

2020 Election Fraud: What Can Twitter Teach Us?

Jill Noorily with Professor Chun and Professor Elkins
IPHS 200 Programming Humanity



Abstract

Through data analysis, we can explore the intersection of social media and politics. Focusing on election fraud in the 2020 presidential election, I collected 36,000 tweets and used topic modeling to analyze the tweets. Using topic modeling to find word frequency and topics, I was able to further understand how social media, specifically Twitter, interacts with individual's views of US democracy and the electoral process. Throughout the project, it becomes clear that social media is a platform that can be used to amplify anxiety, fear, concerns, and misinformation. This study of voter fraud provides insights into the inner workings of social media itself.

Introduction

Claims of voter fraud have been around since America's birth. Although many still worry about election fraud, looking into America's election history makes it clear that US elections have become more and more secure. In the nineteenth century, fraud was an accepted part of part of US elections; coups and social scrutiny commonly and openly pressured an individual's vote. Since then, frequent reforms have been made to secure the validity of the US elections. A 2017 study done by the Brennan Center for Justice found that the recent rate of voter fraud in the US has been reduced to between .00004 %-.0009%. Despite these low rates of voter fraud, allegations of election invalidity remains a recurring topic. During the unprecedented year of 2020, characterized by the Covid-19 pandemic, an awakened movement for racial justice, and an unconventional race for president, many US citizens lost confidence in our electoral system. As technology and social media advance and become integrated into our culture and democracy, they become tools that can be used to understand complex systems. In this project, I have collected tens of thousands of tweets from before, during, and after the 2020 election related to topics of voter fraud, President Donald Trump, and President-Elect Joe Biden. In using topic modeling to understand the data collected, it becomes clear that despite the low statistics of voter fraud people share contrary and diverse views about our democracy.

Since its founding in 2006, Twitter has become a pillar of modern day communication and culture. The USA's political divide has split the country and the popularization of Twitter accentuates the divide. Social media has been portrayed as a benefit to humanity, a tool that can connect and bring us closer. However, the opposite has also been proven. Social media has developed into a tool that is used for mass manipulation of the user. How will this impact our democracy? How has our democracy already been impacted by this? Twitter in particular is a tool used by those in positions of power, sharing information to millions of people every second of every day. This creates social networks in which information regarding conspiracy, rumors, and falsehoods are spread efficiently. Studying tweets about voter fraud allows us to understand the different viewpoints that people can have on one topic.



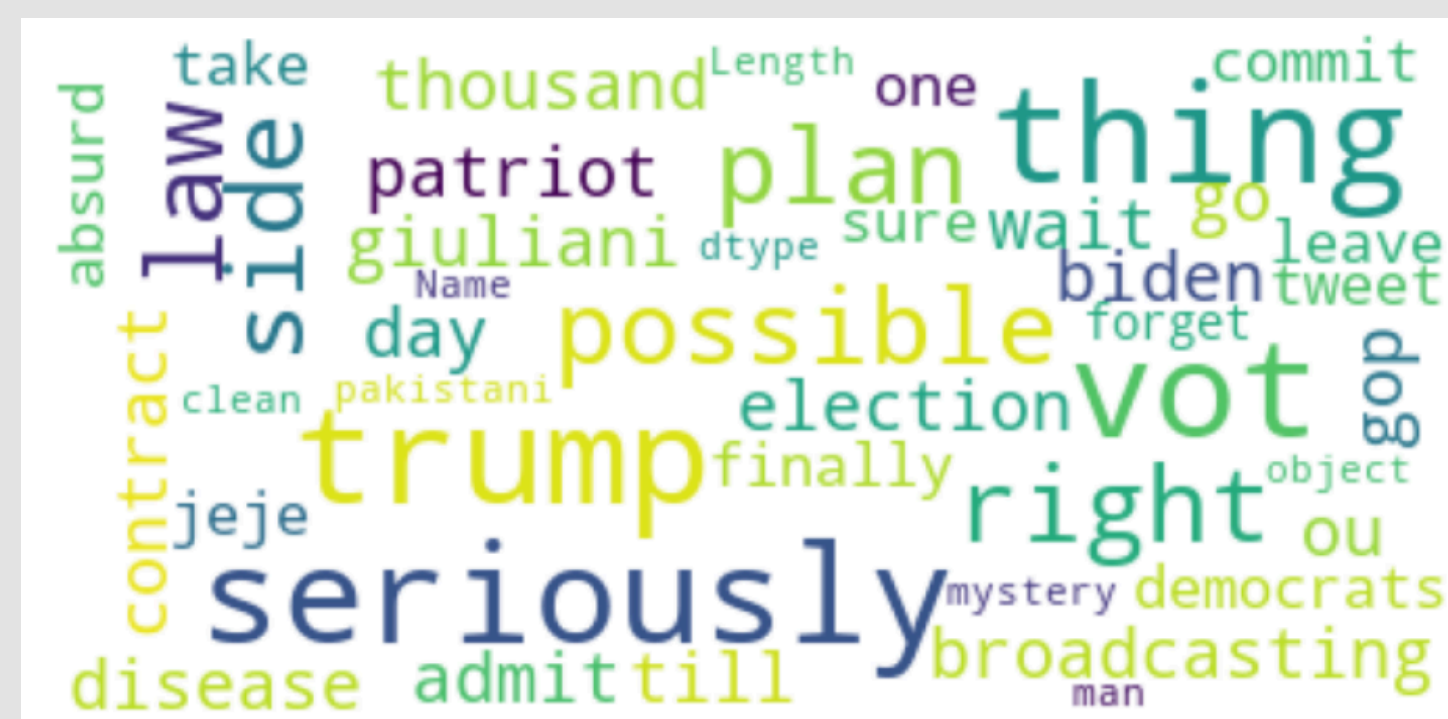
In the graph, above the frequency of the hashtag: '#voterfraud' is graphed from the dates 10/29/2020 - 12/05/2020. Two clear peaks around 11/04 and 12/01 are shown. 11/04 was election day and 12/01 is when Attorney General Barr declared that there was no evidence of widespread fraud in the 2020 presidential election.

