McKinsey&Company

How to win in the age of analytics

January 2017

There's greater potential in big data. What's ahead as the field matures?

Since the concept took hold, big data has made big waves. The field of analytics has developed rapidly since the McKinsey Global Institute (MGI) released its landmark 2011 report, *Big data: The next frontier for innovation, competition, and productivity*. But much value remains on the table as organizations wrestle with issues of strategy and implementation. In this episode of the *McKinsey Podcast*, MGI partner Michael Chui and McKinsey senior partner Nicolaus Henke speak with McKinsey Publishing's Simon London about the changing landscape for data and analytics, opportunities in industries from retail to healthcare, and implications for workers.

Podcast transcript

Simon London: Welcome to this edition of the *McKinsey Podcast*. I'm Simon London, an editor with McKinsey Publishing. Today we're going to be talking about data analytics and how organizations can use the unprecedented volume of data at their disposal to transform industries, create new business models, and, frankly, make better decisions across everything they do. Joining me here in London to discuss the issues is Nicolaus Henke, the global leader of McKinsey Analytics and chairman of QuantumBlack, an acquisition McKinsey made in 2015. And joining us from San Francisco is Michael Chui, a partner with the McKinsey Global Institute.

Nico and Michael are among the coauthors of *The age of analytics: Competing in a data-driven world*, which is a new McKinsey Global Institute research report. If we pique your interest with this podcast, you can download the full report from McKinsey.com. Nico and Michael, thanks for joining me today.

Nicolaus Henke: Thank you very much. Delighted to be here.

Michael Chui: Thanks. It's a pleasure.

Simon London: Before we get into detail on the latest research, I think it might be helpful to take a step back and clarify what we mean in terms of the age of analytics. Cynics would say, "Come on. Companies have been collecting and analyzing data forever, pretty much." So what's really new here? What's driving the data-analytics revolution?

Nicolaus Henke: Thanks, Simon. It's a great question. We think there are three things that have really changed. The first thing that has changed—simply, there's loads more data. Believe it or not, about 90 percent of the world's data existing today didn't exist two years ago. Ninety percent. The second one is we simply have computing power, with the cloud and connectivity, that is much, much lower cost than it was ever before. So we can compute more.

The third is that by leveraging machine-learning techniques, we can analyze much more. To give you an example, in the past it took a statistician to come up with a potential hypothesis for regression, and it took a day or two. You could make, maybe, three a day. With these new techniques, you can add all these things together. We can, in our normal work, do hundreds of millions of calculations a day, which obviously increases the granularity of our work.

Michael Chui: If I could just build on that idea, while all those trends have come together, one of the things that's happened between the time we published our big data report in 2011 and now is the degree to which CxOs and senior leaders have started to understand that this is changing the basis of competition in individual sectors.

While we've discovered there's a lot more work to be done, we've seen an awareness, at the executive level, of the importance of using data and analytics in order to compete and, increasingly, to make decisions in very different ways. For example, people are conducting experiments rather than just basing judgments on the experience they've had in business.

Simon London: Michael, as you mentioned, in 2011 we published a big piece of research flagging the transformative potential, I think it's fair to say, of this new wave of data and analytics. Five years on, how much of the potential you identified back then has been realized? What does the report card look like?

Michael Chui: To be honest, the progress has been mixed. We have seen some industries and some domains—such as location-based services and, to a lesser extent, retail—that really have moved the needle. One of our observations is that those are places where we've seen digitally native companies create competition. And that really forced the industry forward. However, there are a number of other industries—whether it's the public sector, healthcare, or even manufacturing—where some progress has been made. But, honestly, with regard to the total amount of value that could potentially be captured, there's a lot more work to be done.

Less than 30 percent of the value that we identified has been captured. That means we still believe, in fact, that value is there to be captured. We've identified even further ways in which data and analytics can be used to capture value. But there are a number of obstacles that need to be overcome in order for that value to be captured in those industries.

Simon London: The obstacles that need to be overcome, are they primarily technical? Are they organizational? What's getting in the way here?

