Digital Vitriol in the 2020 US Presidential Election

Shawn Chao, Sam Chinnery, Frank Gonzalez, Nyle Sykes, Evan Vogelbaum

17.835 Machine Learning and Data Science in Politics

MIT POLITICAL SCIENCE

Abstract

Are supporters of Bernie Sanders more vitriolic online than supporters of other presidential candidates? We believe this may be the case for two reasons:

- **Cognitive dissonance:** Individuals who voted for Sanders in 2016 may feel the need to validate their past "investment" in voting
- **Far-from-center ideology:** Sanders is on the left extreme of the ideological spectrum

Previous work has shown that social media can be reliable in determining political affiliation. We seek to determine individuals' political ideologies using Twitter posts and to quantify their vitriol using sentiment analysis.

Data

TWINT Project
Open-source Twitter Intelligence
https://twint.io

pandas

We collected Twitter data using the TWINT API with simple integration to the Python *pandas* package.

Data Collected from TWINT:

- 14 usernames of political candidates
- Over 5,000 usernames of followers per candidate
- Over 165,000 total tweets from those followers
- 5,000 usernames of who these followers followed
  - Find between-candidate overlap/affiliation

Sentiment Analysis Methods

Naïve Bayes Models

- Scale Priors by probability of class given input
- Assumes Independence of Features
- **NLTK:** Twitter Data, Stopwords, Noise Reduction
- **TextBlob**: "Vanilla" Model, IMDB Data

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

VADER

- Specialized to Social Media
- Defines heuristics from many datasets
- Accounts for Sentiment Laden Emojis
  - 😍 = smiling face with heart-eyes
  - 😱 = face screaming in fear
- Understands Trigram Sentiment Switches
  - "The food here isn't really all that great

Descriptive Analysis

- Collected top hashtags per party (GOP omitted below)
- Sanders supporters used less aggressive hashtags than supporters of other Democratic candidates

Rank	Sanders (*N* = 57578)		Other Democrats (*N* = 62539)	
	Hashtag	Count	Hashtag	Count
1	#trump	44	#trump	71
2	#nitishfailedcm	36	#trumpliesamericansdie	47
3	#love	33	#yelp	43
4	#runkeeper	30	#biden2020	39
5	#togetherathome	28	#trumpvirus	38
6	#mentalhealth	26	#voicesavetj	36
7	#bitcoin	25	#love	35
8	#stayathome	23	#cbcares	34
9	#family	22	#moscowmitch	28
10	#donaldtrump	22	#palghar	28

Affiliation Matrix

- Matrices describe overlap between the sets of supporters of candidate pairs
- The "incumbency effect" is clearly visible: Trump has a significantly larger follower base and thus has considerable overlap with most other candidates

	Klobuchar	Yang	Sanders	Warren	Weld	Biden	De La Fuente	Harris	Bloomberg	Trump	Bennet	Steyer	Gabbard	Walsh
Klobuchar	363													
Yang	142	358												
Sanders	153	135	436											
Warren	177	137	224	386										
Weld	178	162	79	92	587									
Biden	238	182	238	274	306	811								
De La Fuente	19	21	41	30	11	21	183							
Harris	245	200	182	243	358	469	19	745						
Bloomberg	180	180	103	110	384	290	12	380	570					
Trump	200	206	237	184	293	355	41	292	292	643				
Bennet	133	135	50	53	331	214	2	288	325	231	448			
Steyer	88	84	51	61	147	124	11	134	140	102	111	240		
Gabbard	118	144	74	73	175	119	25	161	160	153	141	65	265	
Walsh	68	56	48	50	90	84	8	86	79	64	65	34	45	178

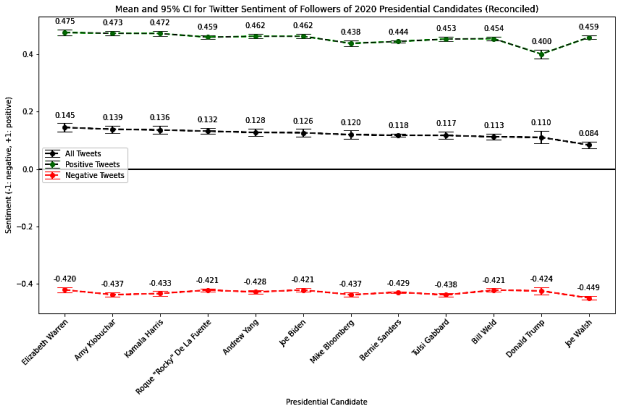
"Affiliation matrix" between candidates: shows the number of users that follow both (row) and (column). Blue cells indicate higher overlap.