Materials

Google Colaboratory. Twitter Database. Python. Topic Modeling. pyLDAvis. Indiana University Observatory on Social Media Trend Tool.

Methodology

With the help of Professor Chun, I collected 36,000 tweets using Google Colab and Python. The program scraped Twitter, collecting Donald Trump's tweets from the past two years, all of Joe Biden's tweets, and all tweets on the topic of voter fraud. The tweets were cleaned of all the hashtags, @, emojis and links. Next, I used exploratory data analysis and pyLDAvis to make visualizations. EDA visualizations made visuals charting tweet length, word cloud (word frequency), unigrams, bigrams, and trigrams. Exploratory data analysis is when a cleaned data set is analyzed by a human and with the external knowledge the analyzer can identify relationships and correlations between different variables. EDA is used to gain insight into the hidden structure and connections amongst the data, letting the data tell the story. Then I used topic modeling, an application of natural language processing, which is a statistical model used to find hidden topics that occur within a group of text. The topic modeling program LDA (Latent Dirichlet Allocation) finds the probability of hidden groups and different topics and insights, by searching the data set (in this case tweets) for common words and word frequency. I reviewed the generated topics and labeled them. I used my context of the world and labeled them appropriately to fit what they are likely being used to talk about. With pyLDAvis I was able to create visualizations. All the different visualizations, topics, and word frequency charts below all tell a story about our electoral system intersection with social media.



The word cloud above was made using EDA. It shows the word frequency in the data set. The bigger the word the more it was used in the tweets. As it can be shown the word trump, seriously and vote are the most frequent.

unigram	count	bigram	count	trigram	count
0	fraud 10522	0	voter fraud 9270	0	evidence voter fraud 527
1	voter 9977	1	donald trump 1158	1	massive voter fraud 454
2	trump 8991	2	president trump 704	2	claim voter fraud 329
3	president 3730	3	fake news 599	3	widespread voter fraud 281
4	election 3623	4	evidence voter 531	4	michigan secretary state 271
5	amp 3211	5	joe biden 531	5	systemic voter fraud 256
6	people 3130	6	united states 509	6	allege voter fraud 246
7	vote 2901	7	white house 461	7	voter fraud detroit 232
8	biden 2686	8	massive voter 456	8	claim systemic voter 229
9	say 2602	9	trump win 345	9	disqualify single ballot 227

The unigram, bigram, and trigram show the most common words and word combinations among the tweets collected.

Results

Below are the results of the topic modeling analysis of the tweets collected. There were 50 different topics generated. After looking at the different groups, I picked out the most unified and interesting ones. I labeled the groups with a title I felt the words encompassed. The topics show different points of view concerning election fraud in the 2020 presidential election.

- MONEY
 - Top 10 words for topic #3: ['india', 'paycheck', 'june', 'unprecedented', 'endorse', 'bob', 'capacity', 'income', 'medicare', 'tax']
 - This grouping is of interest because of the frequency of words relating to money. Many Americans use their vote to reach financial security. This security could be gained through jobs, taxes, or insurance. The fact that this is the 3rd topic makes it clear that many Americans see this election, and the claims of fraud, as a threat to their financial security. This pressing topic for many, especially during the current economic recession leads individuals to vividly speak out about their thoughts and concerns on Twitter.
- TRADITIONALIST
 - Top 10 words for topic #4: ['deal', 'tell', 'white', 'support', 'house', 'americans', 'donald', 'country', 'president', 'trump']
 - In the past four years as Trump has been president a conservative right has risen, again, in popularity, Many of the pillars of their stance lies in family, God, and tradition. With these tweets, it becomes clear that the claim of election fraud empowers, and motivates this conservative-leaning group to speak out against electoral fraud.
- INTERNATIONAL AFFAIRS
 - Top 10 words for topic #5: ['chaos', 'confront', 'eat', 'germany', 'japan', 'steve', 'veteran', 'disaster', 'que', 'honor']
 - Another pillar of election stress is international relations. With Covid-19 and the Trump presidency, US international relationships have become unstable. Leaving many concerned about future conflict, trade, and cooperation. Fear is held from all US citizens from immigrants to active military personnel. There can also be a connection drawn between election fraud and foreign interference, leaving the American people unsure if their democracy is working.

