

An Overview of Big Data – by Jackie Down

Many companies today still do not have a Big Data strategy. Many senior leaders have yet to grasp the significance of what seems to be a purely technology wave. But the hype is hiding the emergence of a new opportunity for basing management decisions on objective insight and unprecedented detail.

Despite the many 'What is Big Data' articles, no consistent definition has so far emerged. But this doesn't matter - what is more important is the potential for this emerging technology to transform the way management is performed in large enterprises and to



change the battleground for the hearts and minds of consumers. At the end of the day, organisations want actionable "Insight" from their data.

Even as organisations are embarking on big data initiatives, many still have several vision and strategy questions regarding how to drive the most value from these vital projects. Organisations are looking to big data to provide an enhanced experience for their customers, reduce costs and enable an information-driven culture.

Big Data can be described as information assets that demand new, innovative forms of processing for enhanced decision making, business insights or process optimisation. But it is characterised in many different ways. There is no predefined threshold of data volume to indicate when an enterprise has actually reached "big data" status. This varies from company to company - and really comes down to the business not being able to achieve its objectives without analysing a broader range and different types of data.

Predictive Analytics and Machine Learning

Predictive Analytics

Surprises are usually not considered a good thing in business! Business leaders prefer accurate business models and forecasts to support their decision making. Predictive analytics is the methodology that makes greater accuracy possible – and in a shorter time.

Predictive analytics uses "historical" data – via data mining, statistical modelling and machine learning – to try to predict what will happen in the future. Now that technology enables us to cheaply process large volumes of very recently collected "historical" data we can operate these models in near-real-time. This means that companies now have the ability to make faster business decisions based on these real-time forecasts. This explains why there is such interest: after all who wants to always be

★ BigData 4Analytics

saying in hindsight, “I should have taken a different decision”? We always want to be anticipating what’s going to happen next.

Predictive Analytics attempts to answer the question “what might happen in the future?” In common usage, Predictive Analytics typically applies more advanced classical statistical techniques such as linear regression to answer a question such as “If I increase my advertising spending by 10%, by how much will my sales increase next quarter?”

Predictive analytics does not provide guarantees. Instead, it is all about better quantifying the probability of different outcomes - fast enough to take corrective action in real-time. These concepts of increased likelihood and timeliness are what make it so enticing to apply it to decision making. Making faster, more educated decisions, and taking quicker action should be at the heart of every business strategy.

Components of the Predictive Model

Data: Lack of good data is the most common barrier to organisations seeking to employ predictive analytics. To make predictions about what customers will buy in the future, you need to have good data on who they are buying (which may be from a loyalty program, or at least a lot of analysis of their credit cards), what they have bought in the past, the attributes of those products (attribute-based predictions are often more accurate than the “people who buy this also buy this” type of model), and perhaps some demographic attributes of the customer (age, gender, location, household income etc.)

Statistics: This is the set of mathematical techniques, ranging from basic to advanced that are applied to the data to derive inference, meaning, and insight. The most common statistical technique used in predictive analytics is linear regression, which simply put is the iterative process of selecting and testing the impact of variables on the outcome.

Assumptions: these are another key factor in any predictive model—the assumptions that underlie it. Every model has them, and it’s important to know what they are and monitor whether they are still true. You have to review these assumptions because things change over time.



Predictive Analytics versus Machine Learning

So where does machine learning fit with Predictive Analytics? Well the fundamental difference is that the predictive analytics model relies on a human expert to formulate and test the relationship between cause and effect, i.e. the hypothesis that advertising is a driver of sales.

Machine learning flips this process on its head; it starts with the outcome, what were my sales and then teaches a computer to automatically uncover the factors that are driving the particular

BigData 4Analytics

outcome. These relationships can be incredibly complex, including hundreds of possible causes, interactions, and non-linear responses. If done correctly, the result is a far more accurate predictive model, which has the ability to automatically adjust and improve over time, thus allowing you to predict who will buy your product.

With rapidly changing and growing volumes of data, you need fast-moving modelling streams to keep up with the pace - machine learning allows you to do this. Typically, a human can create one or two good models a week, but with machine learning thousands of models a week can be created.

Summary

It's pretty amazing that we can use analytics to predict the future! All we have to do is gather the right data, use the right type of statistical model, and be careful with our assumptions!

Businesses today need to identify market trends, understand customer behaviour, tackle inefficiencies sooner, make sense of the explosion of data to stay ahead of the competitive curve, and make impactful and smarter decisions that align with company goals. By using predictive analytics & machine learning, they can get one step closer to their goals while understanding the detail to make better and faster decisions.

For more information contact Jackie at jackie.down@bigdata4analytics.com