Nicolaus Henke: The main obstacle is organizational. In order to really get the value from the data, you need to do five things at the same time. And you need to do them all. If you do not do one of them, you basically lose out on the value.

We have already established that capturing the data is one of them. And doing mathematical models is another one. Now, doing all of these things in itself doesn't create any value. The third thing, therefore, which needs to happen is you need to be very thoughtful on the source of the value. What kinds of use cases are you trying to drive? If you are doing a lot of analysis and modeling without being really focused on the business value, you lose out.

Now let's assume you do these three things. So you are running a bank. You found the top 30 use cases, for example, in revenue management, next product to sell, and so on. The fourth thing that needs to happen is you need to embed it into your processes. Large companies have hundreds of thousands of employees. If you don't embed the result of the findings into the processes, essentially nothing will change.

Finally, there's capabilities. You need to build the capabilities to use all these results in order to figure out how to make decisions in a different way. But you also need to have the capabilities to analyze data and do all of these things first of all.

Michael Chui: I think what we found, in many cases, is that these companies have started to invest and in fact have gotten to the point of doing the modeling and deriving some insights. But it's the organizational difference. The change has to occur between discovering an interesting insight and being able to scale it to the size of an organization—to really embed it within the daily processes of an organization—so that it moves the needle in corporate performance. Again and again, in many companies, we've seen that's where a lot of the gap has occurred. And that does require, as Nico said, really moving across all five of those different components.

Nicolaus Henke: You're right, Michael. And just to illustrate that point, you just spent a couple of days with 200 of the world's leading data scientists. We were talking about the topic of how to interact with a CEO and with the executive team. They were all quite unhappy with how that is going, because they feel that the executives don't quite understand what they're doing. When they don't see the executives, they feel that the data scientists are not focusing on the

key business problems. So there is a translational task. Indeed, we at McKinsey are training 3,000 of our colleagues to become translators, essentially—to know the business problems deeply, but also to understand the data-science and the computer-science aspects of it, so they can tie these things together on behalf of our clients.

Simon London: Back in 2011, we famously predicted quite a big talent gap for real hardcore data scientists. It sounds like that may still be an issue. But you also need this layer of translation, of translators. Is that right?

Michael Chui: That's exactly right. We did look in 2011 and hypothesized and analyzed a potential gap in terms of the number of people with deep analytical skills—the people we now call data scientists—that were being generated at the current course and speed, and how many we would need. We have seen that gap actually occur. Of course, the market has cleared. We're seeing more and more academic programs, more training programs to produce more data scientists. Yet we have seen data scientists' wages increase, which is an indicator of the supply-and-demand dynamic.

Going forward, we'll continue to see this need for more data scientists accelerate. At the same time, as Nico described it, there is also another role. And it's a [need for a] much larger number of people that are able to take the domain knowledge of an industry, of a function, and know enough data science in order to help translate that, to make it consumable by the rest of the organization. We're talking about millions of people here that we'll need in these types of roles.

Simon London: There's something else that I know is a very big deal for companies—this whole issue of data strategy, of actually figuring out what data you need to satisfy the use cases, where you're going to get it, how you're going to govern it. Do you want to say a little about that, Nico?

Nicolaus Henke: We think it's one of the most foundational enablers, close to the importance of talent. And at the end of the day, when you prioritize what kind of areas you want to focus your business improvement on, it is a good time to think about your longer-term master data model. What is the kind of data I'd like to receive? And then you think about how you actually get those.

For example, one bank has gone through a two-year exercise, now, to really build an enterprise-wide data lake. It is one of the few banks in the world where they have that. They created a war room of about 150 people. They identified a number of what they called "golden" sources—you know, where the data comes from. And they went, golden source by golden source, to work with the business on improving the data quality.

To give you one example, they had very poor information, in that particular country, about the names of their customers. All their customers with middle names, last names, first names had three or four differences. The credit card would have [one] name. The bank account would

have [another] name. And the addresses were all kind of misspelled and so on. They had, on average, three or four different descriptions for each customer, which of course makes it hard to make sense of that particular data set.

So they went all the way to the business owners—like the people who opened the accounts, the people who were selling credit cards, et cetera—to make sure the processes with which data was captured were simplified and digitized. That helped them, over time, to get much, much higher-quality data and linkable data.

