

The Making of a President Using Data Analytics and Social Media

Ed Lindoo
Regis University

Today, political campaigns rely heavily on analytics to target potential voters. The algorithms used, as well as bias that may be introduced in the process, may skew the results and cost an election, as seen in the Clinton campaign. Additionally, Social Media plays a huge role in political races, as does fake news. More alarming is the new trend for Social Media sites to censor anyone or anything which can also make or break a campaign. In this paper we research events that affected the 2016 Presidential election and present issues that may have an effect on future elections.

Keywords: big data, bias, analytics, fake news, social media

INTRODUCTION

In Data Sciences and Computing in general, bias can be introduced in a variety of ways. Conscious or subconscious prejudice, bias, can be found in the data collection phase, analysis phase, and insight phase. Bias can also be found in the outcome phase, assessment phase and improvement phase. Analysts must be aware of bias and cautious to prevent it in all phases of their analysis. Analytics and ethics are not necessarily enemies but can be if not properly treated. For example, releasing data gained through analytics, whether it is technically correct or not, could result in a company's loss of their reputation, create competitive weakness, or possibly result in legal sanctions (Noyes, 2015). Big data and analytics extend well beyond organizational benefits and Gartner (Noyes, 2015) predicts that by 2020, big data analytics will cause half of all business violation ethics. The use and or misuse of data also holds true in political analytics.

Today, companies world-wide use big data in many meaningful ways, and recently, so has our political system. In fact, during the 2008 and 2012 elections, the Obama campaign relied heavily on analytics to outperform their Republican counterparts. Going into the 2016 elections, both sides had come to realize that there had been a major shift in how people communicated, moving away from landline phones and going more for mobile phones and Social Media. As a result, it came as no surprise that both the Democratic and Republican campaigns would follow what Obama had done and use analytics in a big way for the 2016 election (Enderle, 2016). With so much data, what went wrong for the Clinton campaign and what went right for the Trump campaign?

LITERATURE REVIEW

Introducing Bias

The Book “Shattered” shares a perspective on why Clinton lost the 2016 election. Highlights include Clinton’s campaign manager, Robby Mook, who put a lot of faith into a self-developed, super analytic system called Ada. Mook felt that Ada was the campaigns “secret weapon.” As for Mook’s algorithm, when approached, Mook’s team said, “we’re the smartest guys around here so let us do our thing” (Jacknis, 2017). The Ada algorithm played a role in just about every strategic decision made by the campaign, including where and when to send the candidate, her staff, and where to place TV ads. Interestingly, Ada was able to run about 400,000 simulations a day, providing a detailed picture of which battleground states would be the tipping point. (Wagner, 2016). Sometimes too-much is too much. Because it’s humanly impossible to sift through 400,000 simulations every day, staff had to rely on Ada to produce summary results of these simulations.

One of the things that Ada missed was the states of Michigan and Wisconsin. The Clinton campaign had paid little attention to these states, assuming they were in the bag, and it wasn’t until just a week before the election that staff realized they had a problem. It appears that Ada had been programmed to underestimate the power of rural voters. The campaign then tried a last-minute ad campaign and a tour of these states with Clinton and the Obama’s, but it was simply too late. (Wagner, 2016)

Was there bias? According to Aurerbauch (2017), Ada was developed by dozens of researchers who were led by Clinton’s director of analytics, Elan Kriegel, in close consultation with campaign manager Robby Mook. Inputs to Ada included polls and surveys as well as the campaigns data from field workers and it’s possible that bias could have been introduced at these points.

The UPB Phenomena

At the beginning of the 21st century, accounting scandals, among other events, led to the collapse of billion-dollar companies. These scandals only confirmed that unethical acts were being conducted, often flourishing within organizations. Research into unethical behavior within organizations has found assorted reasons as to why employees might engage in unethical acts, for example, personal gain, the gain of the organization, or even to cause harm against the organization. (Umphress, 2010). These acts fall under the title of unethical pro-organizational behavior. UPB includes acts of commission (e.g., “cooking” numbers to boost analyst projections and stock values) and omission (e.g., withholding information about the hazards of a pharmaceutical product) that are typically considered unethical.

