RESEARCH
PAPER

Business Analytics in the Machine Learning Era

February 2018

CONTENTS

Executive Summary

As a commodity, data confounds traditional economic analysis. Data volumes have exploded, and so too has the value of that data. Many organisations claim to be "data-driven", and a consensus of opinion exists that decisions driven by data rather than human instinct – and all of the assumptions and biases that power that instinct – are likely to be more accurate and made faster than those based on gut feeling. It is an intuitive concept – the more informed a decision is, the more likely it is to be correct.

What varies across organisations is the quality of information available to decision makers – and the speed with which it is available to them. A huge spectrum of data analytics exists across organisations – from a few canned reports being delivered to the inboxes of decision makers on a fixed schedule to self-service Business Intelligence (BI) and analytics, powered by machine learning (ML) that is available on multiple devices to enable decisions driven not just by data, but by business insights.

Computing surveyed approximately 110 business decision makers representing businesses ranging in size from fewer than 250 employees to those with many thousands and from multiple industry sectors in order to establish how UK businesses were distributed across this spectrum. The paper discusses the barriers to data-driven decision making facing organisations, as well as the impact of self-service BI on data governance. The paper concludes with a discussion of how the new generation of BI and analytics can use machine learning to automate insight and just what the potential of this breakthrough might be.

The Extent of Data-Driven Decision Making

When asked to describe their businesses in terms of its use of analytics, only 27 per cent of our respondents considered their organisation as "advancing." A further five per cent described theirs as "leading." However, the largest single proportion of respondents – 46 per cent – described their employer as "following," with a further 22 per cent stating that they were "falling behind."

In order to gain further insight, we asked respondents to estimate the proportions of business decisions that were driven by data as opposed to by human instinct. The results of this question are shown in Figure 1, which shows that the greatest number of our respondents (31 per cent) said that between 40 and 60 per cent of their decisions were informed by data. Approximately one quarter of respondents said that between 60 and 80 per cent of their decisions were data-driven. A total of 40 per cent of respondents were distributed across the 0 - 40 per cent data-driven decisions bracket.

However, as Figure 1 also shows, our respondents expect this situation to improve a little in the next three years, with the majority expecting in excess of 80 per cent of their decisions to be data-driven increasing from six per cent to 16 per cent, and the proportion of respondents in the sub 40 per cent categories falling slightly to approximately 38 per cent. This rather modest increase in expectations for more data-driven decision making is all the more surprising in the light of the dissatisfaction our survey found with the present situation. When asked, "Are you satisfied with the present ratio of data-driven decision making to gut-feel-based decisions?" only 30 per cent of respondents declared that they were. Exactly half of our respondents stated that they were not happy and a further 20 per cent were simply unsure. This hardly constitutes a ringing endorsement of the status quo.

Fig. 1 : What percentage of decision-making across your organisation, do you estimate, is based on data rather than gut feel?
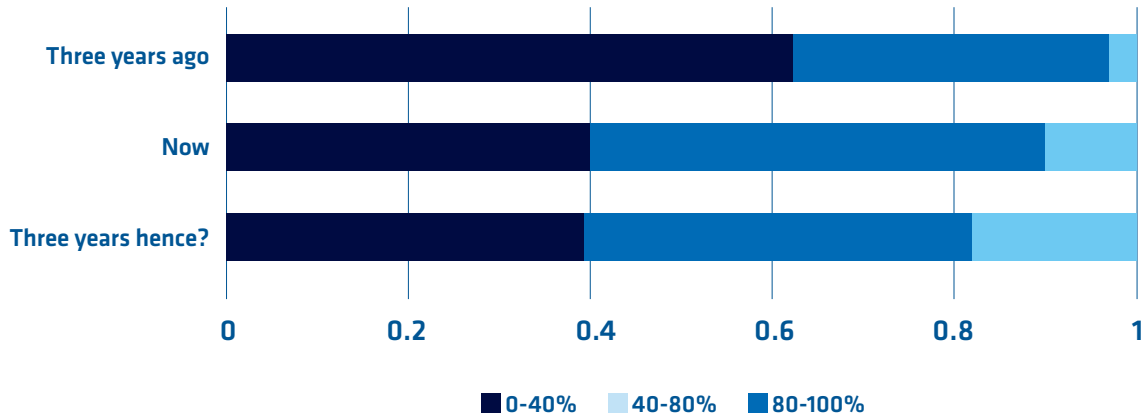

