

Project 1 Hypothesis & Model Plan

Group members: Reese Quillian, Lauren Smith, Ann Sofo

Leader: Ann Sofo

DS 4002

Thursday Feb 9

I. Narrative paragraph

Donald Trump was recently allowed back onto Facebook and Instagram after a two-year ban resurfacing the conversation of the former president's controversial social media presence. His extensive use of Twitter allowed him to communicate with and appeal to voters who also had a heavy presence on social media. His rhetoric online has identified him as having "unpolished character," sending tweets described as "personal" (Kuś, 2020). In the past, Trump has used two Twitter accounts: @POTUS and @realDonaldTrump, and was moderately active on his personal account after opening it in 2009 up until 2016 when he was tweeting almost 375 tweets per month during the campaign (Kuś, 2020). In this project, we will be exploring tweets from the former president's personal Twitter account. Trump had started tweeting attacks on Democrats and Obama in 2012 and even his famous slogan "Make America Great Again" for the first time in that same year (Carr, 2018). After he was elected, he said he wouldn't use Twitter for communication, but he had fewer opportunities to speak personally to Americans since he could not hold rallies as often, leaving him to pour his thoughts into his tweets (Carr, 2018). Even though many viewed his tweets as unprofessional, they seized the attention of Americans and he gained more listeners, an important task for politicians.

II. Hypothesis

We believe that the presence of six words/phrases [*"Wall", "fake news", "media", "Democrat", "great", "proud boys"*] signifies that a tweet was posted after Donald Trump became president.

III. Research Question

Did Donald Trump's Twitter language change over the course of his campaign/election year (before, during, after his election)?

IV. Modeling Approach

We will be using a clustering algorithm to define groups within our data. Clustering algorithms take in unlabeled data (in this case, Donald Trump's tweets between 2009 and 2021) and look for meaningful patterns based on similarities of features. We expect to see clusters defined based on the presence of certain words and time period that the tweet was posted. However, we're most interested in learning from the model's ability to detect groupings in our dataset that we would not be able to easily detect ourselves. The fact that clustering is an unsupervised learning method works to our benefit; we're able to quickly identify the similarities and differences between our data points and determine which features are significant.

V. References

- Anderson, B. (2017). *Tweeter-in-Chief: A Content Analysis of President Trump's Tweeting Habits*.
- Carr, N. (2018, January 26). *Why Trump Tweets (And Why We Listen)*. POLITICO Magazine.
<http://politi.co/2BvZsIV>
- Clustering Algorithms | Machine Learning*. (n.d.). Google Developers. Retrieved February 9, 2023, from <https://developers.google.com/machine-learning/clustering/clustering-algorithms>
- Donald Trump Tweets Dataset*. (n.d.). Retrieved February 9, 2023, from <https://www.kaggle.com/datasets/codebreaker619/donald-trump-tweets-dataset>
- Kuś, R. (2020). Building a Trump Nation: The Rhetoric of President Donald J. Trump on Twitter. *Zarządzanie Mediami*, 8(4), 391–405. <https://doi.org/10.4467/23540214ZM.20.041.12645>
- Masha Maksimava. (2019, January 10). *Analyzing Trump's Twitter: Top themes from 36K tweets*. Awario Blog. <https://awario.com/blog/trump-twitter-analysis/>
- Milecia McGregor. (2020, September 21). *8 Clustering Algorithms in Machine Learning that All Data Scientists Should Know*. FreeCodeCamp.Org.
<https://www.freecodecamp.org/news/8-clustering-algorithms-in-machine-learning-that-all-data-scientists-should-know/>
- Trump Twitter Archive | Kaggle*. (n.d.). Retrieved February 9, 2023, from <https://www.kaggle.com/datasets/headsortails/trump-twitter-archive>