



Can Sales Forecasting Survive Big Data?

"Beware yon Statistician for he has a lean and hungry look."

Big Data and Big Data Analysis are terms we hear a lot these days, and as a matter of fact it's almost impossible to avoid them. Just about every day we read an article or announcement about how Big Data solved another new problem or has provided new insights into old problems. It seems that Big Data is the answer to everything.

My particular interest is sales forecasting; for me and many like me, the question is; "Can Big Data solve the problem of perpetually inaccurate sales forecasts?"

At first glance the application of Big Data analysis to sales forecasting seems like a natural fit. There exist all sorts of sales related data on wins, losses, deal slippage, past forecasts, overall performance, etc., this information is just crying out for some sort of analysis. The result would be the holy grail of sales management; an accurate forecast. But according to CSO Insights and others, sales forecasting accuracy stubbornly remains below 50% and has done for the last number of years. It's hard to imagine that with all the sophisticated analysis tools and CRM systems available today that forecasting accuracy should remain at an anaemic 50%, but it's true. You may as well flip a coin and save all the time and grief that the sales force expends on creating a forecast. And they spend an inordinate amount of time creating those forecasts.

Every week we see articles about companies promising a forecasting accuracy of greater than 90% through some new and innovative application of AI or Predictive Analytics to Big Data; whether it's Watson from IBM, Einstein from Salesforce, Azur from Microsoft or any of a myriad of start-ups.

It begs the question; can we apply these high-powered statistical tools to all this sales-related data and come up with an accurate sales forecast? The answer is a clear and unequivocal 'it depends'.

One of the fundamental requirements of any statistical analysis, using any of the many statistical methodologies available, is that all events are the same. Put another way $n(i)=n(j)$. If the events are not the same, ($n(i)$ is not equal to $n(j)$), then statistics will not work. After all, you can't compare apples and oranges, even though they're both fruits, they grow on trees, they're round and they're about the same size.

And this is where the 'it depends' comes from.



Roughly speaking, sales can be divided into two categories; transactional sales and non-transactional sales. Transactional sales tend to be random, short sales cycle and low value (cost); retail sales are a prime example of transactional sales. Non-transactional sales, which have many other names such as complex sales and consultative sales, tend to be long sales cycle, high value, non-random and expensive. Business to Business (B2B) sales are typically non-transactional sales and usually involve a field sales organization. Transactional sales can be done over the internet while non-transactional sales are done face to face with a sales rep and a prospect.

From an analysis perspective, the most important thing we can say about a transactional sale is that they are all about the same, $n(i)=n(j)$. Which means that transactional sales meet the fundamental requirement for a statistical analysis. And guess what? It works and works extremely well.

But what about non-transactional (or complex) sales?

If you ask any sales person, they'll tell you that no two sales are the same. There's always a number of differences, whether it's the products, the competition, the customer's buying process, the stakeholders or the market for example. This means that B2B sales are all different or $n(i) \neq n(j)$; so, statistics won't work in a non-transaction sales environment, and it never will and big data analysis, while helpful, is also bound to fail.

Applying statistics to sales forecasting is like driving down a busy highway with your eyes firmly fixed on the rear-view mirror. If the road ahead is the same as the road behind, (transactional sales), you're OK. But if the road ahead is not the same as the road behind (non-transactional sales), you've got problems.

If statistics doesn't work for B2B (non-transactional) sales, what does? Probability is what works!

Think back to the recent US election; the Democrats boasted about their Big Data analysis that predicted an overwhelming win for Hillary Clinton. Even the media and pollsters predicted a win, of various degrees, for the Democrats. Everyone was predicting a win for the democrats, except for a lonely professor at the American University in Washington, he predicted Donald Trump would win. What did he do right that everyone else did wrong? He used probability.

So the question now is; "Can we apply probability to non-transactional B2B sales forecasting and come up with an accurate sales forecast?" The answer is 'Yes, we most certainly can.'



*“But, soft! what light through yonder window breaks?
It is the east, and Probability is the sun.*

Arise, fair Probability, and vanquish the envious Statistician”

Above I mentioned the professor who was one of the very few people who predicted Donald Trump’s victory during the recent US Presidential race. The question was ‘How did this professor get it right when so many other got it wrong?’ as mentioned, it was his approach to the question; ‘Who is the most likely person to win the election?’ Rather than relying on statistics to analyze reams of data (Big Data) as the Democrats, polling companies and others, he used a probabilistic approach to the election by examining key aspects of the present (current situation), running it through his model and coming up with the most likely out. And, as we all know he was right and they were wrong. It turns out that he has been using the same probabilistic approach to the US Presidential elections since the early 80’s and has been almost 100% accurate.

And that, in a nutshell, is the difference between statistical predictions and probabilistic predictions; statistics uses past data to predict future events, probability uses current data to predict the most likely outcome of an event.

Confused? Not surprising, as many people blur the boundaries between the two, and, the overlapping terminology doesn’t help either.

Let me provide an example that will hopefully add some clarity. This is an experiment you can do and verify for yourself: tossing a coin.

Everyone knows that if you toss a coin there is a 50% chance of it landing heads and a 50% chance of it landing tails. Statistically, if you toss the coin enough times, half of the time it will land heads (and the half it lands tails).