Whilst few people wish to expunge human beings from the decision making process altogether it is clear that among our survey respondents there is a desire for a greater role for data in the process. The ideal scenario is one where human experience and context can be overlaid onto a data-driven analysis.

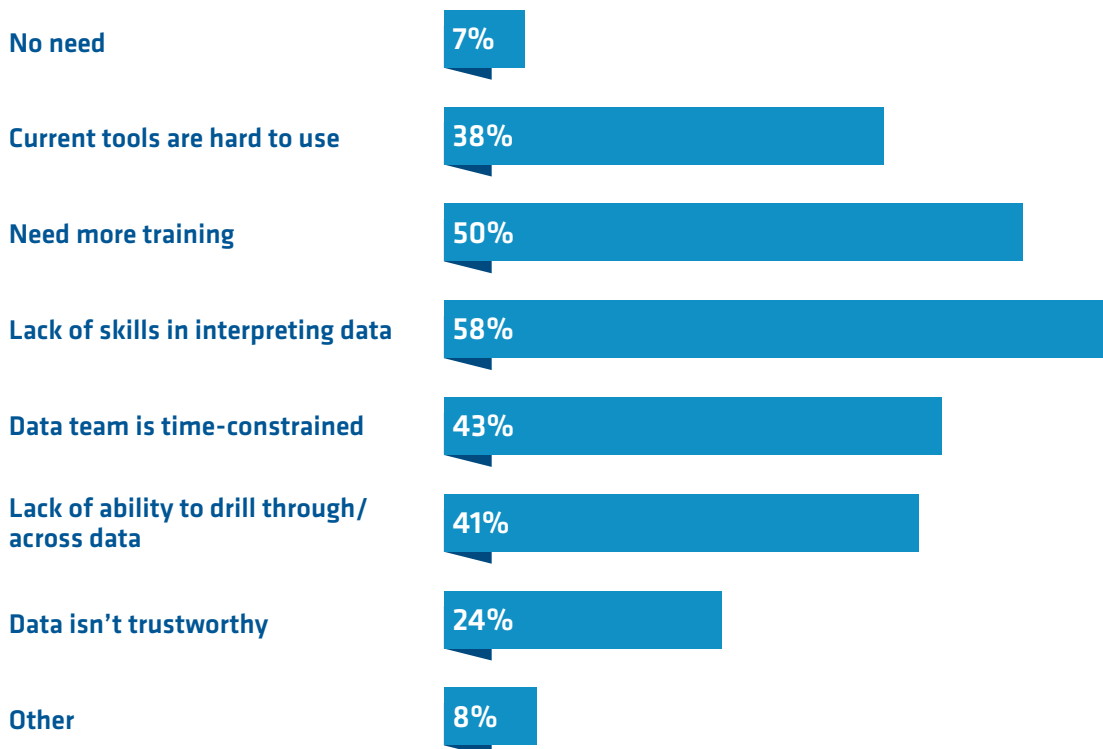# Barriers to Data-Driven Decisions

It would be misleading to infer from these findings that individual business decisions are either 100 per cent based on data or 100 per cent on human instinct. As is most often the case, the truth tends to lie somewhere between the extremes. Each decision will be informed by varying combinations of data and instinct – and the quantity and quality of data will vary. What is causing this variation? In order to establish why the level of satisfaction with the existing ratio of data-driven to human instinct-driven decisions is relatively low, we first have to ask whether people are actually using the tools at their disposal. Figure 2 shows answers to the question, **"In your organisation, of the people who have access to data analytics what percentage are actively using it?"** It is obvious from an even cursory glance at Figure Two that in many organisations, individuals are simply not utilising the BI resources available to them. In only one fifth of respondents' organisations, are over half of people who have access to data analytics actually using them. In a large majority of organisations, somewhere between 10 and 50 per cent of those with access to data analytics are using them.