- CORONAVIRUS
 - Top 10 words for topic #7: ['reality', 'presidency', 'spread', 'washington', 'happy', 'actually', 'action', 'virus', 'health', 'believe']
 - The Coronavirus pandemic has impacted all American's lives and in turn, strongly impacted our election. From opinions to how the Trump administration handled the pandemic to the increase in the mail-in votes, the data has shown that the pandemic has played a role in the election. President Donald Trump has spoken out dismissing the integrity of mail-in votes. This lack of confidence in our election system from the President imbeds fear and mistrust upon an already pandemic stricken country.

Conclusion

An examination of the selected topics, word clusters, and visualizations from the topic modeling process makes it clear that election fraud is being discussed from many different angles. All of the different topics contain a common link of fear. Tweets coming from people who believe the election was fraudulent, in favor of Donald Trump fear that Joe Biden will undeservingly gain power, discrediting our electoral system. Those who believe the election was fair, in favor of Joe Biden, fear that others challenging a fair election will crumble our democracy.

Throughout this project it has been shown that people turn to media spaces, such as Twitter, to have a platform on which they can share beliefs. Or what they think they believe. What becomes clear is that angry, fearful, and anxiety-driven messages tend to be most prolific on social media, so it is understandable why fear is the linking factor amongst the different topics.

This research project raises questions about the interplay between Twitter and politics. It shows the amplification that social media has towards controversial claims and misinformation. With further research, it would be interesting to understand the role that bots play in this amplification. I also concluded that social media is a vehicle for anxiety about other issues. Within one topic, in this case voter fraud, individual's concerns about other issues shine through. The data that social media has is precious. There is much more research to be done surrounding politics and social media. To understand the role it truly plays we must analyze the connection at a larger scale and among different topics. Using topic modeling was very informative in this case, however, it would be interesting to use sentiment analysis to understand the emotions in political tweets. With the fast advancements in technology and data analysis, our abilities to understand social media is becoming more efficient, meaningful, and necessary. More research is needed in order to produce concrete and non-biased conclusions. The models produced in this project are useful in understanding Twitter and politics, however, there is much more work to be done in order to solidify the many different findings.

The US electoral system is a main element of our modern democracy. The integration of social media and technology within our culture has proven its capabilities of hacking the human mind. This generation's major threat to democracy is technological mass manipulation of US citizens. Because of this, we can not stop questioning the intersection of technology and politics. This project not only proves the power and information that social media holds but also places value on our ability to express diverse viewpoints. This is an ability we must protect.

Acknowledgements

Davis CA, Ciampaglia GL, Aiello LM, Chung K, Conover MD, Ferrara E, Flammini A, Fox GC, Gao X, Gonçalves B, Grabowicz PA, Hong K, Hui P, McCauley S, McKelvey K, Meiss MR, Patil S, Pei Kankanamalgale C, Pentchev V, Qiu J, Ratkiewicz J, Rudnick A, Serrette B, Shiralkar P, Varol O, Weng L, Wu T, Younge AJ, Menczer F. (2016) OSoMe: the IUNI observatory on social media. PeerJ Computer Science 2:e87 <https://doi.org/10.7717/peerj-cs.87>

Hajjej, Adam. "Trump Tweets: Topic Modeling Using Latent Dirichlet Allocation." *Medium*, Data Driven Investor, 19 June 2020, medium.com/datadriveninvestor/trump-tweets-topic-modeling-using-latent-dirichlet-allocation-e4f93b90b6fe.

Levitt, Justin. "The Truth About Voter Fraud." SSRN Electronic Journal, 2007, doi:10.2139/ssrn.1647224.

"US Attorney General Finds 'No Voter Fraud That Could Overturn Election'." *BBC News*, BBC, 2 Dec. 2020, www.bbc.com/news/world-us-canada-55153366.

Parmelee, John H., and Shannon L. Bichard. Politics and the Twitter Revolution : How Tweets Influence the Relationship Between Political Leaders and the Public, Lexington Books, 2011. ProQuest Ebook Central, <https://ebookcentral-proquest-com.libproxy.kenyon.edu/lib/kenyon-ebooks/detail.action?docId=912293>.

Viriyakovithya, Ekapope. "How to Combine Multiple CSV Files with 8 Lines of Code." *FreeCodeCamp.org*, FreeCodeCamp.org, 5 July 2019, www.freecodecamp.org/news/how-to-combine-multiple-csv-files-with-8-lines-of-code-265183e0854/.