Simon London: The fascinating thing about this is that it is remarkably unglamorous work in many ways, right? This is not at the bleeding edge of data science. This is not machine learning. This is the real blocking and tackling of management, essentially. It brings home why, in a lot of companies and a lot of industries, the potential that's been realized is only 30 percent or less. Because a lot of this needs to be done before you can realize the value at scale.

Nicolaus Henke: Yes. We think the best way to do that is to begin with a first repository of integrated data, to begin to show the value in the first year, to begin with something, even if it's imperfect. And then you say, "Gee, if we get this and this additional data, or if we had slightly more clean data in this particular area, then we could lift the value we create to a completely new level." Then you take it in an iterative way from there. We think the leaders are doing it like that. They've never planned out, forever, one data strategy. But they have a vision of where they want to go. And they build iteratively to that vision.

Michael Chui: Most data scientists, nowadays, say that over half their time is taken up with data wrangling—just trying to solve some of these problems. But solving those problems is a prerequisite to capturing any value at all.

Simon London: A lot of what we've been talking about so far is applying data and analytics almost within the paradigm of existing businesses in proven organizations—optimizing and so on. Do you just want to talk a little bit, Nico, about what we're seeing out there and some of our favorite examples of really quite brand-new things?

Nicolaus Henke: What we are now seeing is essentially that data actually changed the borders between industries. For example, if you take telephone-ping data in emerging markets in Latin America, they are being used to improve the quality of underwriting of credit cards and credit risk. Because the telephone-ping data are much, much better predictors of certain behaviors. And with that, you can actually tell more than banks traditionally could and improve credit scoring a lot. The implications are vast because you essentially see value between the telecom industry and the banking industry shift. You almost ask yourself who's the right owner for making credit-risk decisions.

Michael Chui: I think another driver of this crossing of industries and these industry disruptions is what we sometimes describe as orthogonal data. Many times, organizations and industries have used data for many, many years. But what can cause disruption is a new source of data that allows either incumbents to drive forward in terms of their performance against the competition or, in fact, new players.

Insurance is a perfect case of that. It's analogous to the underwriting example that Nico described. But then you start to bring new sources of data in. Take, for example, telematics data—or, really, behavioral data about the organizations, people, or devices that you're insuring. Oftentimes, that can allow you to make a much more fine-grained risk decision in underwriting. But you can also make a pricing decision. Furthermore, not only can you make a pricing decision in insurance, but now you can actually help your customers manage their risks better.

I often joke that I only interact, if I'm lucky, with my auto-insurance company twice a year. It's not a great experience, either. I pay the bill. And worse yet, if I have further interaction, it's because I've had an accident, which again is not very happy. On the other hand, imagine an insurance company that provided me with data that said, "You drove very safely today." That can not only change the performance of the insurance product but also the types of interactions you have with your customers. And that can change the basis of competition in that industry. That's because of orthogonal data, because of new sources of data.

Nicolaus Henke: Another example is data-driven discoveries. In one situation, it took us about one week to be as smart as the whole history of clinical research in the world to predict who is going to go to a hospital within a month's time. Basically, using a very good national data set, we could come up with a model that predicted that as well as all the clinical research ever done had. It took another two weeks to, essentially, have a factor-of-three lift in predictions over all clinical research ever done, by linking orthogonally, as Michael was suggesting, data sources to this particular data set, which people hadn't connected before. For example, a feeling of loneliness is a great predictor of ending up in a hospital for elderly people. That's just one example.

Simon London: Nico, you mentioned machine learning. Machine learning and deep learning are sort of on the bleeding-edge technical side of this. Am I right to intuit that a lot of the things you're talking about now are advanced use cases, with machine learning at work?

Nicolaus Henke: Absolutely. The fundamental difference between those and traditional math is that in linear regression, you have a particular hypothesis and then go for the data, and then you find the correlation between them. With these techniques, the machine finds correlations for you. You then look at the output and try to interpret what you are seeing. The power of that is essentially caused by your being able to do hundreds of millions of calculations a day—not necessarily, you know, pursuing a particular hypothesis, but looking at a pattern in a new way.