Fig. 2 : In your organisation, of the people who have access to data analytics what percentage are actively using it?

0 to 10%	11%
10 to 25%	34%
25 to 50%	34%
50% +	20%
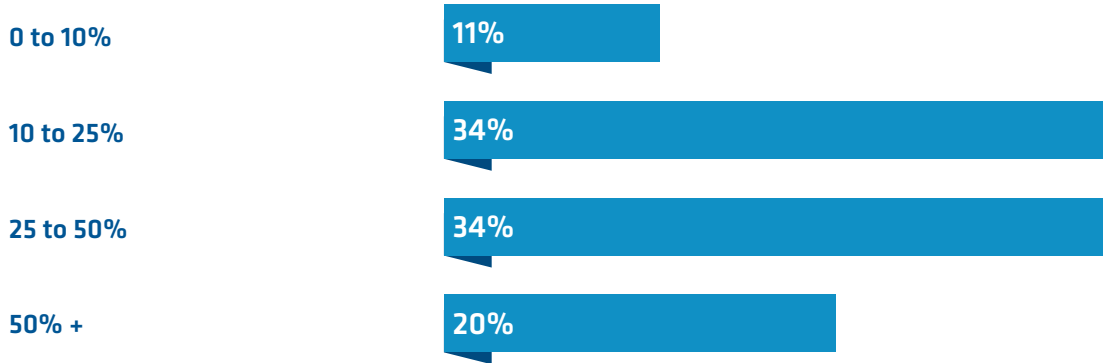
Computing also asked, "**In your opinion, is your existing data analytics/BI toolset being utilised to its full extent?"** 57 per cent stated that it was "probably not" and 20 per cent said "definitely not." Only four per cent of our respondents could state with confidence that their analytics/BI resources were definitely being used to their full extent. What is causing the widespread under-utilization of these resources? Figure 3 provides some insight as to why this is the case. The single most frequently occurring reason was a lack of skills in interpreting data. This was given by in excess of 58 per cent of respondents. Only slightly fewer respondents (50 per cent) stated that users required more training. 43 per cent said that the data team suffered from time constraints and 41 per cent said that users lacked the ability to drill through/across data.

Fig. 3 : What is stopping more people from making use of analytical tools to drive decision-making? (Choose all that apply)

No need	7%
Current tools are hard to use	38%
Need more training	50%
Lack of skills in interpreting data	58%
Data team is time-constrained	43%
Lack of ability to drill through/ across data	41%
Data isn't trustworthy	24%
Other	8%

Business Analytics in the Machine Learning Era

All of these responses point to the fact that existing BI tools tend to require certain specialist skills to use – or are just plain unintuitive. This fact alone makes them unsuitable for use in an environment where speed of decision making can be crucial to the fortunes of business organisations. The problem lies in the transition from data to insight. Data teams or indeed business users may spot patterns in data but in order to derive actionable insight from that data, an investigation becomes necessary. This exploration is often conducted based on a "hunch" or hypothesis of what the answer might already be – which means that any subsequent decision cannot be said to be data-driven.

In addition to the fact that you need to know how to ask the right question, there is also the skills barrier to navigate in order to transform data into insight. The fact that lack of skills was the most commonly given reason for not making the most of existing tools sets indicates that the deficiency in data analysis is hurting businesses. Data analysts – although they may well be able to conduct investigations using the full extent of the resources available to them, sometimes lack the ability to translate the outcome of these investigations into a narrative that a non-technical audience can then take insight from. Exacerbating the issue further is the fact that all of this takes time. Results of investigations will inevitably take time to wend their way back to a business decision-maker, with the result that the analysis is only available sometime after the decision was actually required.

