

A DEMOCRATIC DEBATE ANALYSIS THROUGH THE LENS OF TWITTER

Presented by Sam Videlock

Presentation Outline



- Background - Democratic Primary 2020
- Project Goals
- The Data
- Sentiment Analysis with Deep Learning
- Sentiment Analysis of Debate Night Tweets Results
- Debate Through Hashtags
- Background - Social Spambots
- Bot Detection/Classification
- Bot Analysis on Democratic Debate Tweets
- Next Steps



ELIZABETH WARREN

Massachusetts Senator



JOE BIDEN

Former Vice President of the United States



PETE BUTTIGIEG

Mayor of South Bend, Indiana



BERNIE SANDERS

Vermont Senator



AMY KLOBUCHAR

Minnesota Senator



MIKE BLOOMBERG

Former Mayor of New York City

The Candidates in The Study

Project Goals



SENTIMENT ANALYSIS

Determine the sentiment of each tweet using Deep Learning

Use these sentiments for in depth analysis of each candidates Twitter sentiment during each debate



DATA ANALYSIS

Determine other insights using the gathered twitter data



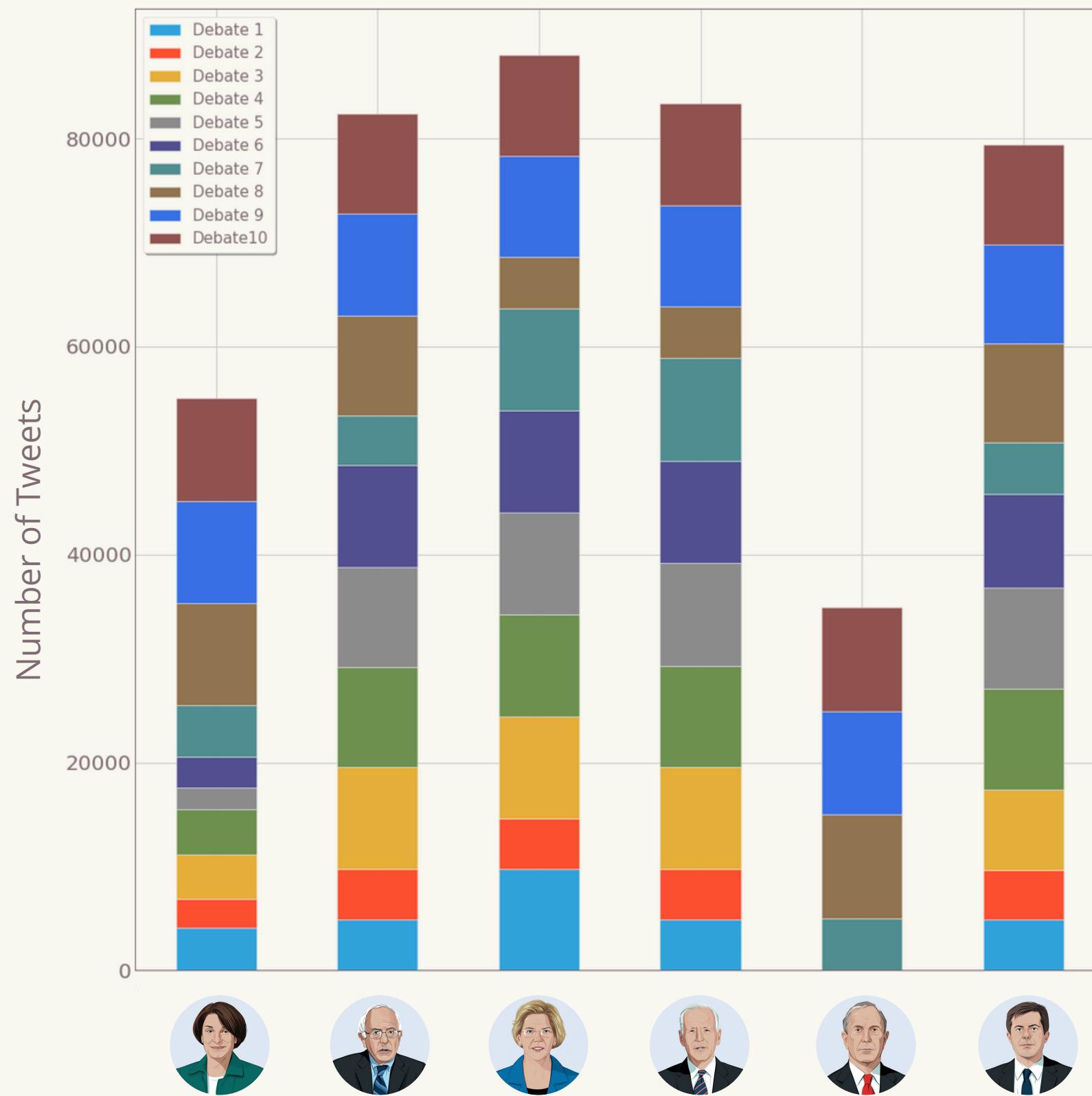
BOT DETECTION

Develop proof of concept using a classifier to identify whether a user is a bot or not

We will look only at Debate 9 tweets for this analysis

The Data

Total Data Points for Each Candidate



GETOLDTWEETS3 PYTHON PACKAGE

Used this python package to obtain debate night tweets for each candidate for analysis-- resulted in over 440,000 tweets for 6 candidates -- Sentiment Analysis

TWITTER API

Used the Twitter API to obtain specific user data for the bot detection model, this resulted in gathering data for ~20,000 users who tweeted during Debate 9 and over 900,000 of their tweets-- Spambot Classification

DATA CONCERNS

Due to limitations of accessibility of Twitter data, not every tweet from every debate is able to be accessed (without paying for it). The hope is that the data that was able to be gathered is representative of the greater Twitter population, however this is difficult to measure without more data

Sentiment Analysis with Deep Learning

Bidirectional Recurrent Neural Network (BRNN)