Michael Chui: In a computer-science sense, one of the ways you might describe it is that it's the difference between programming a machine and training a machine to learn. They're some of the most cutting-edge, most exciting things we're seeing in terms of the use of data. We tried to understand where these types of techniques could actually create the most value.

We expected to find a Pareto curve, where 80 percent of the value might have come from solving 20 percent of the problems—that a lot of the value would be concentrated. What we actually found was the opposite, which is that there's potential for these technologies to really apply across the board. Every single one of the 120 industry problems we identified was identified by at least one expert, and usually by multiple experts, as being one of the top three problems that machine learning could help solve in that industry. So again, what we found was that this is a set of techniques with broad applicability to add value in every industry in the economy.

Simon London: I think it might be helpful to bring out some examples here. What are some of the things where you might not intuitively expect machine learning to have an application?

Michael Chui: We heard from an industry executive who said the three sexiest words in the industrial Internet are "no unplanned downtime." This is the idea of using predictive maintenance to fix something before it breaks. And what we've seen in large, complex assets, whether it's locomotives or whether it's pumps, is that if you get this continuous stream of data—a very detailed set of data, a large amount of data—and then apply machine learning, you can train it to try to discover when this machine is going to break. You can actually discover the signals that allow you to go fix something even before it breaks.

And that has huge amounts of value. Not only can you reduce the cost of fixing something, which usually is more expensive than the preventive maintenance itself, but you can keep that asset from breaking down. Then the trains can actually run. A factory can run. Usually, the benefits of fixing something before you break it have a lot more to do with the avoided cost from having something out of service than with the cost of repairing it itself.

By the way, healthcare you can actually just view as predictive maintenance on the human machine. It's so much more valuable to keep someone from having to go into a hospital—from going to an emergency department—than to try to heal the sick.

Nicolaus Henke: We were working with a company, a retailer in a very large city. It said, "We have a thousand outlets, and we basically feel we can't grow any further. Are there other opportunities to grow? Just help us understand where we can find more spaces, so to speak, to put ourselves."

With artificial-intelligence and machine-learning applications, we found that the stores that are located next to a laundromat, for a particular segment of people, would be highly, highly successful. And we found 850 new locations that the company had never thought about, based on that analysis. It is now heavily growing. So it's an incredible opportunity to link, in this case, geospatial data with things you wouldn't have thought about before.

Simon London: It's interesting. A lot of these use cases we're putting out there are just examples of how to sell more stuff, to make machines more efficient, with less planned downtime. You know, an obvious riposte here—is this going to make the world a better place?

Nicolaus Henke: We think so. There are other examples. It makes some prisons in the world safer places by reducing violence. Hospitals are finding at-risk patients.

For example, I've recently been to a kind of emergency room with 180 very sick people in it, all elderly. And the hospital uses a machine-learning algorithm to predict, of these 180 people, who needs how much intensive care and who needs minute-by-minute supervision versus who needs much less supervision. Because they can target more senior staff to these very sick people, they can actually keep them alive. Their success is a 36 percent lower admission rate. People essentially get turned around in the emergency room and sent home, versus traditional models—which is a resounding success. There are all sorts of use cases where data exist about human behavior.

Michael Chui: Another case where machine learning can greatly improve the human experience is the ability to understand natural language. I'm a former artificial-intelligence researcher. For a long time, it was so hard to try to get machines to understand spoken language. They're not perfect at it now. But we've seen great advances through using more and more data and machine learning in order to better understand voice.

That can enable all kinds of people—say, the elderly, where it might be more difficult for them to use a traditional interface, to be able to look at a small, mobile screen and type. [They can simply] speak into the phone and ask for directions to a place, or to have their phone just call the person they want to call.

Simon London: The obvious comeback is that this makes me think quite a lot of jobs could be replaced, as well. In customer service–type jobs, clearly, natural-language processing is part of what you do. What do we think about the labor-market impact of all of this?

Michael Chui: One of the things that we would note is that as this technology continues to increase in its ability, it does enable more and more activities we currently pay people to do in the economy to be automated.