In order to gain greater insight into some of these barriers we asked respondents to our survey to indicate the extent to a series of statements were true or untrue. Two statements tied for first place as to which was the most strongly disagreed with. The first was, "It's easy to drill down into data using our analytics tools, even for non-techies." 39 per cent marked this either as a one or two with one being completely untrue. Once again, the requirement that businesses have of their BI tools to be able to drill down into areas of interest without specific coding skills being applied is brought into sharp focus. The second statement that was considered the least true was, "Predictive analytics is working for us – we trust the recommendations." The extent of disagreement with this statement suggests strong levels of dissatisfaction with the analytical status quo, as does the extent of disagreement with the statement, "We are an insight-driven organisation."

On the flipside, the statement that our participants most strongly agreed with was, "We are analyzing data from more sources these days." This finding is very much in line with expectations. Data on the external business market environment cannot be left out of BI if decisions are going to be data-driven. However, much external data does not come in a neat, machine readable format and data analytics skills are usually required to repackage it in a visual format and combine it with data from other internal and external sources to answer the questions required.

Another statement to register strongly as true was, "Our analytics processes are trustworthy and free from bias." This finding is positive, but begs an obvious question in response. *If analytics processes are trustworthy why aren't they being used to make more decisions?* This leads us back to the issues of usability, visibility and skills deficits. The tools for data-driven decision making might be there – they're just very difficult for decision makers to actually use.

Data Governance & BI

It is difficult to think of a time when data governance has had a higher profile within commercial and public organisations but on the whole, our survey respondents were guarded on this issue. Approximately five per cent said that "our data is clean and of good quality" was completely true of their organisation. A further 34 per cent rated it as slightly less true. However, the largest proportion of respondents to this question – 42 per cent – chose the middle response – neither true nor untrue. 28 per cent said it was either fairly or completely untrue. When questioned about investment priorities a large majority of respondents – 65 per cent stated that their top priority was getting data quality and master data management correct. Taken together these findings suggest that data governance as a whole in many organisations is, at best, patchy.

Is it possible that data governance has been affected by self-service BI? We asked respondents **"Do you believe that self-service analytics pose a challenge when it comes to data governance?"** Fewer than one quarter of our respondents – 23 per cent in total – said that in their view the answer was either probably or definitely not. A huge 53 per cent said that maybe they did and the final quarter of respondents said that they definitely did.

It is easy to see why self-service BI and data governance seem mutually exclusive. The emphasis in just about every organisation has, necessarily, moved away from centralization of all aspects of data and all of its management, security etc. to a distributed infrastructure where control is decentralized to enable greater agility and empowerment of the end user. In many organisations this democratization of data has led to issues with governance. This has occurred with self-service analytics for several reasons. The first is that Workbook solutions create new copies of data (mainly so the end user is spared the trouble of having to ask for it elsewhere) and the more copies of data that exist, the more likely mistakes are to occur. Issues can also arise due to the need for third party-plugs that are often required, and scalability beyond the initial workgroups can be hampered.

It would seem that companies are being challenged from multiple angles. On one hand, there is clear evidence of a bottleneck in the process of analysis where users try to establish why a particular event has occurred. End users may be able to see *what* happened but when the user invariably asks *why*, further data modelling from data analysts is required in order to investigate. At the same time, empowering end users with self-service BI to try and sidestep this issue is likely to lead to issues with data governance and possibly security.

