

Documentation:

Data

1. biden_tweets.csv
 - a. Raw data of Biden's tweets (outdated)
2. trump_tweets.csv
 - a. Raw data of Trump's tweets (outdated)
3. allcombined.csv
 - a. Cleaned, combined data (outdated)
4. cv_data.csv
 - a. Used for training and cross-validation (outdated)
5. valid_data.csv
 - a. Used for validation (testing) (outdated)
6. However, we will only be using the data that Neil sent (which is the allcombined.csv file with additional factors that we got from OpenAI's moderation endpoint).
 - a. So, we will only need to download Neil's file called **backup_final.parquet** when refining the project.

Mining

1. BERT Emotion Analysis.qmd
 - a. Mined for values of SADNESS, ANGER, JOY
2. VADER Sentiment Analysis.qmd
 - a. Mined for positivity, negativity, and compound scores (-1.00 being most negative and 1.00 being most positive)
3. Violence Analysis.qmd
 - a. Mined for violence factors
4. Toxicity Analysis.qmd
 - a. Mined for toxic factors

Cleaning/EDA

1. Preliminary Data Cleaning.qmd
 - a. Where we combined Trump and Biden data sets; not too useful
2. EDA.qmd
 - a. EDA (outdated since Neil's new data)
3. Word Cloud.rmd
 - a. (outdated since Neil's new data)

*In general, most of the mining files are not too useful except for understanding the factors. The Cleaning/EDA files are also not too useful (we only changed variable names and combined data sets). EDA.qmd is decent for visualizing data.

*Our models might need some improvement

Models

1. Linear Model - Relaxed LASSO.rmd
2. Neural Network.rmd
3. Random Forest.qmd

Abstract:

https://docs.google.com/document/d/1-1Mx4D8xHJ6fxz-1iARstPhEv_NoB_gOonn_WRY93p8/edit?tab=t.0

Google Slides of Presentation:

https://docs.google.com/presentation/d/1oKXYm9rev8Z78YTLryasOn5EtTGgM_L7FxTvCu4ue5c/edit#slide=id.g2ef4b8054fd_1_0

Documentation:

Data

7. biden_tweets.csv
 - a. Raw data of Biden's tweets
8. trump_tweets.csv
 - a. Raw data of Trump's tweets
9. allcombined.csv
 - a. Cleaned, combined data
10. cv_data.csv
 - a. Used for training and cross-validation
11. valid_data.csv
 - a. Used for validation (testing)

Mining

5. BERT Emotion Analysis.qmd
6. VADER Sentiment Analysis.qmd
7. Violence Analysis.qmd
8. Toxicity Analysis.qmd

Cleaning/EDA

4. Preliminary Data Analysis.qmd
5. EDA.qmd
6. Word Cloud.rmd

Models

4. Linear Model - Relaxed LASSO.rmd
5. Neural Network.rmd
6. Random Forest.qmd

Google Slides of Presentations:

https://docs.google.com/presentation/d/1oKXYm9rev8Z78YTLryasOn5EtTGgM_L7FxTvCu4ue5c/edit#slide=id.g2ef4b8054fd_1_0