ANALYSIS OF TRUMP TWEETS

Joey Livorno

INTRODUCTION

- During the second day of his impeachment hearing, Donald Trump sent a record 142 tweets (MarketWatch)
- Twitter is an excellent source for Linguistic data
- One of the most notorious Twitter users just so happens to be our President
- Donald Trump's Twitter page is filled with data that is simply begging to be analyzed

MY MOTIVATION

- I'm an Information Science Major
- I'm addicted to Twitter
- Donald Trump is constantly in the limelight
- I'm not very interested in politics
- But what better way to learn more?

BACKGROUND INFORMATION

- Donald Trump is the 45th President of the United States
- He began his first presidential campaign in 2015
- Donald Trump uses his Twitter in for ways that no other President has done
  - Obviously, Barack Obama was the only other President to have this tool
  - He uses it to keep the American people constantly engaged in what he is thinking and doing

BACKGROUND INFORMATION (CONT.)

- Donald Trump has been tweeting since 2009
- Most of his tweets were made after 2015
- He is notorious for using his Twitter recklessly
- Are there any conclusions we can draw from all this data?

PROJECT PROPOSAL

- Analyze Donald Trump's Twitter feed

- Draw conclusions about his tweeting habits:
  - Trends
  - Sentiment
  - Machine Learning Model

HYPOTHESES:

- Donald Trump's twitter influence most likely has grown overtime
- A sentiment analysis of his tweet's regarding certain issues should be consistent with the idea that he "flip flops" (i.e. he changes stances on issues when convenient)
- Donald Trump's unique tweeting style should allow for a successful prediction model when compared to other tweets

TOOLS

- Pandas, NumPy, nltk
- TextBlob
- Scikit-learn
  - TfidfVectorizer
  - Multinomial Naïve Bayes
  - GridSearchCV

A SHORT DESCRIPTION OF TEXTBLOB

- A python library for processing textual data
- Provides a simple way to access elements of Natural Language Processing
  - Part of speech tagging
  - Noun phrase extraction
  - Sentiment analysis
- I use TextBlob to expedite sentiment analysis

DATA PROCESSING

- First option to gather this data would be TweePy
  - Much more work-intensive
  - Limited to 3200 tweets
  - Trump's twitter feed is already well documented on the internet

- Trump Twitter Archive
  - 44,000 Tweets
  - Includes retweets

- Followthehashtag
  - 200,000 tweets
  - Comes from random Twitter users

DATA PROCESSING (CONT.)

- Download tweets as .csv
- Add necessary columns
  - Polarity
  - Subjectivity
  - Year
- Tokenize and lowercase

	source	text	created_at	retweet_count	favorite_count	is_retweet	id_str	polarity	subjectivity	year
0	Twitter for iPhone	Crazy “Nancy Pelosi you are a weak person. You...	2020-04-16 13:33:08	9593.0	34528.0	FALSE	1250779261595783168	-0.3875	0.514583	2020
1	Twitter for iPhone	.@OANN Poll “Gives President Trump a 52% Appro...	2020-04-16 13:12:40	5915.0	25884.0	FALSE	1250774109144875008	-0.1250	0.375000	2020
2	Twitter for iPhone	RT @WhiteHouse: The Federal government has ord...	2020-04-15 22:51:36	8035.0	0.0	TRUE	1250557416078532617	0.0000	0.000000	2020
3	Twitter for iPhone	RT @WhiteHouse: In 11 days @GM sourced materia...	2020-04-15 22:51:22	7494.0	0.0	TRUE	1250557356238389248	0.0000	0.000000	2020
4	Twitter for iPhone	RT @WhiteHouse: We are using every available a...	2020-04-15 22:51:20	7360.0	0.0	TRUE	1250557346193031168	0.4000	0.400000	2020

DATA PROCESSING (CONT.)

