Original Article

Bid Data Analytics and Predicting Election Results

Ajith Singh

Assistant Professor, Patna Women's College, Bihar, India

*Corresponding author: Ajith Singh, Assistant Professor, Patna Women's College, Bihar, India;

E-mail: ajit_singh24@yahoo.com

Copyright: © 2019 Ajith Singh

Abstract

The best way to predict the future is to study past behavior. This is the underlying idea behind Big Data Analytics. The 2008 Obama election campaign was one of the first to take advantage of data-driven methods in the race to an elected office. The Obama campaign had a data analytics team of 100 people. This shows how deeply data analytics impacts the world. From recommending products to customers on e-commerce websites (i.e. using predictive analytics) to electing the most powerful official of the free world. Big Data Analytics is indeed everywhere. Data analytics has evolved itself to become the brain of every election campaign since the Obama campaign. Data analytics helps the election campaign to understand the voters better and hence adapt to their sentiments. Now let's find out how data analytics affects the elections and how election campaigns use it.

Keywords: Big data analytics; Election result prediction; Election result prediction model

Introduction

The 2016 race to the White House had data at its center and made itself an unstoppable force [1-3]. The question here is how it affects the outcome of the election? Positively or Negatively? In other words, does data analytics have the ability to turn election results? Social and polling data can affect the voters [4-20]. Social websites such as Facebook and Twitter optimize their feeds to the target audience to promote voting. Conversely, you see Hillary Clinton leading by 73% chance to take over the White House in the polls (released by some analytics firm) to Trump. In reality, would you agree that a good number of people would feel that the election result is obvious now?

As a result, I feel that there will be a negative impact on the voter turnout in such situations. Websites like Five Thirty-Eight, primroses and Real Clear Politics use social, polling data to predict the election results. To emphasize, if they tweak those results for a single candidate, then it can give an altogether different perspective to their millions of followers, who now after knowing the probable outcome may not turn out at the booths to support their candidate. Hence, it is crucial to realize the downside as well [21-32].

How election campaigns use data analytics

There are two subdivisions of extracting data for an election campaign. Firstly, social data and polling data and

secondly public data which becomes a part of Big Data. It helps the candidate to understand the voters better and design the campaign accordingly (Figure 1). Moreover, this brings more clarity to the election campaign. Both Clinton and Trump campaigns are relying on technology for reaching out to the voters in the 2016 race for The White House [27]. The Campaign job distribution for both Hillary and Trump Campaign obtained from ValuePengiun is shown below. We can observe from the graph that Data Analytics and the resulting Strategic Operations takes up a huge chunk of the workforce of both the presidential campaigns [30].
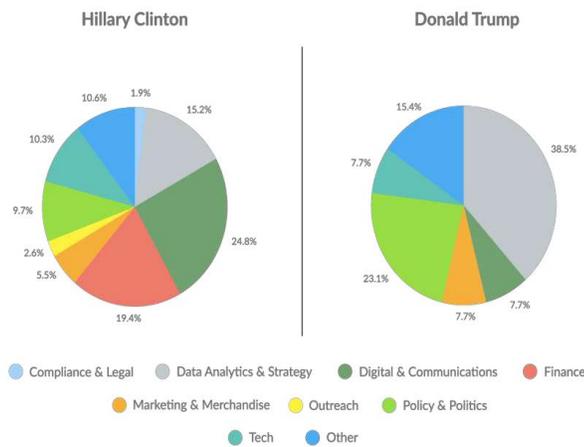

Figure 1: Data Analytics election campaign.

Identifying the Swing States

A swing state is a state where the two major political parties have similar levels of support among voters, viewed as important in determining the overall result of a presidential election. Swing states are one of the most important factors in the US elections [3]. Red states are ones that are dominated by the Republicans (i.e. Trump party) whereas the blue ones signify the dominance of the Democrats (i.e. Clinton party). Hence, swing states are also known as purple states as both parties have similar electoral support in these areas. Large amounts of public data, polling data, sentimental analysis of Twitter and Facebook feeds are used to determine the swing states. In particular, winning the swing states can make a big difference in the electoral votes. These are the best opportunity for a party to gain electoral votes. So, political parties majorly focus on these states while strategizing their election campaign [31].

