Election results 2015: How did the polls get it so wrong?

By Caroline Preece

Were data analytics or the data responsible for predicting a hung Parliament?

The Conservative Party is staying in government after winning with a majority almost nobody predicted, least of all the election polls.

YouGov, Ipso Mori and other polling groups all forecast that the Tories would be neck and neck with Labour, with no party having an overall majority in what they predicted would be a hung parliament.

Then last night's exit polls told a different story, showing the Conservatives far in front - they have now won the election with 51 per cent of seats.
As a result, the British Polling Council today confirmed that an independent investigation into the accuracy of election polls will be carried out.

But with fine-tuned Big Data analytics tools widely available, how did the polls turn out so wrong?

Data issues
"People are wonderfully unpredictable," says Martin Lee, cybercrime manager at Alert Logic. "You can create mathematical models that describe everything that is likely to happen to the finest detail, and then people go and do something you didn't expect."

So, yes, there's human error at work here. It's really impossible to accurately predict voting results without leaving a significant margin of error.

This is largely because the voters are the data, and people aren't always truthful - the only absolute fact is the box they mark on the day.

"The problem was not data analytics but the quality of the data itself," explains Nicholas Lansman, group managing director at Political Intelligence. "The election results have made evident that the electorate did not clearly reveal their voting intentions to pollsters.

"There were several reasons for this. The first is the "shy Tory" phenomenon - people reluctant to reveal they wanted a continuation of the Cameron-led Coalition government's austerity measures."

But opinion polls and the graphics they fuel for the media are visible throughout the entire campaign, and so they're understandably presumed by many to reflect what's going on among the British public in the lead up to election day.

Andy Cotgreave, senior technical evangelist at Tableau Software, says: "As with all elections, the press and media developed elaborate visualisations and models of the opinion polls, investing large amounts of time and developer resource.

"The problem lays not in the failure of data analytics but in the fact that the data didn't change in real-time. Despite all efforts by all political parties, the polls were static throughout the campaign. The dashboards became dead ends."

So what happened?

Referring to his Election Wipe programme which aired on BBC2 earlier this week, journalist Charlie Brooker summed up the situation by tweeting:

Philomena Cunk's hypothetical Election Wipe forecast was more accurate than most of the polls.

*— Charlie Brooker (@charltonbrooker)*May 8, 2015

Collecting data and turning it into statistics isn't a remotely new thing in politics, of course, and the polls have also been wrong before. But this election, specifically, was being dubbed the UK's first 'data-driven' vote and, with analytics currently fuelling so much of how we understand business and the world at large, it stood to reason that we could trust what we were being shown.

As Alert Logic's Lee says, the error has served to highlight the importance of developing analytic models, with the right experts working to refine how we understand and interpret results.

"This is why it's so important to continuously monitor the environment to spot the changes and understand how the situation is evolving so that analytic models can be refined," he said.

"It's also why it's necessary not to solely rely on analytical models but to involve domain experts to update and interpret your predicted results.

"No matter what domain you work in, if you get the right technology, the right analytics and the right experts working together then you'll be able to make progress and get the results that you need."

"The real value in data analytics, especially in the run up to a general election, lies in being able to hunt down answers to fresh questions," adds Tableau Software's Cotgreave.

"If data isn't changing, then stop tracking that metric and look for one that is. Dashboards should be regarded as evolving objects that need to change as you seek out the most important answers."

As the general bemusement of the public and politicians revealed last night, many are of the opinion that it is the pollster's methods, rather than a discrepancy with turnout, that needs addressing.

"It would be highly unlikely that we saw a sudden change of heart in voters last night," said Richard Cassidy, technical director EMEA at Alert Logic.

"We shouldn't however sit on our laurels and put it down to a poor turnout, this highlights the need for pollsters to rework their modelling to the reams of data we have over the past several general elections at a typically poor turnout nationally at the polls."

Social media

As well as misleading polls, many were citing the amount of social media discussions that could be attributed to certain parties as reason to believe the ongoing rhetoric of a tight, neck-and-neck race to Downing Street.

Social media has become a valuable source of data over the last few years, with many people pointing to President Obama's 2012 re-electionin the US as an illustration of how political data science can become an integral part of a successful campaign.

Data collected by ElectUK, an app from Tata Consultancy Services, found that most conversation exchanged on Twitter was about Labour (30.9 per cent) and leader Ed Miliband (29.6 per cent) – both positive and negative – with UKIP in second place with 27.3 per cent of mentions.

Tying into this, the economy (33.1 per cent) and healthcare (28.4 per cent) were the top two political issues being talked about on the social media platform ahead of today, based on more than 10 million tweets.

Results from a Proofpoint Nexgate analysis between 1 April and 4 May showed similar data for Twitter – Labour had the most engagement on the site with more than 210,000 followers – while the Tories won on Facebook with more than 467,000 likes.

With Twitter entering the frame for this election as a relatively new means of engaging in discussions around politics, no one knows quite how this data factors into the electoral race itself.

However, it's clear from these results that it has little bearing on how we should understand who people are voting for.

"News is sensationalised and travels at the speed of light, so opinions can change in an instant," says Alexei Miller, managing director at DataArt. "Data analytics need to be developed to recognise nuance more quickly. This is possible, but the need has to be recognised first by those using them to predict elections."

So, while this could be painted as a failure of data analytics to accurately predict the outcome of political campaigns, the fact that it is so widely used across the entire election process poses an issue.

"The science behind forecasting for pollsters isn't exact at all, and we the people are fickle indeed," Alert Logic's Richard Cassidy adds. "It's not the first time we've seen such inaccuracies this year, or indeed over the past decade."

"All in all this general election highlights yet again that the pollsters need to review how the farm the data pre-election and perhaps need to increase their demographic to help prevent such events from reoccurring."

Polling remains the primary method we have for understanding how an election might go.

If we are to improve the results, the ways data scientists understand and interpret the data they are faced with needs to evolve if they are to avoid the same mistakes in the future.