some producers are unsuccessful resulting in their products. Continued investments are often contingent upon problems be addressed
4. **Slope of Enlightenment:** The technology's potential for additional applications will often become broadly understood resulting in an increasing number of companies test the product in their environments.
5. **Plateau of Productivity:** The technology will then become widely implemented. The products place in the market well understood. Standards will often soon arise for evaluation.

In order to have a general understanding of Gartner's Hype Cycle the below diagram is a visual representation that can often for any product or service.

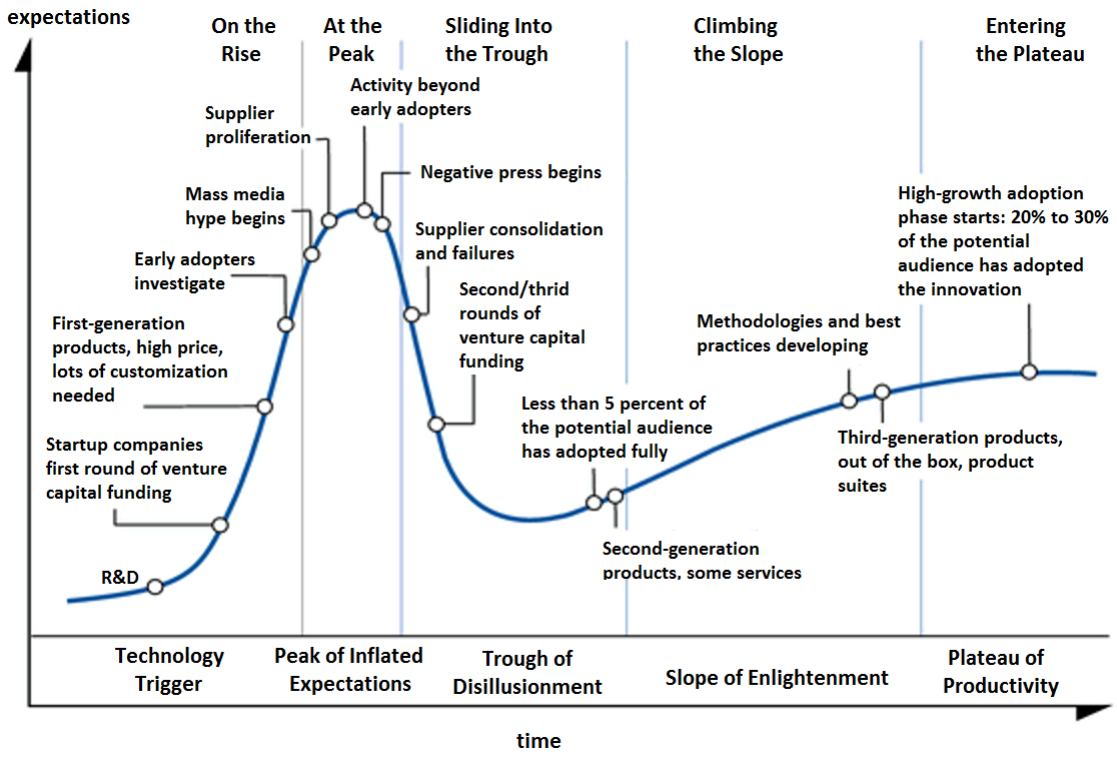
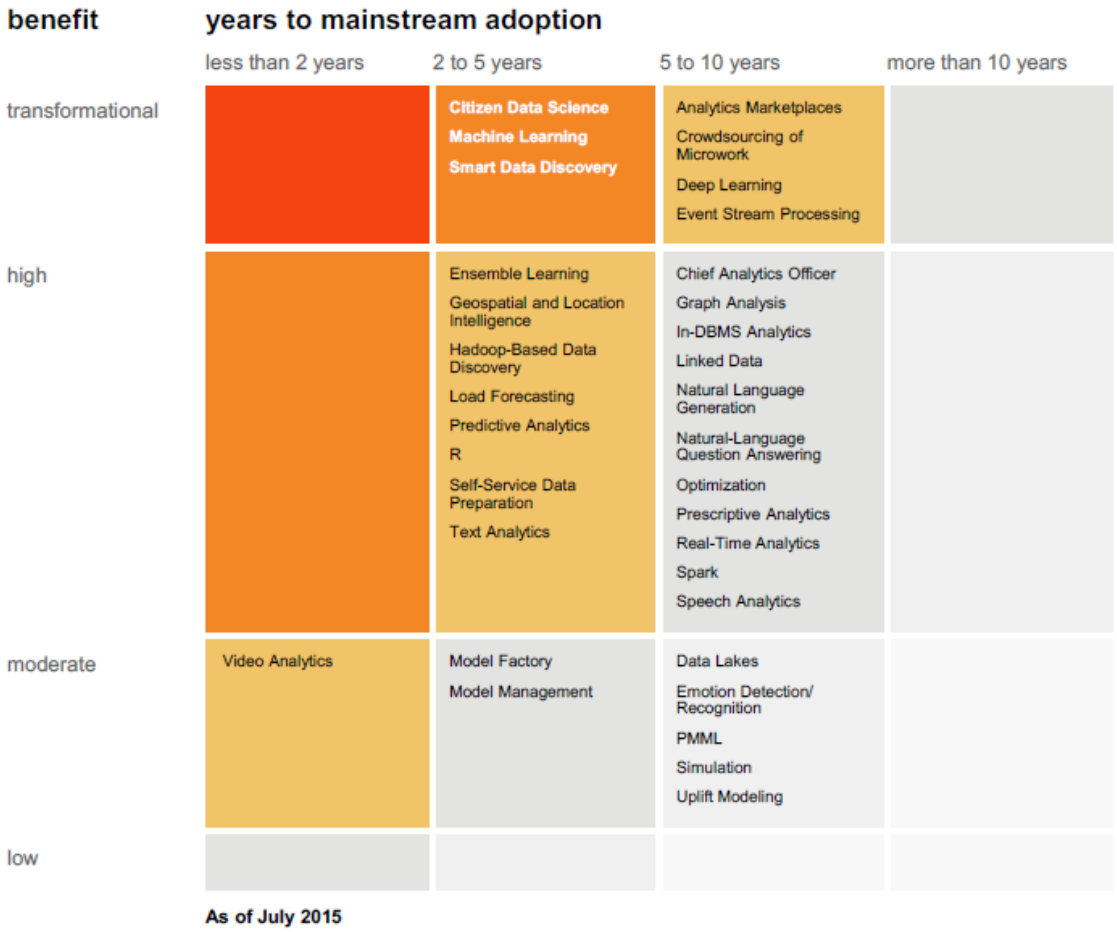