In 2016, US Presidential elections the 12 swing

states are – Wisconsin, Minnesota, Nevada, Pennsylvania, New Hampshire, Colorado, Ohio, Iowa, Virginia, Florida, Michigan, and North Carolina. ─Tipping-point chance‖ as described by FiveThirtyEight is the probability that a state will provide the decisive vote in the Electoral College (Figure 2). This is a good indicator of the Swing states [32]

Figure 2: Tipping point chance.

Online and offline marketing

Using big data analytics, the election campaign analyzes the demographics of the states where they fall behind their opposition. Offline marketing like billboards and television ads is deployed strategically to target the audience using data analytics. It helps them to understand the states where the campaign needs to improve on the marketing and hence turn the voter sentiments around [12].

Methodology-The Primary Model-Statistical Model (2016)

87%-99% Certain Trump Will Be President.

It is 87% to 99% certain that Donald Trump will win the presidential election on November 8, 2016; 87% if running against Hillary Clinton, 99% if against Bernie Sanders.

These predictions come from primarymodel.com based on a statistical model that relies on presidential

primaries and an election cycle as predictors of the vote in the general election.

Winning the early primaries is a major key for electoral victory in November. Trump won the Republican primaries in both New Hampshire and the South Carolina while Hillary Clinton and Bernie Sanders split the Democratic primaries in those states (Figure 3).

What favors the GOP in 2016 as well, no matter if Trump is the nominee or any other Republican, is the cycle of presidential elections. After two terms of Democrat Barack Obama in the White House the electoral pendulum is poised to swing to the GOP this year [32].

	Primary Vote of Party Candidates in Selected Elections Nominees in Caps, Winners in Bold						
	Republican Party			Democratic Party			
Year	Candidate	(%)	Rival (%)	Candidate	(%)	Rival (%)	
1912	Taft	33.9	51.5	**WILSON**	44.6	41.6	
.....							
1928	**HOOVER**	49.2	31.2	SMITH	29.5	20.8	
.....							
1964	Goldwater	22.3	35.5	**JOHNSON**	95.3	1.6	
.....							
1980	**REAGAN**	49.6	22.7	CARTER	47.1	37.3	
.....							
2012	ROMNEY	39.3	22.9	**OBAMA**	82.0	1	

Figure 3: Primary vote of party candidates in selected elections nominees in caps.

In a match-up between the Republican primary winner and each of the Democratic contenders, Donald Trump is predicted to defeat Hillary Clinton by 52.5% to 47.5% of the two-party vote. He would defeat Bernie Sanders by 57.7% to 42.3%. For the record, the PRIMARY MODEL, with slight modifications, has correctly predicted the winner of the popular vote in all five presidential elections since it was introduced in 1996. In recent elections the forecast has been issued as early as January of the election year. Presidential elections going back as far as 1912 are used to estimate the weight of primary

performance. It was in 1912 that presidential primaries were introduced. That year the candidate who won his party's primary vote, Woodrow Wilson, went on to defeat the candidate who lost his party's primary vote (Figure 4), William Howard Taft. As a rule, the candidate with the stronger primary performance wins against the candidate with the weaker primary performance. For elections from 1912 to 2012 the PRIMARY MODEL picks the winner, albeit retroactively, every time except in 1960[16].

Figure 4: How the white house party has fares. (Since 1960)

For elections prior to 1952 all primaries were included. Beginning in 1952, only the New Hampshire Primary has been used, as a rule. South Carolina has been added to gauge primary performance this year. Hillary Clinton enjoys strong support in a large group of voters, African-Americans, who are few in numbers in New Hampshire. Her combined performance in New Hampshire and South Carolina gives her a higher primary score than Sanders. As a result, in the general election Clinton fares less badly against Trump than does Sanders [15].