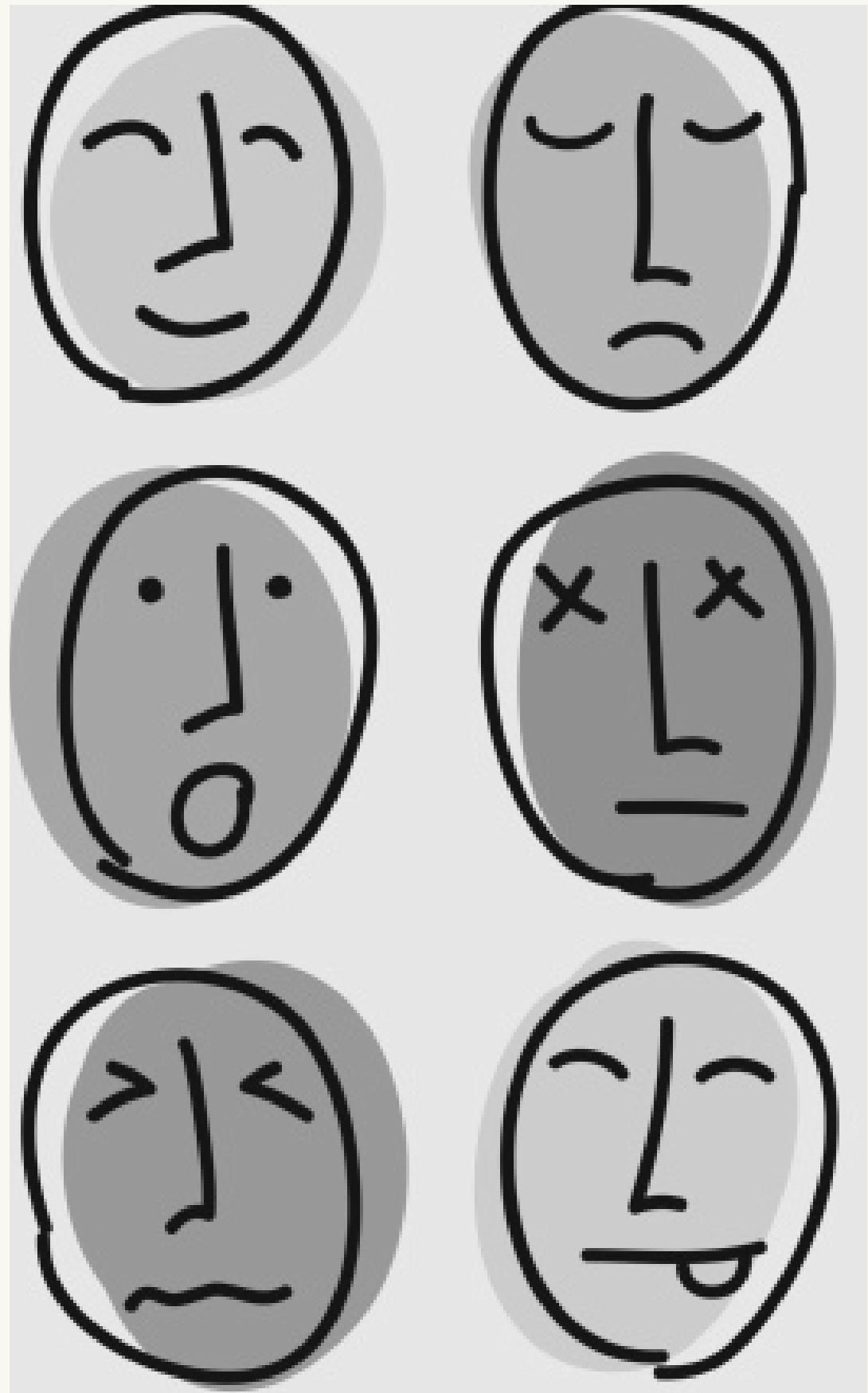
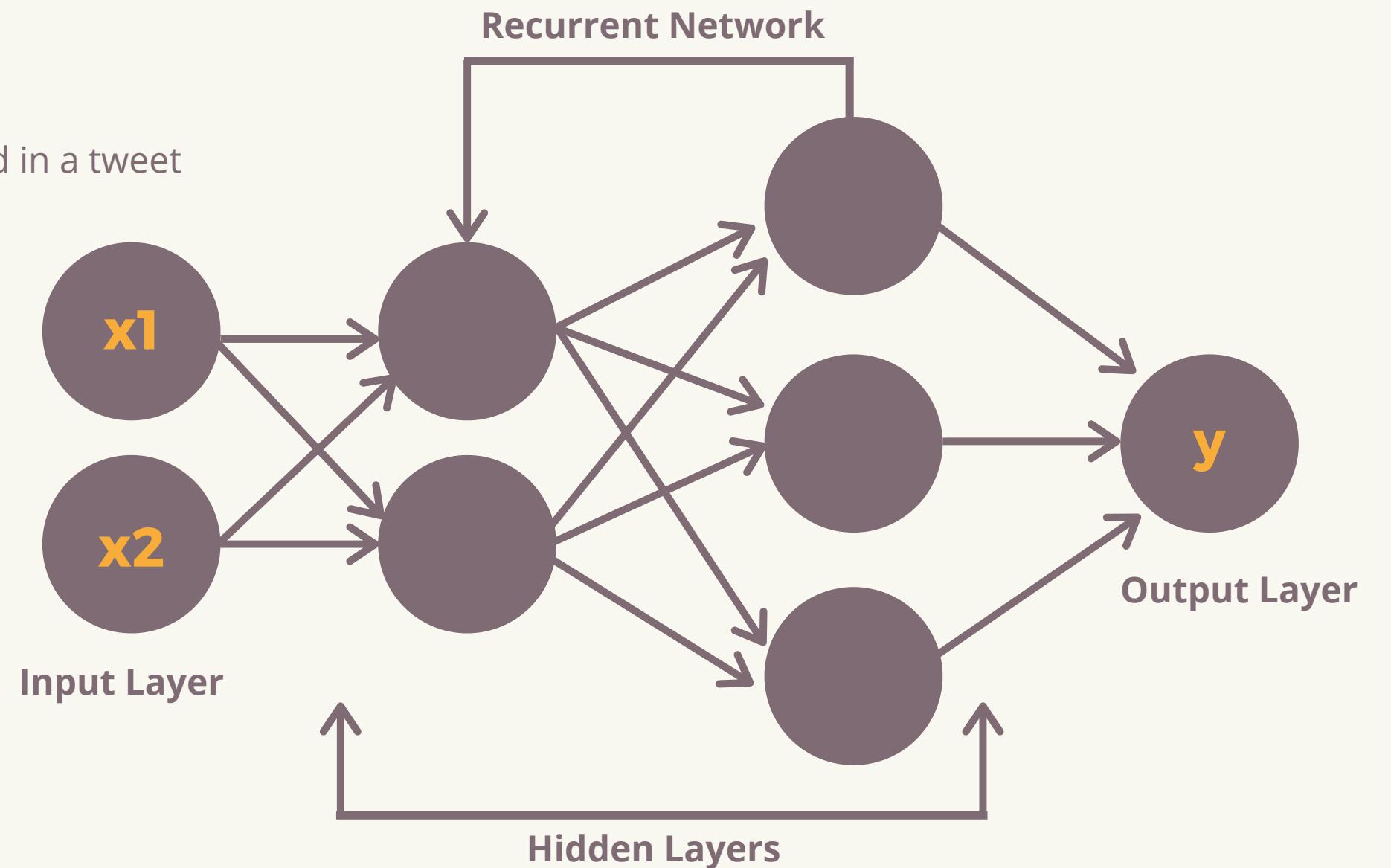
RNNs have in them a sense of some memory about what happened earlier in the sequence of data. This helps the system to gain context. In the case of analyzing sentences this becomes important to a successful sentiment analysis