	Klobuchar	Yang	Sanders	Warren	Weld	Biden	De La Fuente	Harris	Bloomberg	Trump	Bennet	Steyer	Gabbard	Walsh
Klobuchar	1.00	0.39	0.42	0.49	0.49	0.66	0.05	0.68	0.50	0.55	0.37	0.24	0.33	0.19
Yang	0.40	1.00	0.38	0.38	0.45	0.51	0.06	0.56	0.50	0.58	0.38	0.24	0.40	0.16
Sanders	0.35	0.31	1.00	0.51	0.18	0.55	0.09	0.42	0.24	0.54	0.12	0.12	0.17	0.11
Warren	0.46	0.36	0.58	1.00	0.24	0.71	0.08	0.63	0.29	0.48	0.14	0.16	0.19	0.13
Weld	0.30	0.28	0.14	0.16	1.00	0.52	0.02	0.61	0.65	0.50	0.56	0.25	0.30	0.15
Biden	0.29	0.22	0.29	0.34	0.38	1.00	0.03	0.58	0.36	0.44	0.26	0.15	0.15	0.10
De La Fuente	0.10	0.12	0.22	0.16	0.06	0.12	1.00	0.10	0.07	0.22	0.01	0.06	0.14	0.04
Harris	0.33	0.27	0.24	0.33	0.48	0.63	0.03	1.00	0.51	0.39	0.39	0.18	0.22	0.12
Bloomberg	0.32	0.32	0.18	0.19	0.67	0.51	0.02	0.67	1.00	0.51	0.57	0.25	0.28	0.14
Trump	0.31	0.32	0.37	0.29	0.46	0.55	0.06	0.45	0.45	1.00	0.36	0.16	0.24	0.10
Bennet	0.30	0.30	0.11	0.12	0.74	0.48	0.00	0.64	0.73	0.52	1.00	0.25	0.32	0.15
Steyer	0.37	0.35	0.21	0.25	0.61	0.52	0.05	0.56	0.58	0.43	0.46	1.00	0.27	0.14
Gabbard	0.45	0.54	0.28	0.28	0.66	0.45	0.09	0.61	0.60	0.58	0.53	0.25	1.00	0.17
Walsh	0.38	0.32	0.27	0.28	0.51	0.47	0.05	0.48	0.44	0.36	0.37	0.19	0.25	1.00

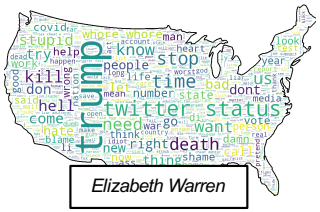
"Agreement matrix" between candidates: shows the probability that a follower of (row) also follows (column). Red cells indicate higher overlap.

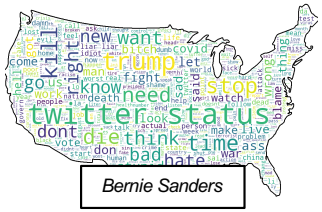
Sentiment Analysis Results

- Average sentiment for all tweets, negative tweets, and positive tweets by followers of candidates shown below
- Combination of Vader and TextBlob results



- While Sanders' supporters are more negative than other Democratic candidates, we cannot attribute this to material differences in the negative vocabulary used





Limitations and Future Work

- **Data collection:** Tweets were collected following the outbreak of COVID-19. Future work would limit the data to tweets posted prior to the outbreak
- **Overlap between candidates:** Future work would consider alternate data sources (e.g. Reddit) where political affiliation may be more easily distinguishable
- **Sample bias:** The ideological distribution of Twitter users is not uniform. Future work would cross-reference election contribution databases to determine ideology

References

1. Bond, Robert, and Solomon Messing. "Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook." *American Political Science Review* 109.1 (2015): 62-78.

May 10, 2020