An earlier forecast, which garnered much notoriety, predicted a Trump victory over Clinton with 97% certainty. That prediction was made before the Democratic primary in South Carolina was held and relied on polling reports for that state. Clinton wound up beating Sanders by a much bigger margin than was indicated by pre-primary polls. Still, it is 87% certain that Trump will defeat Clinton in

November. It is possible, of course, that the Republican Party will not nominate Trump as its presidential candidate. If the nomination were to go to Marco Rubio instead the PRIMARY MODEL would predict:

It is 86% certain that Hillary Clinton will defeat Marco Rubio. Clinton will get 52.4% and Rubio 47.6% of the two-party vote [11]. According to the model, Ted Cruz or any other candidate (except Trump) would fare the same way against Clinton. In the event that the Democrats nominate Bernie Sanders instead of Hillary Clinton, while the GOP nominates Marco Rubio, the PRIMARY MODEL would predict: It is 89% certain that Marco Rubio will defeat Bernie Sanders. Rubio will get 52.8% and Sanders 47.2% of the two-party vote [32].

According to the model, Ted Cruz or any other candidate (except Trump) would fare the same way against Sanders.

Big Players in the Election Forecast

Now, let looks at some of the notable players who use Data analytics on polling, social and big data for the forecast.
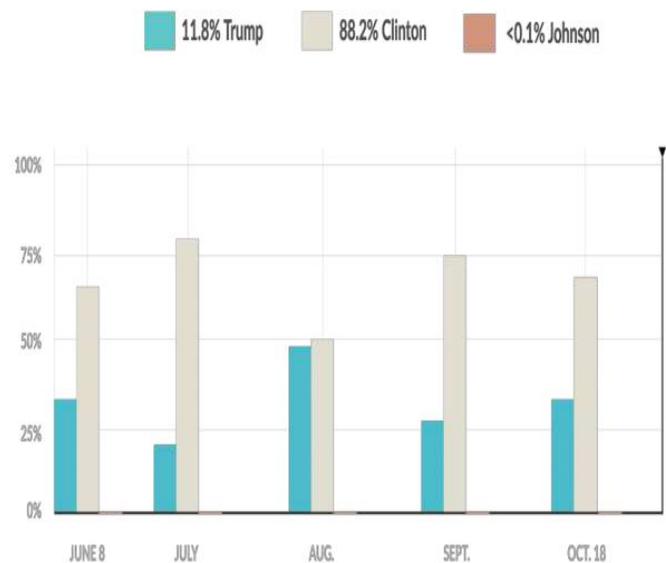
Five Thirty-Eight

In 2007, Nate Silver launched Five Thirty-Eight. Silver made data analytics super cool with his famous 2008 US Presidential election predictions. Five Thirty-Eight 2008 presidential election forecast had 98.08% accuracy in predicting the winners in each of the states.

Notably, they correctly predicted the winner of 49 of the 50 states including the District of Columbia. Overall, Indiana is the only state in which they missed out. Five Thirty-Eight prediction on-chance of winning‖ for the 2016 election cycle is shown below. Before the advent of model-mania, a common way to think about polling was that if the polls are close, the race is a "toss-up," and it could go either way. RealClearPolitics, for instance, continues to classify any race where the polls average out to a 5-point lead or less for one candidate as a toss-up, instead of making more specific forecasts of some kind. Though based on most of the same underlying polls, FiveThirtyEight and its kin try to come up with a numerical percentage of how likely each candidate is to win, with the help of historical data and mathematical

modeling (Figure 5).

"Almost all statistical models are grounded in history," Silver recently explained in an appearance on The Ezra Klein Show. "The implicit idea is that you are hoping that history will repeat itself, at least in a probabilistic way." So, they start from polling averages of the current races. They modify that a bit with certain other technical tweaks depending on the particular version of the model. For instance, in sparsely polled House races, they incorporate other polls that could be helpful (similar districts, state, or national numbers). Some versions of the model incorporate other factors that appear to be historically important, like fundraising: the "fundamentals."

