

Investigating the correlation between twitter sentiment with campaigns and election results

DS4002 Case Study by Arjuna Bazaz



In a famous debate between John F Kennedy and Richard Nixon, radio listeners and TV watchers had different opinions on the outcome. This discrepancy, resulting from Kennedy's more photogenic presentation, highlighted the importance of understanding the medium you are in. The new frontier of this phenomenon is social media. Campaigns are run and won on social media, and it is the most common place people go to see other's reactions to things, and express their own. You are tasked with designing a way to gauge the reactions of social media to particular events, with the second 2020 presidential debate and the occurrence that you are conducting the initial experiment on.

Your goal is to gather a dataset of tweets from 2020 around the time of the presidential election and conduct a natural language processing analysis to assess the sentiments of the tweets. Afterwards, statistical analysis should be conducted to assess the impact of the debate both before and afterwards. This can then be compared to actual results of the debate from polling, and the actual election results to see whether the results hold.

Despite the complexity of the task, your results will affect the way many campaigns from here on out are run and how they engage with social media. It will also affect the way that social media companies should be covered and regulated in the modern world. As a result, this project should be taken with seriousness and consideration of your steps.

Everything you need can be found in the documents in this folder and through the link:
https://github.com/ArjunaBazaz/CS3_2020_Presidential_Approval_Rating_Analysis.git