This was a pretrained BRNN that was trained on Movie Reviews.

Limitations:

Multi-polarity

Two candidates mentioned in a tweet

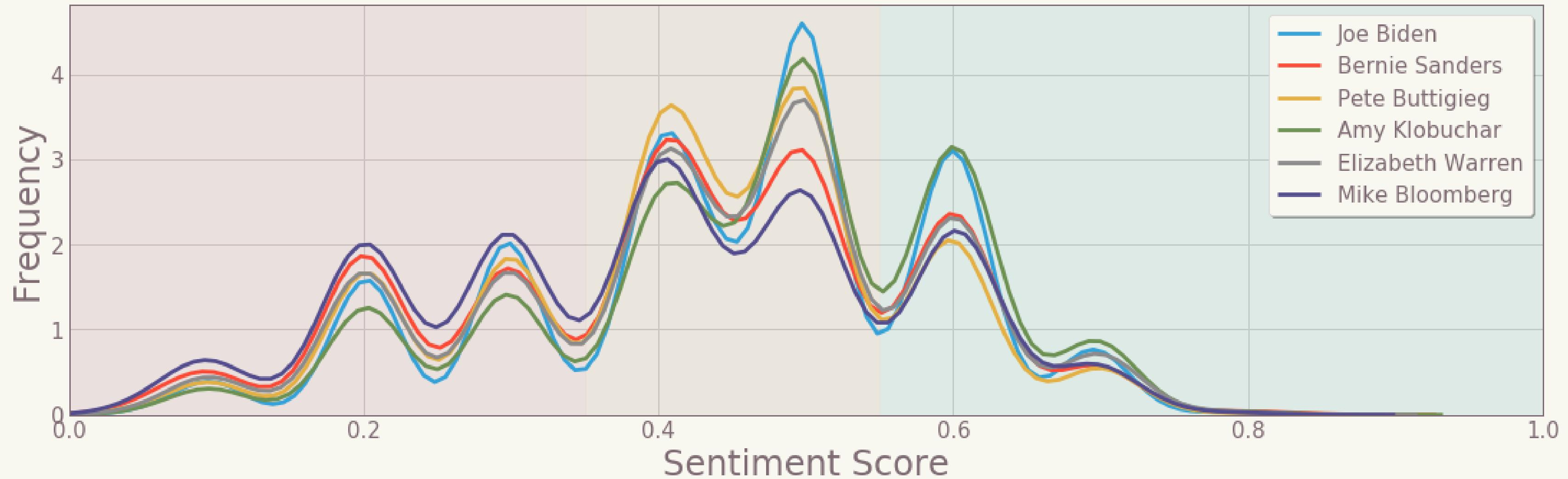


Average Sentiment Over Time for Each Candidate



Sentiment Analysis

Debate 9



Sentiment skews negative for all candidates in all debates. I have adjusted tweets in the following categories:

- Negative: $<= 0.35$
- Neutral: $0.35 - 0.55$
- Positive: $>= 0.55$

A reminder that what's most important right now is the momentum and energy that @BernieSanders + the #NotMeUs movement is creating. It's AMAZING! Our dedication to the #Bernie2020 ground game is bearing incredible fruit. Keep it going! #ITrustBernie #WomenForBernie

0.81



Bernie **Sanders** wants 77% death tax. 77%. So I can leave my son 23% of what I have worked my entire life to acquire?

0.05

Biden is the best for 2020

0.80



Sorry, **Biden** was a mess in the last debate. I found myself getting so frustrated trying to listen to him as he tripped over every other word, & I WANTED to hear what he had to say. I like him, I'll vote for him if he's the nominee, but he'll be a disaster against Trump.

0.04

This is truly fantastic. @nytimes asked deep, detailed foreign policy questions of each candidate. Reading @ewarren answers was uplifting, incredibly inspiring- and again reminded me how brilliant & profoundly well informed she is.

0.85



It's a ridiculous low blow smearer. I thought Elizabeth was above this. A dirty Clinton trick. I have lost all respect for Elizabeth **Warren**. Where have 30 years a film footage of Sanders , we know his track record.

0.04

This is Worth reading. **Bloomberg** has been a great supporter of progressive causes and candidates without asking anything in return. Dems need his support and he's proven capable of winning and being president.