Figure 5: Big players in the election forecast.

Out of all that, FiveThirtyEight calculates a candidate's expected vote share. This looks pretty similar to a polling average—for instance, in the Florida Senate race, they show Democratic Sen. Bill Nelson leading his Republican opponent Rick Scott by 3 percentage points (as of Tuesday afternoon).

Then, though, comes the big leap. Working off that expected vote share, they simulate the election thousands of times — comparing to a wealth of historical data on how past elections turned out, and trying to incorporate uncertainty. What comes out at the other end is the candidate's projected chance of victory. For Bill Nelson, that's 69 percent [6].

Once they do that for every House and Senate race up this year, they come up with an overall estimate

of the chance each party will win control of each chamber. Something like an 81.3 percent chance Republicans will keep the Senate, and an 85.6 percent chance Democrats will take the House.

What happened with the forecast models in 2016?

An 80 percent chance of winning compared to a 20 percent chance seems like a huge advantage. Yet in commentary about their models, Silver and the rest of the FiveThirtyEight team (I recommend their podcast) repeatedly stress that an 80 percent chance of victory is not a done deal — not even close. Their argument is that an outcome that's 20 percent likely will happen 20 percent of the time.

The most infamous example of something like this was, of course, the 2016 election. FiveThirtyEight's model gave Hillary Clinton a 71.4 percent chance of winning, and she, er, didn't win [32].

But though this put an end to the legend of Silver as the forecaster who "called" all 50 states "correctly" in 2012, he ended up something of a winner anyway. In the days before the election, even though his model showed Trump as the underdog, it gave Trump a substantially higher chance of winning than any other mainstream model out there. For that, some accused him of "panicking the world" and "putting his thumb on the scales," and others opined he was "cautiously" trying to ensure his forecast looked right whatever happened [3].

In the days before the election, I dug into FiveThirtyEight's modeling choices and concluded they made "a whole lot of sense." Most broadly, I wrote, this was because "Silver's forecast is just more uncertain that the result will match what the current polling data shows." He understood quite well that there could be a polling error, a last-minute swing, or both.

In the end, the worst error made by other 2016 forecasters were that they drew far-too-confident conclusions from Clinton's relatively narrow, single-digit poll leads nationally and in key swing states [7].

Conclusion

To sum up, we have seen how highly data analytics is used by election campaigns and how it affects elections as a whole.

Additionally, this also opens a whole world of possibilities on how someone can be a part of such a technological field with great impact. For elections from 1912 to 2012 the PRIMARY MODEL picks the winner, albeit retroactively, every time except in 1960. For elections prior to 1952 all primaries were included. Beginning in 1952, only the New Hampshire Primary has been used, as a rule [9]. "Almost all statistical models are grounded in history," Silver recently explained in an appearance on The Ezra Klein Show. "The implicit idea is that you are hoping that history will repeat itself, at least in a probabilistic way."

References

1.   www.intheimageofthecreator.com
2.   www.researchgate.net
3.   www.pinterest.com
4.   www.en.wikipedia.org
5.   www.citylens.blogs.rutgers.edu
6.   www.law.justia.com
7.   www.frontloading.blogspot.com
8.   www.en.wikipedia.org
9.   www.123helpme.com
10.   www.pinterest.com
11.   www.researchgate.net
12.   www.pt.scribd.com
13.   www.researchgate.net
14.   www.coursehero.com
15.   www.pt.scribd.com
16.   www.tarbabys.com
17.   www.offthekuff.com
18.   www.stat.columbia.eduepdf.tips
19.   www.offthekuff.com
20.   www.bizjournals.com
21.   www.msn.com
22.   www.liebertpub.com
23.   www.pdfs.semanticscholar.org
24.   www.quizlet.com
25.   www.researchgate.net
26.   www.patents.google.com
27.   www.youtube.com
28.   www.in.answer.yahoo.com
29.   www.debate.com
30.   www.sites.google.com
31.   www.pinrest.com