UPB is predominantly considered pro-organizational behavior which is neither specified in formal job descriptions nor ordered by superiors, yet, is carried out to benefit or help the organization. Many theorists believe that the more an individual identifies with their organization, the more likely they are to disregard their personal moral standards and engage in acts that favor the organization. Further research has found that the stronger the employees’ organizational identification, the more influence it will have on productive work behavior, such as increased extra-role behaviors, and job performance. Umphress et al., believe that individuals who identify strongly with their organization may be more likely to engage in UPB. (Umphress, 2010).

As a public example, we have recently noted journalists who have been found to be lying or “bending” the truth. In 2015, NBC Nightly News anchor, Brian Williams, was caught lying about riding in a helicopter that was shot down in Iraq in 2003. NBC suspended him for six months without pay and the Peabody Award-winning newsman, became the top story himself. In 2015, Fox News Channel star, Bill O’Reilly, was caught lying about experiencing war involving Argentina and England in 1982. In 2011, MSNBC’s Rachel Maddow, was caught lying when she accused Rush Limbaugh of racism. She was later forced to apologize to Limbaugh on the air. Be it public figures, or an analyst within an organization, people tend to get caught up in the moment, the organization, personal gain, or a number of reasons that might fall under UPB. Imagine then, how an analyst could add an extra day of revenue to the monthly or quarterly reports for a company. Depending on their position within the company, unethical behavior may be committed (Umphress, 2010).

What does the UPB phenomenon have to do with political analytics? Probably more than we know! In fact, Aurerbach (2017) determined that Clinton analysts did have bias, and programmed that into Ada, assuming that Michigan, Wisconsin and Pennsylvania were a “lock”. As Aurerbauch (2017) points out, “If a piece of code crashes, it’s broken, but at least you know it’s broken. The most dangerous kind of code—as I learned too many times in my years as a software engineer at Google and Microsoft—is the kind that breaks but appears to keep working. The worst part is that you have only yourself to blame” Aurerbauch (2017). In essence this was Ada’s failure.....it went wrong early and no one caught it.

Passive Bias

Enderle (2016) found another type of bias was introduced, passive bias. As the campaigns matured prior to the November elections, things heated up and Clinton began a very negative campaign against Trump, one that would see her call his followers “deplorables”. This caused many folks on both sides to either not respond to surveys or not respond truthfully. As a result, this introduced massive pro-Clinton bias, a bias that Ada could not account for. Enderle went on to state that “As an analyst, a big part of the job is identifying and mitigating bias otherwise you are driving the people who pay you to make bad decisions and, given the outcome, that would seem to be the case here” (Enderle, 2016).

If that were not enough, even more bias was introduced by complacency. Because Ada was reporting great numbers to the Clinton campaign, Mrs. Clinton reportedly became complacent and started laying back while Trump who didn’t believe any of the numbers pushed even harder. As Enderle (2016) pointed out, when people don’t challenge the results they like, it almost always leads to bad results.....this certainly proved out in Clinton’s case. In fact, Enderle pointed out the three rules of analytics. First, assure your data source. You must have a strong sampling methodology to obtain accurate results. Secondly, it is a must to identify and eliminate any and all bias; any bias will invalidate the results. Finally, decision makers must challenge the analysis, especially if the analysis is telling you what you want to hear.

Even more bias was introduced via polarization. Smith (2017) found that as campaigns target their supporters with data, there is less effort put forth to persuade these same supporters, causing polarization. Polarization happens because the data puts these potential voters into predetermined buckets which reflects the public’s ideology and voting preferences. As Smith (2017) points out, this can be dangerous and may lead to self-induced bias.