- I then created subsets for various political topics

- Russia
  - Russia, Moscow, putin
  - 716 tweets

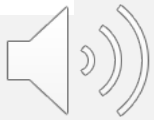
- Iran
  - Iran, Tehran, nuclear deal, rouhani
  - 437 tweets

- China
  - China, Beijing, xi, jinping
  - 1638 tweets

DATA PROCESSING (CONT.)

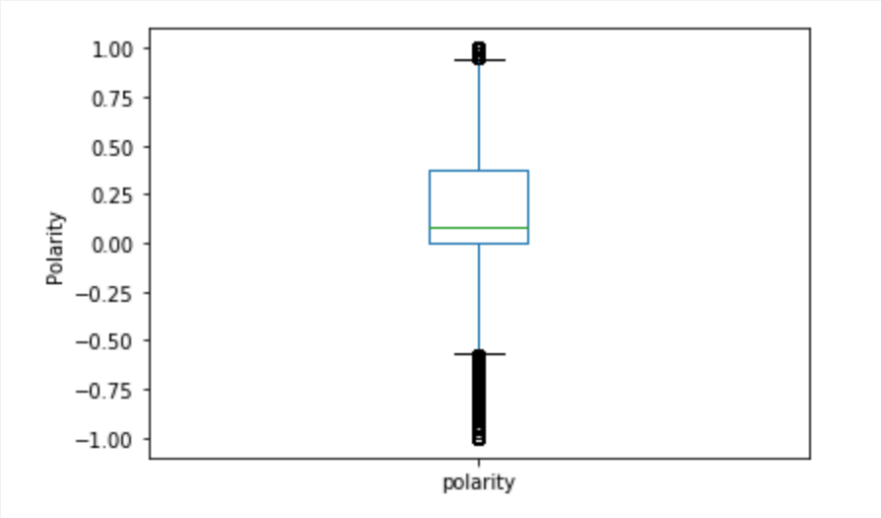
- Read in generic tweets as .csv
- Place tweet content of each set in separate dataframes
- Tag Trump tweets as T and others as NT
- Concatenate the dataframes
- Pickle

	text	code
0	Wind 3.2 mph NNE. Barometer 30.20 in, Rising s...	NT
1	Pausa pro café antes de embarcar no próximo vô...	NT
2	Good. Morning. #morning #Saturday #diner #VT #...	NT
3	@gratefuldead recordstoredayus _____ @ TOMS M...	NT
4	Egg in a muffin!!! (@ Rocket Baby Bakery - @ro...	NT