Automating Insights

The challenges being faced by our respondents in trying to optimize BI have partly come about because in many organisations, the understanding of the different areas necessarily involved in the process, such as platforms, automation, the devices and the importance of data governance, have all been evolving separately. All of *Computing*'s research into data analytics conducted throughout 2017 pointed to the fact that these paths are now converging.[1] Did our respondents agree? The short answer is yes. When asked, **"When do you think machine learning will have a significant impact on your organisation?"** the majority of respondents believed that it would in two to three years' time. This is certainly consistent with the idea that we are standing at the beginning of an era defined by machine learning (ML).
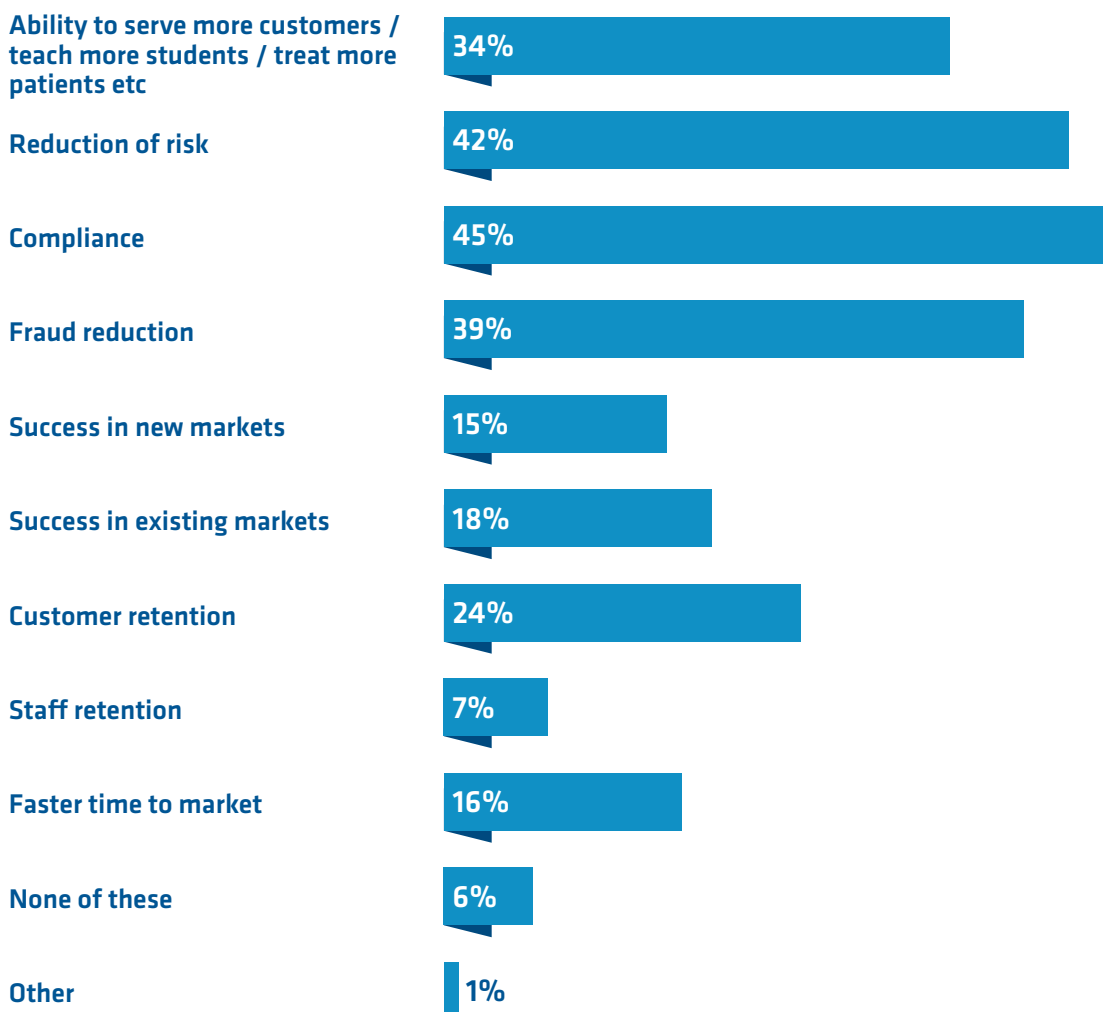
Computing asked, **"Which of the following business outcomes are likely to be achieved through the use of Artificial Intelligence-driven analytics in your sector?"** and asked respondents to