0.72

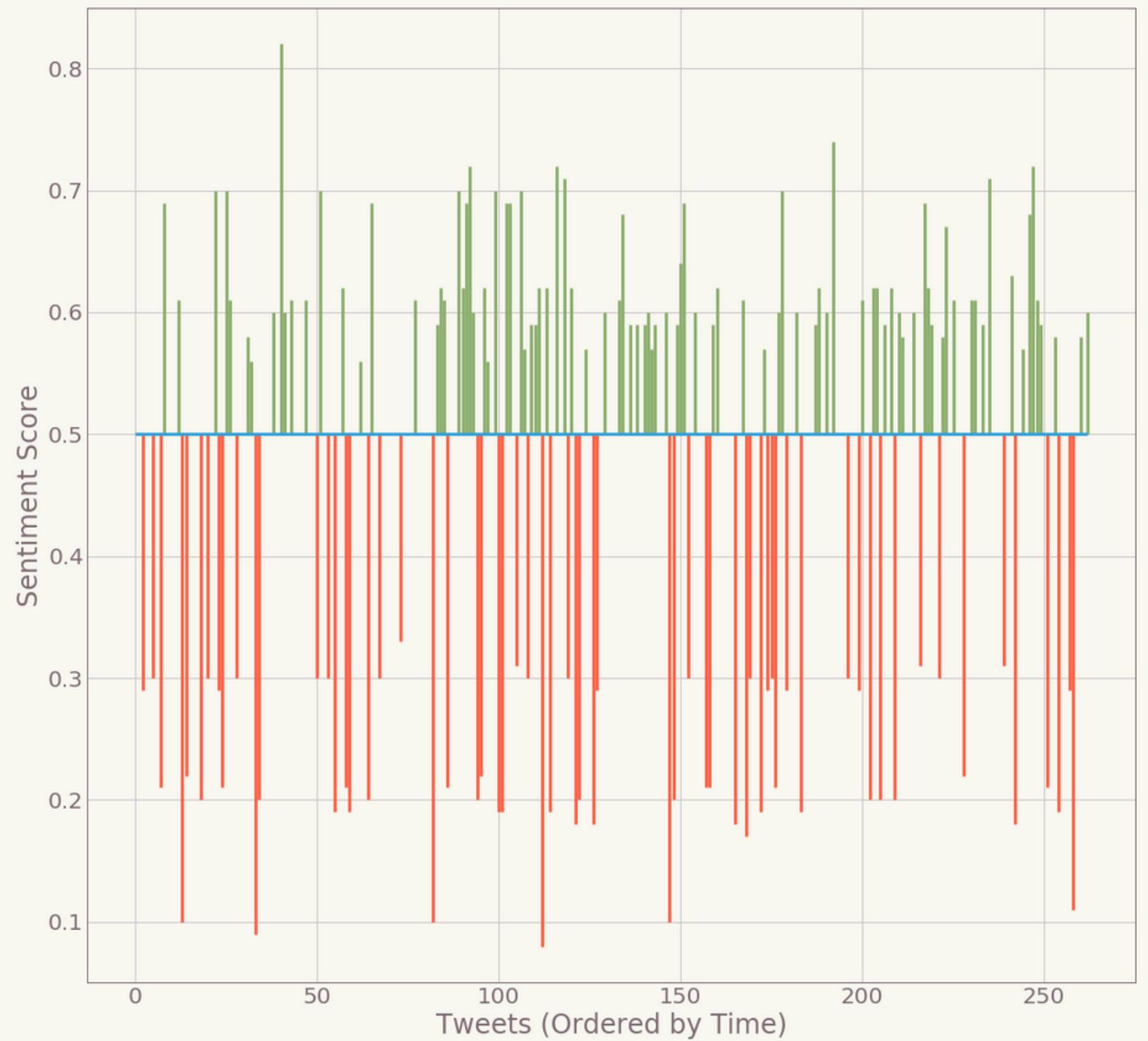


Me: lolol These ads Mike **Bloomberg** runs are just a waste of money. What kind of f*** idiot would actu- *Polls at 15% Me: **SPITS OUT TEA**