The Use of Facebook

While the Trump campaign relied on voter data owned by the Republican party, they also engaged the services of Cambridge Analytica, using their models to help make decisions on advertising and how to better reach donors (Kirchaessner, 2017). Because the Clinton campaign was using a highly sophisticated analytical approach, Trump’s campaign knew they needed to combat that with their own analytics. One of the things that helped Cambridge Analytica turn the tide was in their ability to acquire data on people who had voted early, data that it collected from local counties and states. This data was then linked to individual Facebook accounts and then analyzed to determine the political advertisements that they had been subjected to (Kirchaessner, 2017). In his book “Outnumbered” David Sumpter, a professor of applied mathematics at the University of Uppsala in Sweden, analyzed the Cambridge Analytica’s models and found they were using a regression “that takes the data we already have about a person and uses them to predict something we don’t know about him or her.” (Rathi, 2018) However, as Rathi (2018) points out, it is quite difficult to predict a person’s personality from their Facebook page using this regression model. Rathi went on to state that a better method is to use Facebooks advertising platform which enables developers to build audiences similar to a manually selected audience.

Although Trumps campaign was investigated for using Cambridge Analytica and their use of Facebook, the Clinton campaign also gathered tons of information from Facebook users by creating and distributing an app that asked users to pair their Facebook friends and families with the smartphone contacts. In this way the Clinton campaign could reach friends of friends (Jacknis, 2017)

The Analytical Focus on Social Media

It used to be that elections would almost certainly be decided by about 20 percent of voters who fell somewhere in the “ideological middle”, but that logic has now disappeared as a result of big data analytics. And because of this, Todd (2017) believes that Big Data is hurting American politics, not because it’s wrong or that it can’t be used for good, it’s the cries of “fake news” and “unskew” that are accurate.

Prior to the 1900’s, local, national and world news came to us primarily by way of newspapers. In the early 1900’s, radio introduced a new medium by which people obtain news. These mediums had editorial boards that scrutinized articles written by trained professionals. These professionals often went through years of college preparatory courses and internships prior to attaining a job in the world of news. Television appeared in the 1950’s, and the footsteps of radio were followed when it came to news integrity. When the Internet arrived in the late 90’s, so did Blogs, Facebook, Twitter, YouTube, and a host of other Social Media sites. With Social Media, came reporter “wannabees”, with little or no training in the world of news reporting, and frequently, with little, if any, reporting ethics, presenting information that is often hard to substantiate.

Fake News

It’s not known which came first, poor reporting on the Internet, or, poor reporting on the major cable networks such as CNN, MSNBC, FOX, BBC. As witnessed prior to the 2016 presidential election, everyone appears to be scrambling to be the first to report news, be it accurate or not. As a result, we regularly see, companies such as The New York Times and CNN retracting stories upon realizing the stories are incorrect. How much damage is done when a wrong (Fake News) story goes out? Researchers at MIT researched this question, and recently concluded a five-year study on “Fake News on Twitter” (Meyer, 2018).

The study found that a false story reaches 1,500 people six times quicker, on average than a true story does. They also found that false stories outperform the truth on every subject including business, terrorism and war, science and technology, and entertainment, determining that fake news on politics routinely outperform the other categories. Ultimately the researchers found that between 2006 and 2016, about 126,000 tweets had been retweeted more than 4.5 million times. Some of these were linked to “fake” stories hosted on other websites while others started rumors themselves, either in the text of a tweet or in an attached image. For example, a rumor was tweeted in February of 2016, reportedly by a recently deceased elder cousin of Donald Trump, and opposing the presidential bid, stated in his obituary, “As a proud bearer of the Trump name, I implore you all, please don’t let that walking mucus bag become president”. However, Snopes could not find evidence of the story and rejected it as false. Nonetheless, roughly 38,000 Twitter users re-tweeted the story. It’s not known how many might have shared this in the form of an email or other medium (Meyer, 2018)

In a 2017 study of Higher Education Students, Chandra, Surjandy and Ernawaty found that higher education students can become emotional and be misguided by fake news. Alarmingly they can disseminate this fake news faster because of Social Media and instant messaging applications. Because Social Media and instant messaging are used by almost everyone who has a smartphone, to interact and communicate with just about anyone, they can quickly disseminate information, motivations, and promotions, good or bad. Through their research, it was determined that a large proportion of their sample population of Higher Education students actually modify the news and made it look legitimate by including a reliable source! (Yakob, C., Surjandy, Ernawaty, 2017).