---

[1]   Computing Big Data & IoT Review 2017

choose up to three. Figure 4 illustrates some of these possibilities. Compliance was the most commonly selected outcome, underlining the importance of good governance yet again. Related factors also scoring very highly were (in ascending order of importance) the reduction of risk, a reduction in fraud and the ability to simply do more – serve more customers / teach more students / treat more patients etc. These are all relatively modest aspirations given the potential of ML. In an attempt to delve a little more deeply into the degree to which organisations are getting to grips with the possibilities of ML we also asked, **"Which use cases of analytics powered by ML/AI could provide the quickest benefits for your organisation or sector?"** and asked participants to respond via a free text box. A relatively small amount of individuals participated here – and around 35 per cent of them admitted to not really knowing what ML could do for them. Other responses focused on operational benefits such as, "procurement of materials and distribution, better prediction of financials, i.e. cashflow," "predictive analytics for equipment maintenance," and "understanding shop floor process data." Only one respondent focused on growth with a response of "where to focus our expansion." The response to these questions suggests that our respondents find the potential of ML for their businesses difficult to comprehend.

Fig. 4 : Which of the following business outcomes are likely to be achieved through the use of Artificial Intelligence-driven analytics in your sector?

Ability to serve more customers / teach more students / treat more patients etc	34%
Reduction of risk	42%
Compliance	45%
Fraud reduction	39%
Success in new markets	15%
Success in existing markets	18%
Customer retention	24%
Staff retention	7%
Faster time to market	16%
None of these	6%
Other	1%

Many of our respondents have told us that they are at an impasse with BI. In order to establish why events occur, investigation by specialists is usually necessary in order to derive insight from data. It takes precious time and uses equally precious resource. If businesses take the self-service BI route there is likely to be a negative impact on data governance. Analytics powered by ML enable organisations to break out of this limiting position. ML can be utilized to discern patterns within a variety of data sets and react accordingly by suggesting a course of action. With every suggestion or action it can learn and fine tune its variables according to the results. If ML is combined with a web-based, visual analytics platform, decision makers can use the learning capabilities to uncover anomalies, trends, patterns, and relationships within the data without coding skills being required to find out why events happened. Because results can be presented visually the need for a specialist interpreter of the results becomes less crucial. The pitfalls relating to data governance when end users try to undertake data analyses to discover why events happened are also sidestepped because data preparation can be performed centrally by data specialists, in a single environment, rather than by individual users at a desktop level.

ML should not replace skilled data analysts and scientists – it should be used to assist and automate some of the more mundane aspects of their role. Data scientists will gain a more granular data set to work with and a set of contextual tools to fine tune algorithms while analysts gain more time to define tasks, curate findings and find better ways to share them.

Where do ML powered advanced analytics reside on the priority lists of our respondents? Given that a majority of our respondents expect ML to start having a meaningful effect on them within the next two to three years we would expect to find it around the middle of the list – and this is broadly the case. For the approximate 13 per cent of respondents who believed that ML was already affecting them or would do so in the next year, we would expect to see ML powered analytics fairly high on this list. In fact, approximately 17 per cent of respondents stated that ML analytics were either high or very high on their to do list. 39 per cent came in at a mid-level position. However, for approximately 43 per cent of respondents, the ranking of how ML analytics is affecting them was either low or very low.

This position may well be subject to change over the next 18 months. When asked, **"What do you think might be the cost of NOT investing in smarter BI tools and processes that use machine learning and AI?"** the most popular answer chosen by respondents was "loss of competitive edge/market share." In excess of 65 per cent of respondents chose this answer. It may well be that we see organisations moving to embrace ML powered BI as the costs of inaction become clearer over the course of the next year or two.

Conclusions

The role of data in the process of business decision making has become ever more important as volumes of data, as well as the sources from which it is generated, have proliferated and diversified. Decisions made in ignorance of data external to organisations, or on the basis of data of dubious quality, are less likely to be winning decisions than those made in full knowledge of the facts. The concept of intuition being more powerful than facts is appealing for many, very human reasons, but those who lead the organisations are usually aware of the limitations of human instinct and wish to make decisions on the basis of data as well as their own feelings.