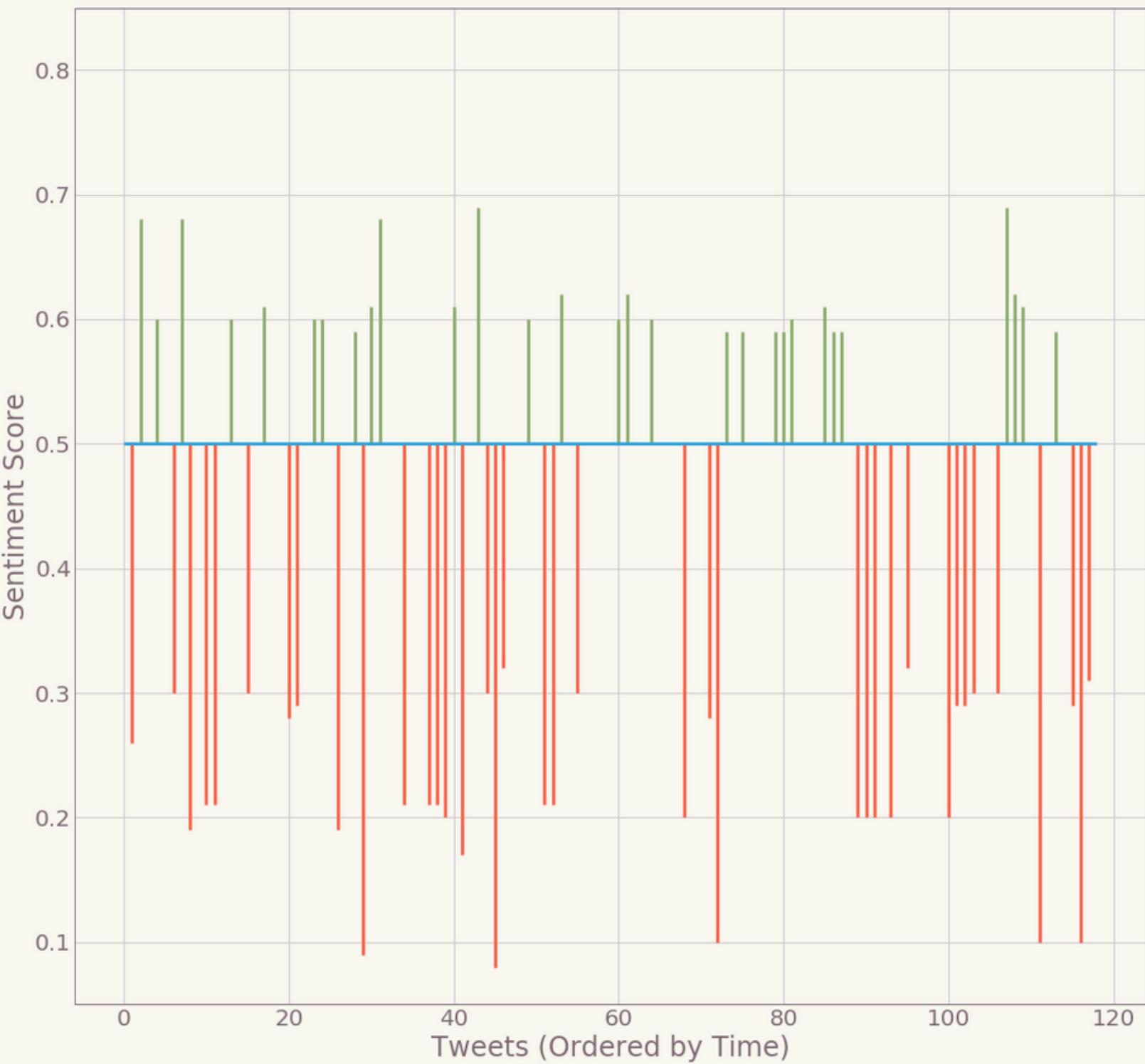
0.04

Positive & Negative Tweet Examples

Elizabeth Warren Debate 6 Sentiment



Elizabeth Warren Debate 7 Sentiment

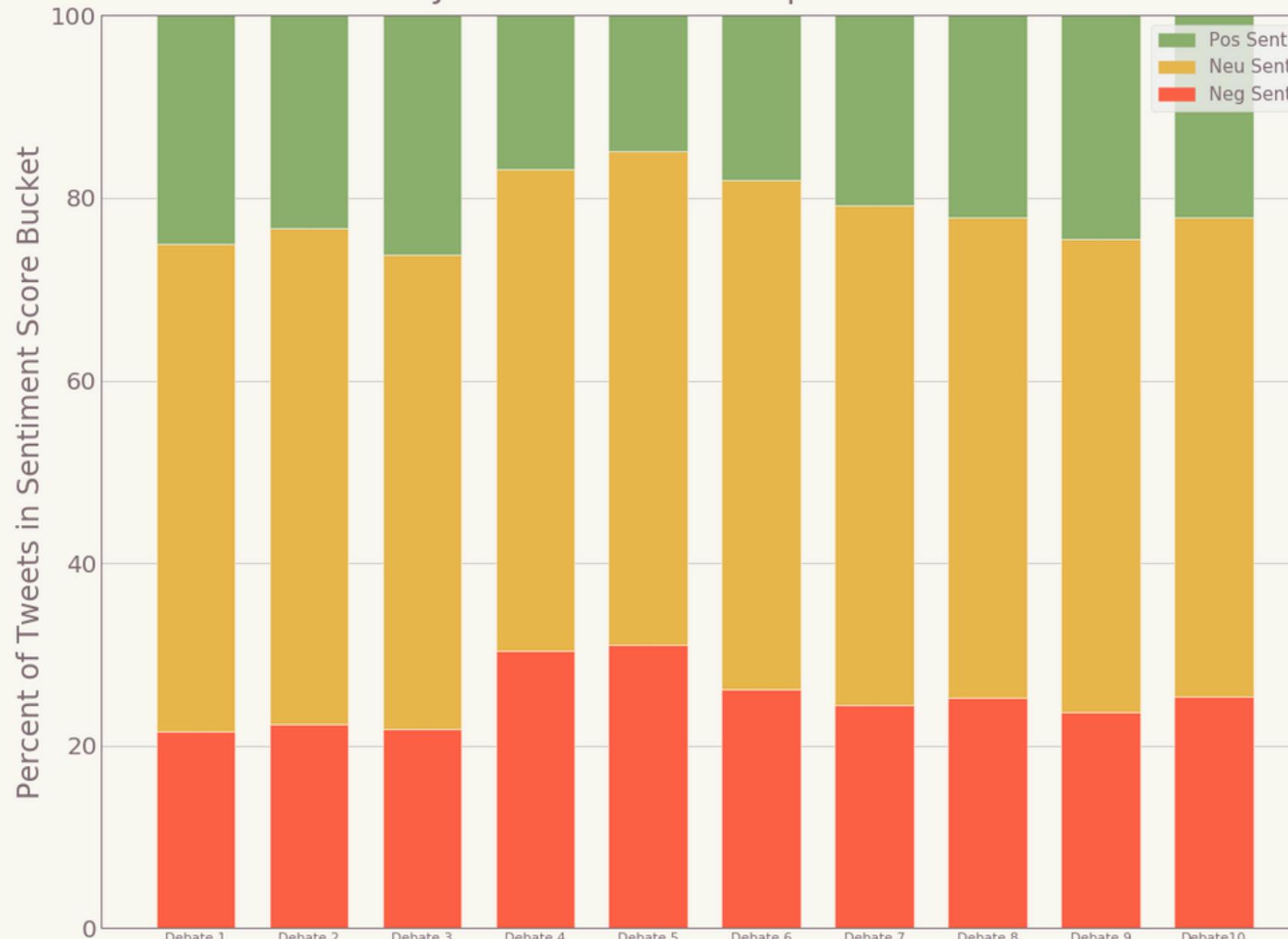


Sentiment Analysis - A Deeper Look

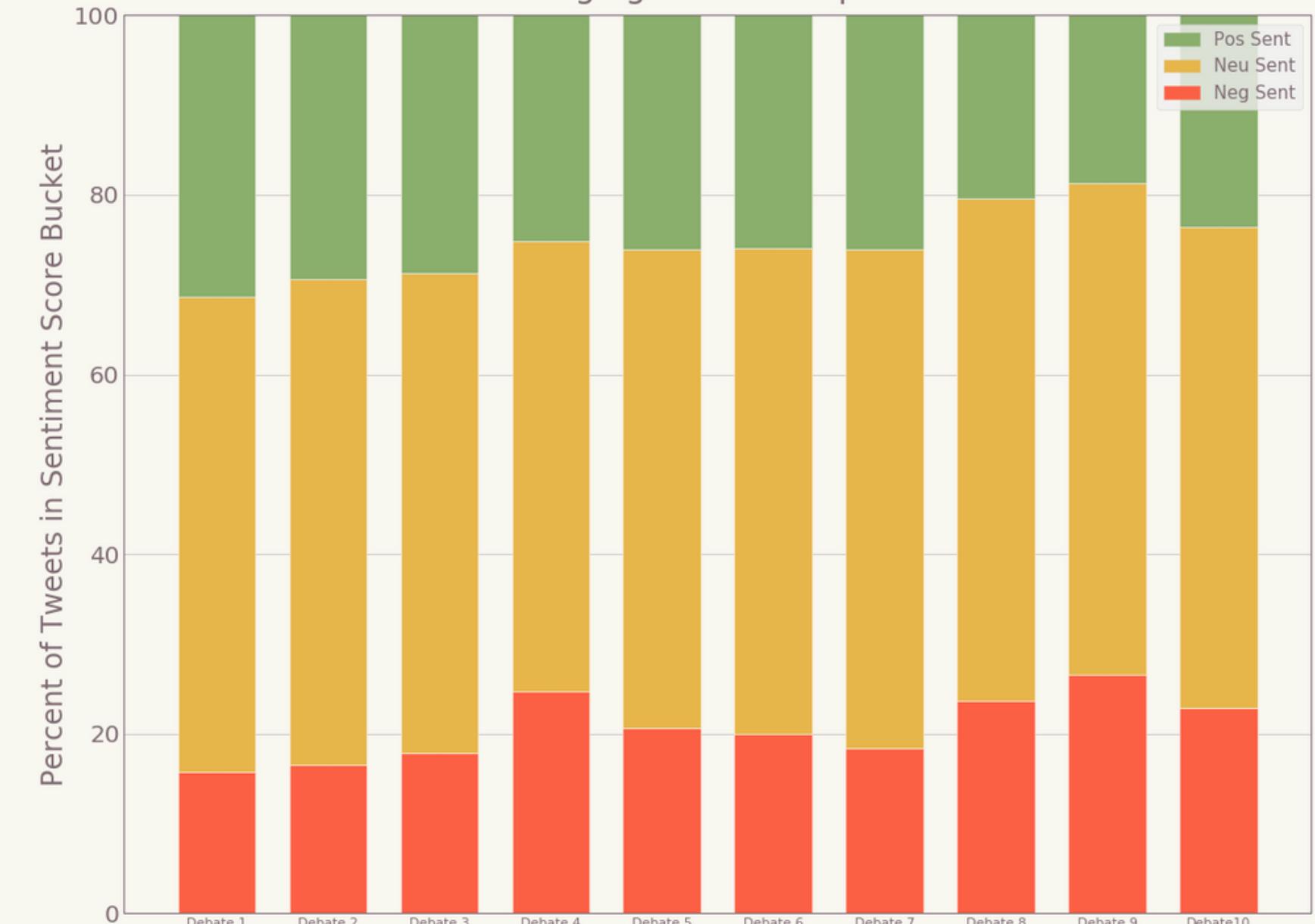




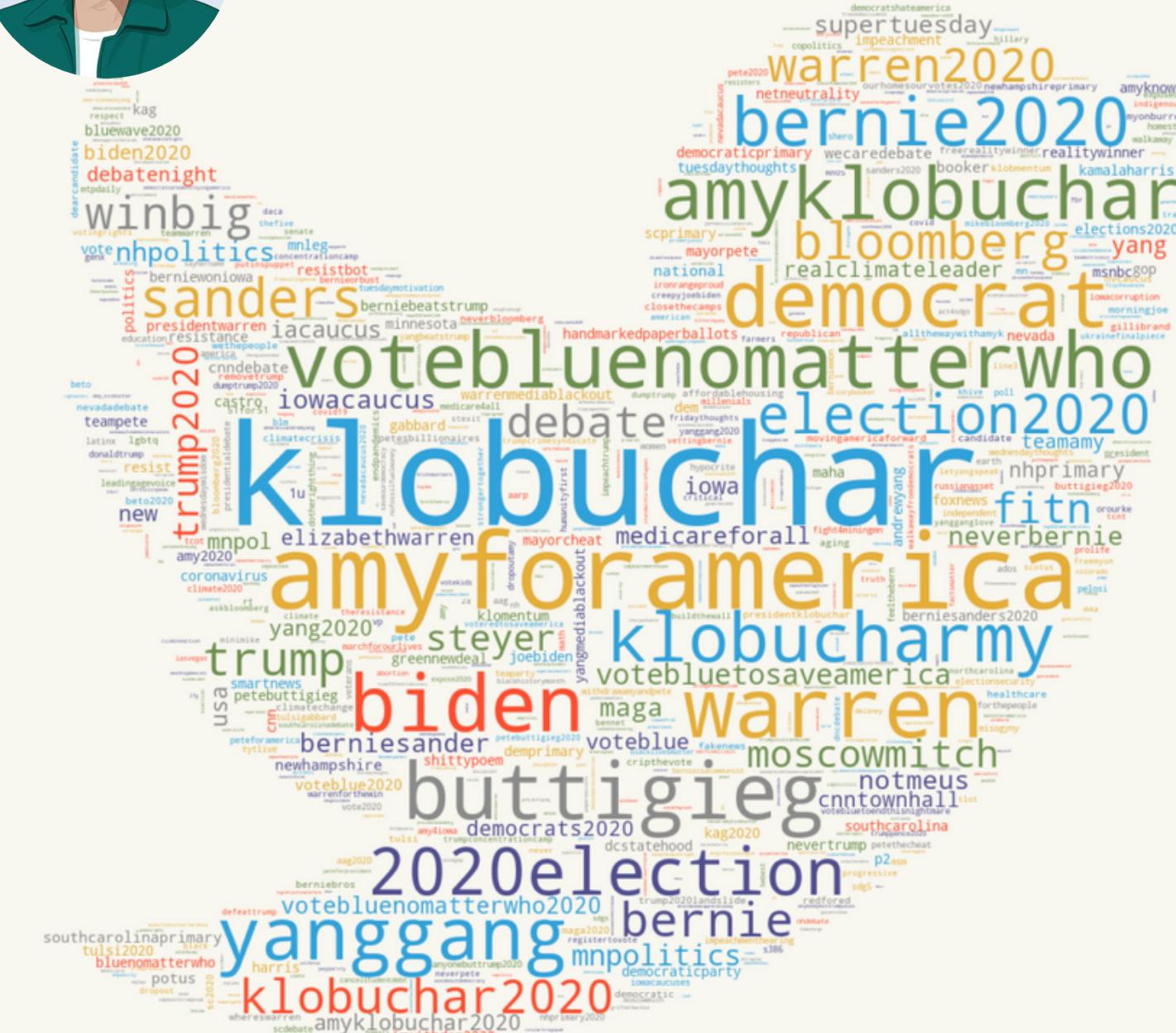
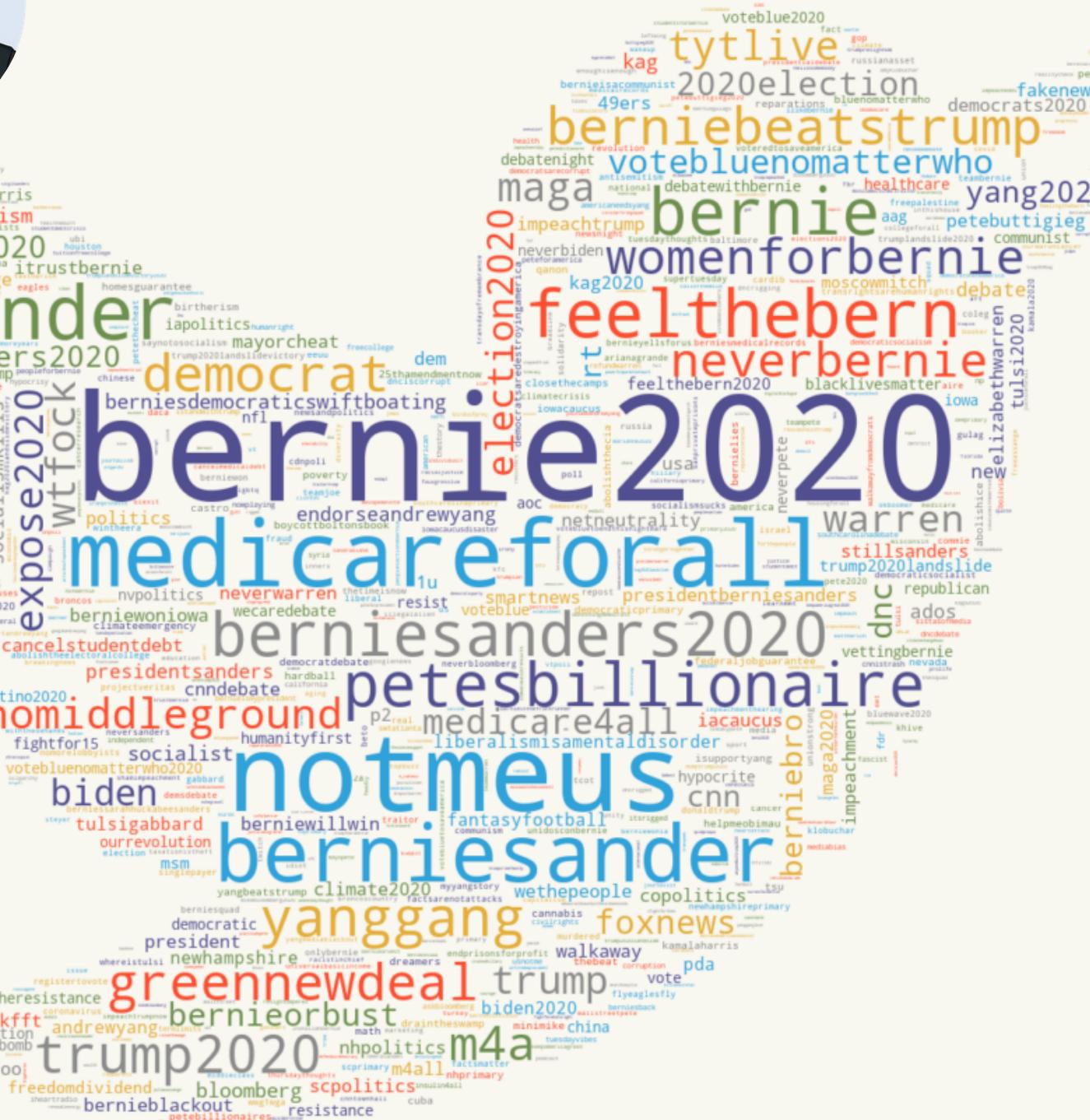
Joe Biden Sentiment per Debate



Pete Buttigieg Sentiment per Debate



Sentiment Analysis - A Deeper Look



#Most Popular Hashtags For Candidates

Background - Bots