Misinformation and Disinformation

False information comes in two forms: misinformation and disinformation. Disinformation is false information that is purposely spread to deceive. Misinformation is simply incorrect information, for example, “I was misinformed about when to meet you for lunch, however I know it was not deliberate.” One study, as reported by Kshetri (2017) found that roughly 62% of US adults get their news from Social Media sites, and of those 40%, from Facebook. Kshetri (2017) also determined that in the final three

months of the 2016 US presidential campaign, the top performing fake election news stories on Facebook attracted more views than the top stories from major news outlets such as the New York Times, Washington Post, Huffington Post, or NBC News. During that time 8.7 million shares, reactions, and comments were generated on Facebook as compared to just over 7.3 million from 19 major news websites. As we all know now, Russia was involved.

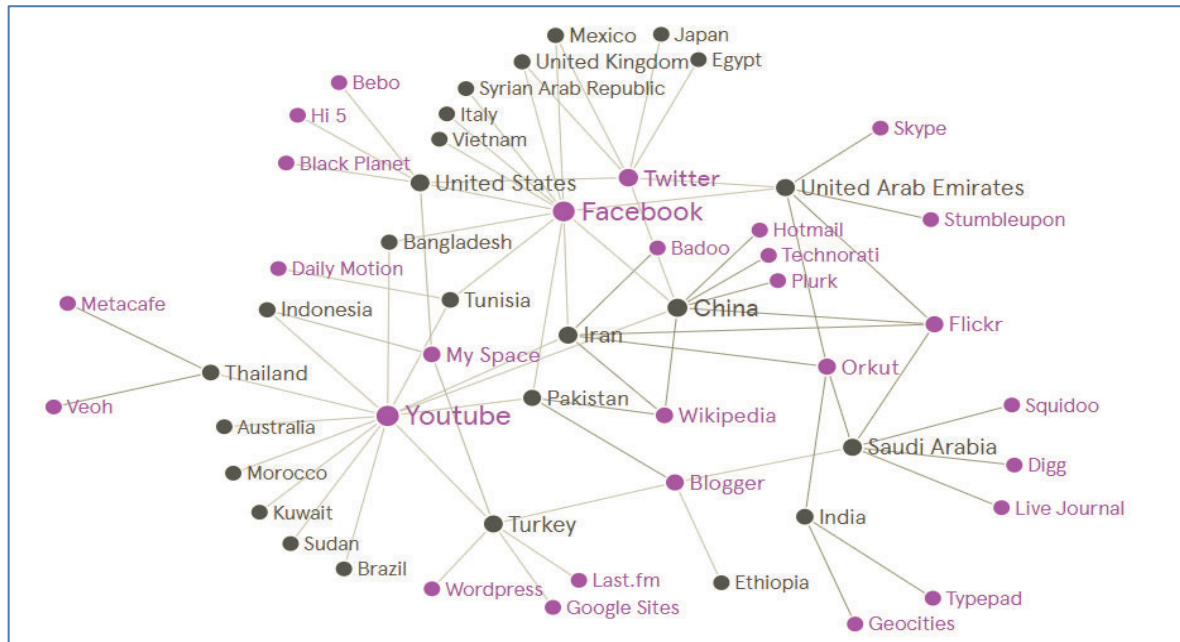
Russia, however, was not the only country spreading fake news. We have recently learned that other fake news creators were known to be operating from countries such as the Republic of Georgia and Macedonia. For example, during the one-year period before the 2016 US presidential election, residents of the Macedonian town of Veles (population 45,000) launched more than 140 US political websites. Most of the domain names looked American, such as WorldPoliticus.com, TrumpVision365.com, USConservativeToday.com, DonaldTrumpNews.com, and USA DailyPolitics.com Kshetri (2017).