Our research has shown that whilst the desire for data-driven decision making and an insights-driven culture is prevalent in UK organisations there is a considerable gap between the ideal and the present reality. Only one third of respondents to our survey said that between 60 and 100 per cent of the decisions made at their organisations were driven by data. This is probably why only 30 per cent were satisfied with their present ratio of data-driven decision making to gut-feel-based decisions.

Business Analytics in the Machine Learning Era

This dissatisfaction is underpinned by the fact that in 80 per cent of organisations, fewer than half of the people who have access to BI tools are actually using them. Only four per cent of our respondents could state with confidence that their BI/analytics resources were definitely being used to their full extent. There are a number of related reasons for this under utilisation. Firstly, there exists a significant skills shortage in the data analysis/data science arena – skills which are essential for interpreting conventionally generated BI data. The shortage of specialists is particularly acute because they tend to specialize in very specific areas such as risk analysis. A second, related problem is a need for training in end users which would enable them to drill through and into relevant data – training that overstretched teams are struggling to carry out. This situation encapsulates almost every shortcoming that existing BI toolsets seem to have. End users might well be able to identify what has already happened but in order to find out why they have to go back to specialists. Furthermore, they often tend to do so based on an idea of what the answer might already be. In summary, the process is slow, unwieldy and subject to bias. There are short falls in usability, visual depictions of data and the skills required to take that data out of its model and convert it into insight.

The limitations of relying on data teams to prepare and present data, has led some organisations to adopt self-service BI. This reflects the move away from data centralization to the greater empowerment of individuals and distribution of data – the democratization of data. However, the same rules apply to democracy of data as they do to democracy as a whole. On balance, it's better than any of the alternatives but it's not perfect. The necessity of the creation of new copies of data for self-service BI means that the desired state of "a single version of the truth," cannot, by definition, exist. This is why in excess of three quarters of our respondents stated that they believed that self-service BI either probably or definitely posed a challenge to good data governance.

Traditional BI has led organisations to somewhere between a rock and a hard place. On one hand, reliance on a centralized architecture does not empower the end user and requires the input of skilled data specialists to provide insight as to why a particular event has occurred – and translate the findings back into a language that non-specialists can use. Workbook based BI empowers users but at the likely expense of good data governance and security – neither of which are "nice to have."

The way to break this deadlock is to use machine learning to enable the jump from data to insight to be executed by the end user – whilst keeping the actual data preparation and spinning up of dashboards in the realm of data teams. When business users spot something in their data they wish to investigate further they can do so on their own terms. They can perform their own analysis and get answers when they need them. The object of the exercise is not to render data teams obsolete but to automate the more mundane aspects of their role and allow them to focus on their own areas of expertise such as the preparation of data from multiple sources and the mining of data for unknown unknowns. End users can focus on why an event occurred but a data scientist has time freed up to focus on framing new questions. Equally, the more human aspects of decision making are not redundant. Context can be combined with data for the most informed decision.

The subject of ML is often framed in extremely negative terms – as every technological innovation has been since the Industrial Revolution. However, the really exciting aspect of ML in data analytics is not its potential to replace human decision making but to work in partnership with it. Until now self-service BI has effectively tried to turn end users into data analysts – and our research here suggests this approach has failed. ML in BI allows each employee to do what they do best. The roles of data specialists will evolve so that they can add even more value to their employers and business users will make better decisions that are driven by data but still made by a human being.

About the sponsor, Yellowfin

Yellowfin provides a Business Intelligence (BI) and analytics platform aimed at solving real enterprise analytics challenges and helping business people understand not only what happened, but why. Founded in 2003 in response to the complexity and costs associated with implementing and using traditional BI tools, Yellowfin is an intuitive, 100 percent web-based reporting and analytics platform. More than 25,000 organisations and more than three million end users across 75 countries use Yellowfin every day.

For more information:

Visit: www.yellowfinbi.com