WHAT IS A SPAMBOT

- A bot is a piece of software that completes automated tasks over the internet
- Good Bots disclose their identity to the web servers they access while Bad Bots do not

1/2

OF ALL INTERNET
TRAFFIC IS PERFORMED
BY BOTS

56%

OF BOTS ARE
CATEGORIZED AS
BAD BOTS

~15%

OF TWITTER ACCOUNTS
ARE ESTIMATED TO BE
BOTS

Studies show in the months leading up to the 2016 election 1/5 of all tweets that were election related came from a legion of bot accounts

When working together in large cluster, bots have the ability to push narratives that could be false or misleading

DARPA (The Defense Advanced Research Projects Agency) wants to identify and remove bots whose goal aims to influence others and protect information exchange on the internet.

Spambot Relevancy - Why Should We Care?



"Lawmakers Are Warned That Russia Is Meddling to Re-elect Trump"

New York Times (Feb 20, 2020)

"Russia Is Said to Be Interfering to Aid Sanders in Democratic Primaries"

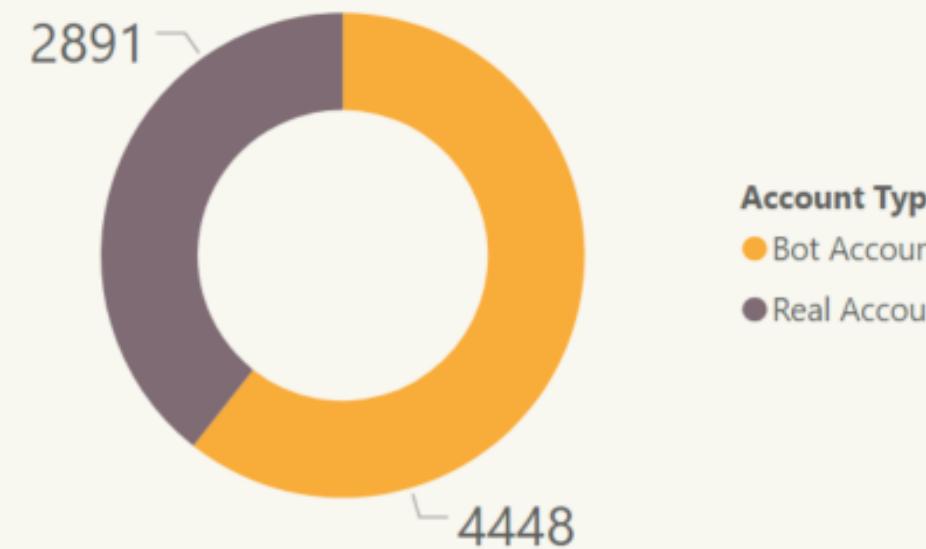
New York Times (Feb 22, 2020)