The China Syndrome

Realizing that Social Media can sway stock markets and has the potential to overthrow governments, the Chinese central government has placed bans on many Social Media URLs. It is common knowledge that Facebook, Google, YouTube and Twitter are banned, in fact, over 8,000 sites are banned in China (Greatfire.org). Studies (Kuang, 2018) suggest that all media content harmful to the legitimate ruling of the communist state is censored. To protect its rule, the Chinese Communist Party (CCP) has identified a few priorities including the sustaining of economic growth, nationalism, social stability, and rational legal authority and electoral legitimacy. This is also what sustains the current political regime. Therefore, the news most likely to be censored is news that the propaganda authorities believe will have a negative effect on the legitimate rule of the Communist Party state. Further, Chinese law mandates that all Chinese news media must contribute to the enhancement of party ruling (Kuang, 2018).

China, however, is not the only country to ban Social Media sites. Figure 1 is a good representation of the major sites banned by numerous countries around the world (Yasaklanan and Ađlar, 2018). North Korea is not shown as everything is blocked. Until recently, Cuba was very much like North Korea, although presently more and more sites are being allowed by the Cuban government. However, all the Internet access is controlled by the Cuban government and opposition sites are blocked (San Pedro, 2016).

FIGURE 1
A HIGH-LEVEL VIEW OF COUNTRIES THAT BLOCK VARIOUS WEBSITES



Yasaklanan, 2018

The Push to Curtail Social Media

As we have seen, some governments simply ban Social Media sites while others such as the U.S. are either trying to pass laws to ban certain content or have asked the leaders of these organizations to do their own censoring. After the 2016 election, attention was given to the false, inflammatory and misleading information being presented on Facebook, Twitter, Google and other online services who now find themselves in a position of how to appropriately operate on a global platform (Lima, 2019). In a recent and certainly controversial move, Facebook banned the accounts of far-right extremists Alex Jones, Milo Yiannopoulos and Louis Farrakhan. To most, this seems like the right thing to do, however it begs the question, where do you draw the line. For example, when questioned by Congress, Zuckerberg stated that he that “he’d allow posts denying the Holocaust as long as the people behind them sincerely believed what they were saying” (Lima, 2019).

In a Senate hearing during April 2019, Republicans accused Facebook, Google and Twitter that they censor conservative user’s content online. They went on to threaten regulation of these and other online services. In the hearing, Senator Ted Cruz stated "Not only does big tech have the power to silence voices with which they disagree, but big tech likewise has the power to collate a person's feed so they only receive the news that comports with their own political agenda" (Romm, 2019). During this hearing it became clear that Republicans believe that Facebook, Google and Twitter are biased against the party. Political parties or not, "Every time Facebook makes the choice to remove content, a single company is exercising an unchecked power to silence individuals and remove them from what has become an indispensable platform for the speech of billions" (Lima, 2019). This should put fear into a lot of people because it is possible that just a few Social Media companies have the power to define elections in the United States.

CONCLUSION

As demonstrated in this paper, data analytics is often swayed by bias. Managers of data and analytics need to be aware of this and ensure that the findings they present are accurate and without bias. It remains to be seen if the UPB phenomenon will continue to grow, or if those in charge of data will come to realize the ethical and moral values of accurate reporting. It also remains to be seen if college students, news organizations and others will continue to spin “fake news” across Social Media sites, or if they too will come to realize the ultimate ramifications of putting out un-validated stories.

This paper presented the UPB phenomenon, pointing out that it is quite easy to fabricate data, be it for personal reasons, or to benefit a company or a political campaign. Data can be analyzed and presented in a way that may be incorrect and more importantly, analysts need to be aware of personal bias, or biases that may come from above. Finally, it was determined that Social Media and “Fake News” are presenting and will continue to present dangers to our society, especially through our young, college-level students. This danger was seen not only in the use of Social Media, but also television, radio, print and cable news. As pointed out, Fake News travels at a rate six times faster than real news, yet Fake News is often more believed and therefore opinions formed with no real basis.

With the 2020 Presidential elections closing in, it remains to be seen as to what effect analytics will again have on the election. Will campaigns be able to data mine Facebook users accounts as they did in the 2016 elections? Will they find other ways, legal or not to get to the data that they’re looking for? Will they be better able to target market those swing voters? Most importantly, will Social Media sites introduce bias or cutoff certain messages in an effort to sway the election?