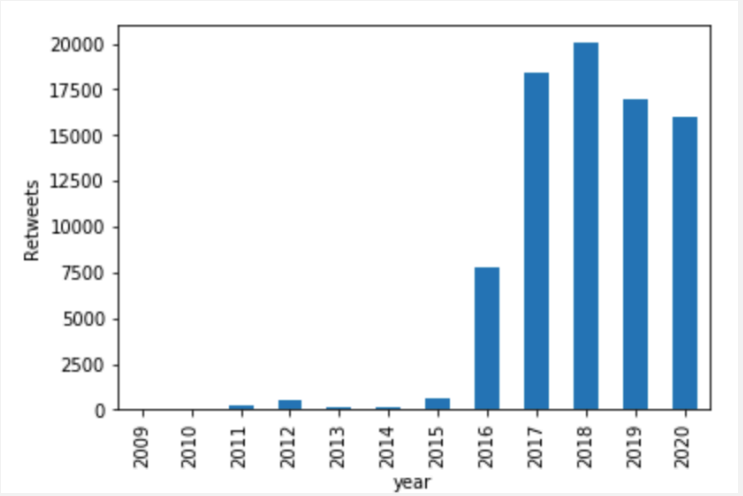
EXPLORATORY DATA ANALYSIS

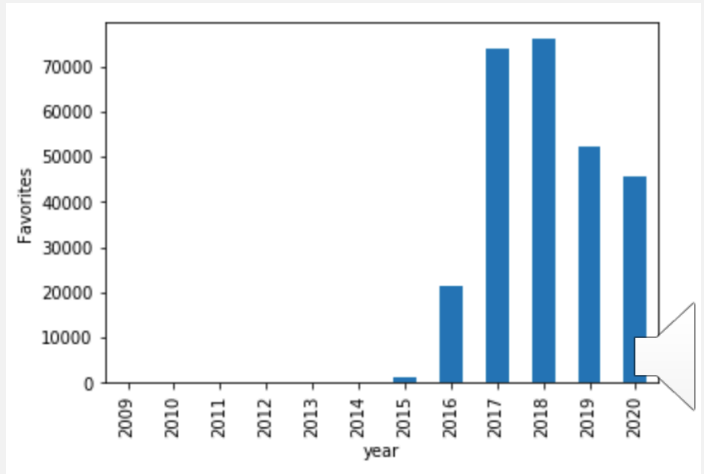
- Began with .describe()
  - Average sentiment was about .17
  - Median was about .08
  - Because median is less than mean, we can expect the data to be skewed right
- Counterintuitive to what we would expect from Trump's twitter
- An interesting first discovery



EXPLORATORY DATA ANALYSIS (CONT.)

- 2019 was when he tweeted the most, but average tweet was in August 2015
- Data is incomplete, but 2018 showed his most retweets and favorites
- Most favorited and retweeted tweet
  - 369,530 retweets
  - 879,647 favorites





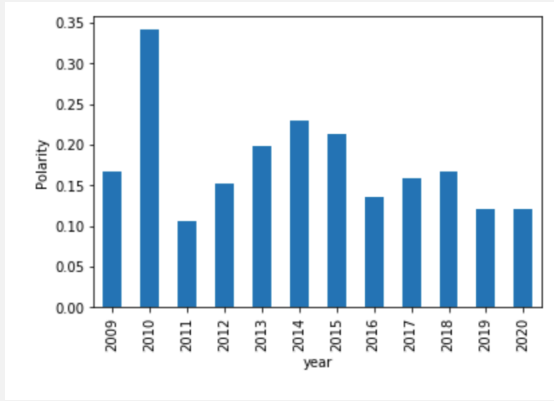
EXPLORATORY DATA ANALYSIS (CONT.)

- Learned basic statistics of Trump's Twitter
- Trends are consistent with the idea that his influence has grown overtime
- Interesting, but nothing is very conclusive

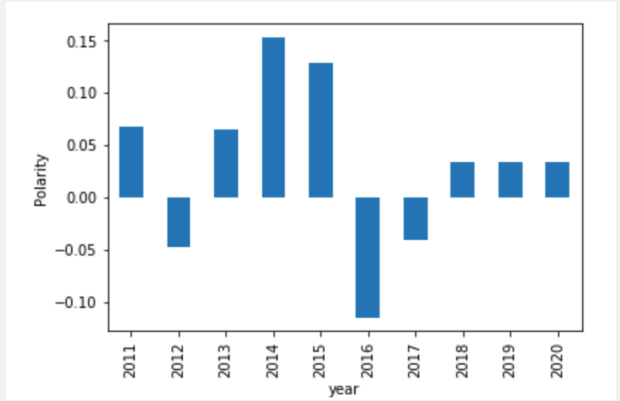
LINGUISTIC ANALYSIS

- Overall sentiment by year shows that other than 2009 and 2010, twitter has remained mostly high-neutral
- Russia
  - Big drop off from 2014-2016
  - 2019 was a significant increase
- Iran consistently low
  - 2016 showed the lowest
  - 2019 abnormally high
- China was consistently high neutral