Two other things that we've discovered about these technologies. One is that it will actually take quite some time for the activities we currently pay people to do in the economy to be completely automated. So there's time to adapt as we adopt. But there's no time to wait. We actually have to start understanding how these technologies might be used in the economy. The other thing that we've discovered is that, while we have time to adapt as we adopt, it doesn't look likely that we'll actually have a surplus of labor.

In order for us to have the type of economic growth that we need, in both the developed as well as the developing markets, not only do we need all the machine learning that we can get, we need everybody to be working, as well. We'll need to make sure that, as people are displaced by technology, we find productive things for them to continue to do in the economy. We need to find things for people to do in order to have the economic growth that we need.

Nicolaus Henke: There are a number of areas where this really can help to solve problems that otherwise couldn't be solved. For example, in healthcare, if the trend of the past 80 years would continue—where healthcare has outgrown the economy by two percentage points a year, roughly—then by 2100, 98 percent of the US economy will be for healthcare. Now, that obviously cannot happen. The need may be there, but some other things need to be found in order to deliver all that. That's where robot sensors, automation, and big data monitoring can make healthcare much better and more sustainable.

Michael Chui: All that being said, while we think data and analytics can drive tremendous value for companies and can drive great benefits for individuals, there are real risks. And there are things that we'll need to manage. People have an interest in their own privacy. We'll need to try to find that balance to understand, you know, when people can value the use of data and analytics and when they will want to think about uses of data that they actually don't want to have happen.

Cybersecurity is a huge issue, as well. We think there's great value in combining data from multiple sources. But if that data or those analytics are used in ways or by actors, whether they're criminals or others—that's a risk that needs to be managed.

Simon London: There's a question that you do see written about a fair amount in the media. I think it's a legitimate concern. If we have algorithms making decisions about more and more aspects of our lives—whether it's how we're deployed in an organization, for example, or the level of healthcare that we might be offered—how do we know that those algorithms are constructed in a way that is fair and transparent?

Michael Chui: A couple of thoughts about this. First of all, again, the use of data and analytics itself doesn't mean you're going to get good answers. You have to use it well. And one of the things we often find is a problem is that the underlying data set you use can sometimes have issues in itself.

You have to understand the data. We've seen multiple examples of this being an issue—Internet of Things data being used, for instance, in Boston, in order to identify where there are potholes by using the accelerometers within smartphones. Well, one of the issues there is who has smartphones. Again, that biases the data toward places where there were simply more sensors looking for those types of potholes. Unless you understand the provenance of data, unless you understand the metadata, as you might describe it—the data about data, how it's collected, what are the underlying assumptions behind that data—you are likely to discover that you have issues there.

One of the biggest problems that we find now is model opacity. What do you do when this extremely complex machine-learning model seems to perform very well, but it's difficult to figure out how it discovered the things that it discovered? And then we actually find some regulations where you're not allowed to use these types of models unless you're able to explain them. Those are going to be some of the challenges going forward.

Nicolaus Henke: Exactly right, as Michael was saying. At the end of the day, machine learning is pattern discovery. It discovers patterns that are shown to have been true in data. If you then act on those rules, you first need to assume that these patterns are going to be consistent in the future. That's why machine learning is frequently not applied to problems under true uncertainties—for example, investment problems. There are certain types of investment problems these techniques will not help you much with, where heuristics are much better.

Then there are other problems where the system counteracts. In human performance management, when an organization finds out how, essentially, performance is measured, that has an implication. That's sometimes why models age—not just humans age, but models age as well. You need to readjust them all the time.

Simon London: That's all we have time for today. Thank you very much, Nico Henke, here in London. And in San Francisco, we thank you, Michael Chui, for joining us. To download the report, *The age of analytics: Competing in a data-driven world*, please visit us on McKinsey.com. ◻

Michael Chui is a McKinsey Global Institute partner based in McKinsey's San Francisco office, and **Nicolaus Henke** is a senior partner in the London office. **Simon London** is a member of McKinsey Publishing and is based in the Silicon Valley office.

Copyright © 2017 McKinsey & Company. All rights reserved.