Suppose you're tossing the coin and it lands heads 4 times in a row, what are the chances of it landing heads on the 5th toss? Statistically there’s a 50% chance the next toss will be a head, so the answer is 50%. But, intuitively that doesn’t feel right, sooner or later it must land tails, so wouldn’t there be a greater chance of tails coming up? Statistically the answer is no.

Now let’s apply some probabilistic thinking to the coin toss; the coin has landed heads 4 times in a row. The question is; “What is the probability of the coin landing heads a 5th time in a row?” The answer is 3.1%! Try it at home and see what you result you get.

We can say that with statistics we know the final outcome, but don’t know the initial conditionals and the challenge is to determine and understand the initial conditions that caused that outcome and we do so by looking for patterns and correlations. And this is where Big Data and huge amounts of computing power jump into the fray. The problem is that if you look hard



enough you can find correlations between just about anything. For example, I refer you to the following article; [‘Hilarious graphs reveal how statistics can create false connections.’](#)

However, with probability we know the initial conditions but don’t know the final outcome and the challenge is to determine the most likely outcome given the initial conditions and the probability model used.

But that’s not the only difference between probability and statistics that makes probability the superior option for B2B sales forecasting, with probability we can update our information in real time, with statistics we cannot.

Let me give you an example that we have all lived through, and for most of us, continue to do so daily, and that is the commute time to get to work in the morning.

On average, it takes me 60 minutes to commute to work. 90% of the time I can leave at 7:00AM and get in by 8:00AM, plus or minus 3 or 4 minutes. But what is the first thing I do when I get in the car? I tune the radio to the traffic channel. Why? To find out what the road conditions are like. If they are different I ask myself, ‘If I follow my normal route, what is the probability of getting to work by 8:00AM?’ Obviously, it depends on the severity of the problem(s), but I quickly assess my chances and if I believe they are unacceptably low I will change routes. But if I don’t change routes I’ll get stuck in traffic; and interestingly, I can be stuck in traffic, with 0% chance getting to work by 8:00AM and the statistical probability remains at 90%. This is what we all do; we approach the problem of the daily commute probabilistically not statistically and update the probability in real time.

So how does all this apply for B2B sales forecasting? The similarity is quite strong.

As a sales manager, I must submit a weekly forecast to my VP of sales. This is a time-consuming task for both me and the team, but it has to be done.

We make extensive use of our CRM (Customer Relationship Management) system and have a sales process that assigns a percent chance of winning at each stage as a deal progresses through the sales process. The actual percentage assigned to each stage is based on historic company-wide performance, so it’s very much a statistical approach. You would think that the easiest thing to do would be take that percent and multiple it by the amount of the deal add it all up and there’s my forecast. Not so!

There a number of problems with this type of statistical forecasting (it’s actually called a Factored or Gated Forecast system), two of which are;

1. Sales is binary; we win it all or we lose it all, we don’t have partial wins in sales.
2. If we make the short-list on a deal with 2 other companies, our CRM tells us that we have a 80% chance of winning. The problem is that our 2 competitor’s CRM systems are also telling them that they too have a 80% (or some number like that) chance of winning. $3 \times 80\% = 240\%$; that’s just impossible!



In addition, my VP wants to know which specific deals we will win (these are called 'Commits'), which ones we could win (called 'Upside') and which ones we're not going to win, at least not this quarter. Unfortunately, the CRM does not provide any information like that. So, it's back to the spreadsheets.

Historically, my team has a win rate of just under 50%. When I look at the deals in our sales pipeline, I know that some of the deals shouldn't be there, some will slip into the next quarter, some we will lose and some we will win. The problem is that I don't know which is which and the CRM does not give me any insights into this. To get the information about the deals I have to 'grill' each member of the team to uncover which deals I can commit in my forecast. As you can imagine this activity is time-consuming and fraught with error, not to mention that sales reps tend to be overly optimistic about their deals and are often very secretive when the deal is in trouble. Then our VP of Sales has a manager's meeting where she grills us on our forecast, which we must vigorously defend.

And the following week I repeat the process and send in my updated forecast. This is the process that I and other sales managers live with, it's no wonder our forecasting is only about 50%.

Then along came Big Data with a list of promises to fix the problem; they will apply new & innovative statistical methodologies to identify patterns in our sales data (which is resident in the CRM) that will identify which deals we are going to win, which deals will close on time and other stuff. The problem was, it didn't work. I'm still using an Excel spreadsheet and grilling the sales reps.

The next iteration of Big Data introduced the idea of searching external databases in addition to our databases to identify much more complex patterns and correlations. But I, and all the other sales managers out there, still have our spreadsheets.

No matter how hard you work or how much computer power is applied to B2B forecasting a statistical approach will not work; you cannot change the fundamental characteristics.

What is needed is to move beyond a statistical approach to B2B forecasting to a probabilistic approach, and recognize that B2B deals are not identical, even though they have many similarities, and allows continuous updates to the initial conditions in real-time. A radically new way of approaching B2B forecasting is required. To quote Orin Harari;

"The electric light did not come from the continuous improvement of candles."

Mel Harding is Vice President of Product Marketing at [Occulus Inc.](https://www.occulusinc.com) and can be reached at Mel.Harding@OcculusInc.com



Mel Harding
VP Occulus Inc.