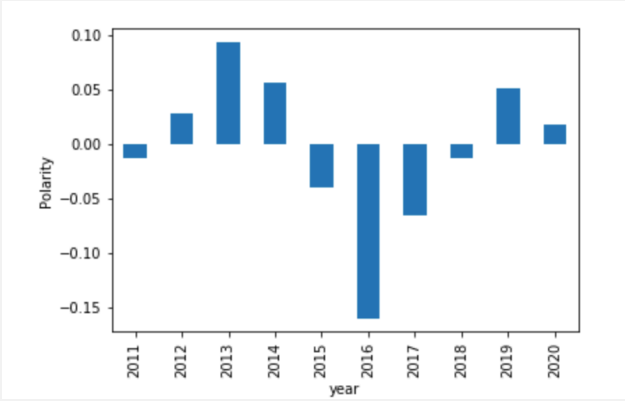
Overall



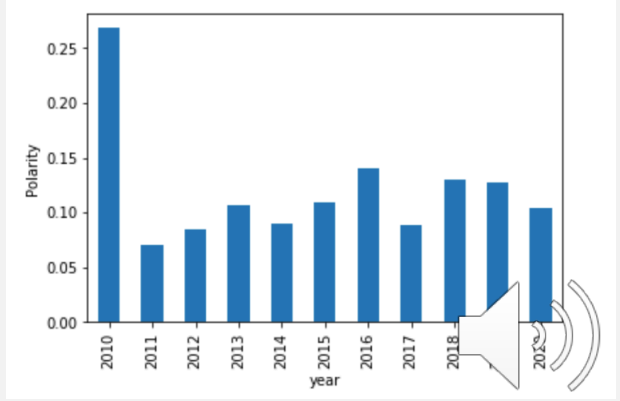
Russia



Iran



China



LINGUISTIC ANALYSIS (CONT.)

- I would've expected China to be lower
- I cannot yet explain why Iran was so high in 2019
  - Conflict began in early 2020
  - Perhaps they were not yet on Trump's "shitlist"
- Russia is the most conclusive
  - Dropped significantly at the start of his campaign
  - Increase after 2017 might relate to the Mueller Investigation and Report

MACHINE LEARNING

- I split the concatenated dataframe into two series
- Created the pipeline consisting of TfidfVectorizer and MultinomialNB
- Created parameter dictionary
- Fit the GridSearchCV with the data
  - Changed Tfidf max_features and MNB alpha
  - 5 folds

MACHINE LEARNING (CONT.)

- Best result was about 93%
  - Mnb__alpha: .01
  - Tfidf__max_features: 9000
- MNB alpha value had a small impact, but .01 yielded slightly higher accuracy
- Accuracy increased directly with max features
- Possibly true for many people, but significant nonetheless

```
Best Results:
Score:
 0.9296214626975376
Parameters:
mnb__alpha: 0.01
tfidf__max_features: 9000

All Results:
  param_mnb__alpha param_tfidf__max_features  mean_test_score
0             0.01                      1500         0.906052
1             0.01                      3000         0.917273
2             0.01                      4500         0.922381
3             0.01                      9000         0.929621
4            0.001                      1500         0.906003
5            0.001                      3000         0.917138
6            0.001                      4500         0.922087
7            0.001                      9000         0.928972
```

KEY TAKEAWAYS

- Trump's Twitter influence has increased drastically since he started his first presidential campaign
- Trump uses this influence and strong language to defend his friends and attack his enemies
- Trump's tweets are easily identifiable when compared to others

STILL TO DO

- Incorporate random tweets into analysis to form a baseline
- Group tweets more generally by sentiment to show broader trends (low, low-neutral, high-neutral, high)
- Apply machine learning to new labels
- Become more familiar with events that may better explain the data

CONCLUSION

- Politicians have a lot of online influence
- Our tweets don't directly affect people
- With influence like Trump, tweets can have a direct effect on certain groups
- It's important to be aware of this influence and check those who are in positions to use it negatively

SOURCES

- http://www.trumptwitterarchive.com/archive
- http://followthehashtag.com
- https://www.marketwatch.com/story/this-is-trump-unleashed-these-charts-show-that-the-president-is-tweeting-and-speaking-more-than-ever-2019-09-23
- https://www.axios.com/mueller-russia-investigation-timeline-indictments-70433acd-9ef7-424d-aa01-b962ae5c9647.html
- https://textblob.readthedocs.io/en/dev/