Spambot Classification

Proof of Concept

Number of Bot Accounts vs Real
Accounts in the Data Set



DATA

Data for training the model came from hand labeled bot and human accounts from the website The Bot Repository. This data consisted of 4448 bot accounts and 2891 human accounts.

LOGISTIC REGRESSION MODEL

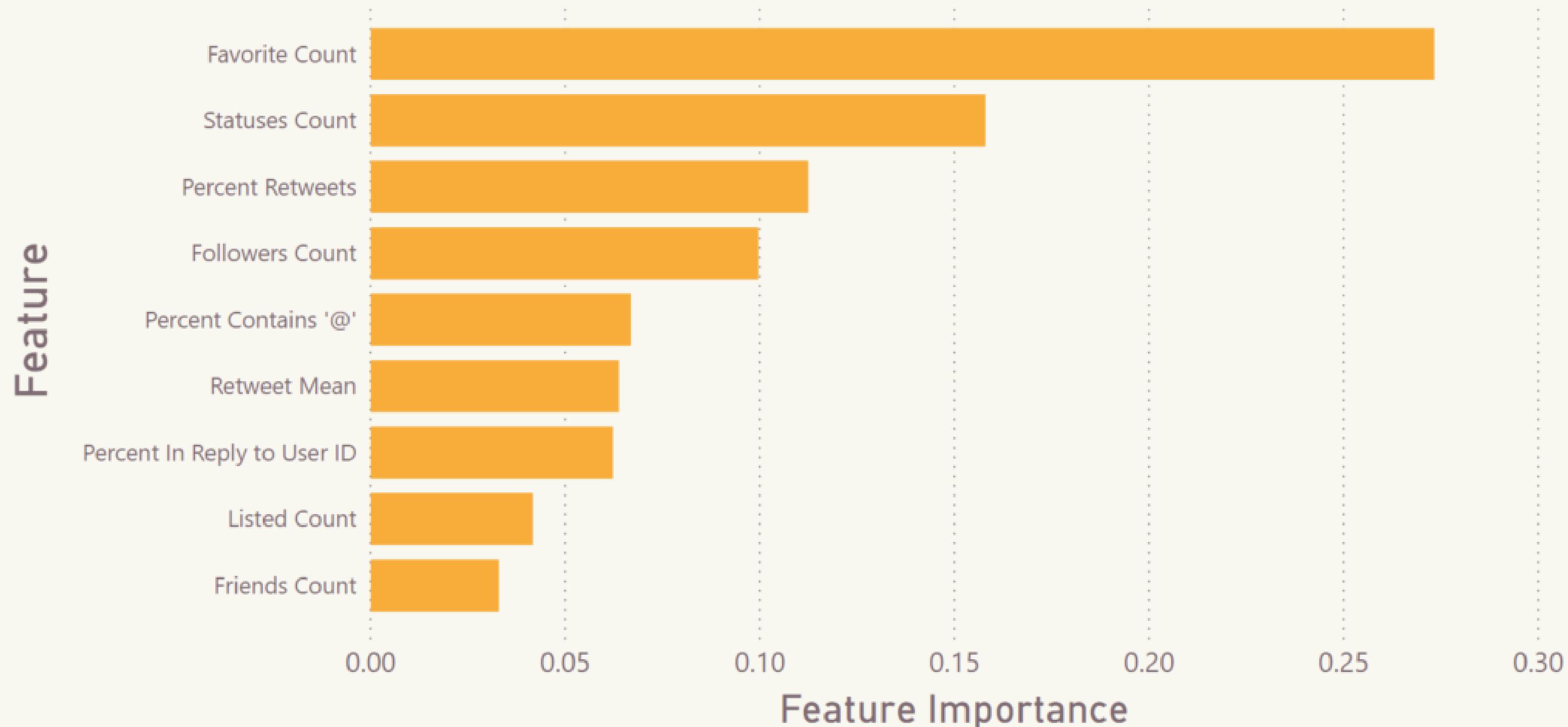
Used a logistic regression machine learning model with user features. Additionally aggregated tweet data to engineer more features

RESULTS

This model is accurate on this training data 98 % of the time in identifying correctly human vs bot accounts

Due to the vast amount of account types on twitter, this model serves more as a proof of concept

Most Important Tweet Features



Spambot Classification - Feature Importance

Bot Classifier Test Case

SCENARIO

For the proof of concept problem the test case has been simplified to **20,000** users from **Debate 9**

RESULTS

The results from this test case produced an estimated **36** bots from the **20,000** users

This number appears to be low, however there are a few reasons this could be.

- 1.** Twitter may already have filtered out bots considering I was grabbing historical tweets
- 2.** The data in the analysis was not representative of the greater Twitter population
- 3.** The data in the train was old

VERIFICATION

Of the **36** accounts identified as bots:
28 appeared to be bots
4 were likely human accounts
4 were difficult to tell

There is also a website called Botometer, and I was able to compare my results to theirs and we agreed on accounts being bots **70 %** of the time

Next Steps

NEURAL NET

Add layers to sentiment Neural Net, using labeled tweet data to see the effect on sentiment scores, and see if we can boost more confidence in the Sentiment Analysis

DATA

More data for both debate night tweets and for training the bot classifier would be beneficial to the analysis

BOT ANALYSIS

Identify more bots (with more data) and do analysis to identify if there are specific candidates that are being favored by bots

Questions?

CONTACT INFORMATION

Sam Videlock

svideloc@gmail.com

<https://www.linkedin.com/in/samuelvidelock/>