expectations
On the Rise
At the Peak
Sliding Into the Trough
Climbing the Slope
Entering the Plateau
Activity beyond early adopters
Supplier proliferation
Negative press begins
Mass media hype begins
Supplier consolidation and failures
Early adopters investigate
Second/thrid rounds of venture capital funding
First-generation products, high price, lots of customization needed
Methodologies and best practices developing
High-growth adoption phase starts: 20% to 30% of the potential audience has adopted the innovation
Less than 5 percent of the potential audience has adopted fully
Startup companies first round of venture capital funding
Third-generation products, out of the box, product suites
R&D
Second-generation products, some services
Technology Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time

Figure 2. Priority Matrix for Advanced Analytics and Data Science, 2015



Digital research giant comScore's March 2015 study "U.S. Digital Future in Focus" provides perspective because of its four-year view. It reveals that mobile access to the Internet has quadrupled over the past four years, and desktop has risen 37%. (Because tablet Internet access began with almost no share in 2010, it increased by almost 1,800%, mostly because tablet adoption increased exponentially in a short period.) The study also found that the predominant Internet activity on mobile was social networking. Podcasting is now gaining ground. Podcasting is now between content marketing and social. It's content because of the useful information or entertainment it provides; it's social because it provides personal interaction with one of the voices of a brand.

In March 2015, Edison Research and Triton Digital released "The Infinite Dial 2015," their study of the digital consumption habits of 2,002 Americans. Monthly audio podcast consumption grew from approximately 39 million monthly users in 2014 to approximately 46 million in 2015, an increase of 18%.

BitNavi is a blog conceived by Karl Motey in the heart of Silicon Valley, dedicated to emerging technologies and strategic business issues challenging the industry.

Follow them on Twitter: @bitnaviblog

David Darrough has over 30 years of experience in high technology products, services and education.

Follow David on Twitter: @DarroughDavid and his OpenStack website daviddarrough.com