REFERENCES

- Auerbach, D. (2017). *Confirmation Bias: Did big data sink the Clinton campaign?* Retrieved from <https://nplusonemag.com/online-only/online-only/confirmation-bias/>
- Bringing transparency to the Great Firewall of China. (2018). GreatFire.org. Retrieved from <https://en.greatfire.org>
- Enderle, R. (2016). *How Trump defeated Clinton using analytics*. CIO Magazine. Retrieved from <https://www.cio.com/article/3138972/how-trump-defeated-clinton-using-analytics.html>
- Jacknis, N. (2017). *Campaign Analytics: What Separates The Good From The Bad*. Retrieved from <https://startupsventurecapital.com/campaign-analytics-what-separates-the-good-from-the-bad-974fb273456a>
- Kirchgaessner, S. (2017). *Cambridge Analytica used data from Facebook and Politico to help Trump*. The Guardian. Retrieved from <https://www.theguardian.com/technology/2017/oct/26/cambridge-analytica-used-data-from-facebook-and-politico-to-help-trump>
- Kshetri, N., & Voas, J. (2017). *The Economics of “Fake News” IT Pro-IEEE*. Retrieved from <https://ieeexplore-ieee-org.ezproxylocal.library.nova.edu/stamp/stamp.jsp?tp=&arnumber=8123490>
- Kuang, X. (2018). *Central State vs. Local levels of government: Understanding news media censorship in China*. Fudan University and Springer Nature Singapore Pte Ltd. Retrieved from <https://link.springer.com/article/10.1007/s41111-018-0091-5>
- Lima, C. (2019). *Facebook wades deeper into censorship debate as it bans 'dangerous' Accounts*. Politico. Retrieved from <https://www.politico.com/story/2019/05/02/facebook-bans-far-right-alex-jones-1299247>
- Meyer, R. (2018). *The grim conclusions of the largest-ever study of Fake News*. The Atlantic. Retrieved from <https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/>
- Noyes, K. (2015). *Mind your analytics, or you could soon be in ethical hot water*. IDG News Service. Retrieved from <https://www.computerworld.com/article/2993922/mind-your-analytics-or-you-could-soon-be-in-ethical-hot-water.html>

- Rathi, R. (2018). *Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential election*. Retrieved from <https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d>
- Room, T. (2019). Senate Republicans renew their claims that Facebook, Google and Twitter censor conservatives. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/04/10/facebook-google-twitter-under-fire-senate-republicans-censoring-conservatives-online/?utm_term=.beef09528527
- San Pedro, E. (2016). *Cuba Internet access still severely restricted*. BBC. Retrieved from <https://www.bbc.com/news/world-latin-america-35865283>
- Smith, A. (2017). *Campaigns know more than ever about you — And it may be breaking our politics*. Business Insider. Retrieved from <https://www.businessinsider.com/campaign-big-data-use-politics-analytics-2017-6>
- Todd, C., & Dann, C. (2017). *How Big Data broke American politics*. NBC News. Retrieved from <https://www.nbcnews.com/politics/elections/how-big-data-broke-american-politics-n732901>
- Umphress, E., Bingham, J., & Mitchell, M. (2010). Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *Journal of Applied Psychology*, 95(4), 769 –780.
- Wagner, J. (2016). Clinton's data-driven campaign relied heavily on an algorithm named Ada. What didn't she see? *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/post-politics/wp/2016/11/09/clintons-data-driven-campaign-relied-heavily-on-an-algorithm-named-ada-what-didnt-she-see/?utm_term=.b61a6889691b
- Yakob, C., & Surjandy, E. (2017). Higher Education student behaviors in spreading Fake News on Social Media: A case of LINE Group. *International Conference on Information Management and Technology (ICIMTech)*. Retrieved from <https://ieeexplore.ieee.org/document/8273511>
- Yasaklanan, D., & Ađlar, S. (2018). *Mapping the banned Social Media by country*. Retrieved from <https://graphcommons.com/graphs/52eadc97-6bc3-48a2-8526-741f358c9146>