

# Political Ideology Detection

## Gradient Dissenters

Maharshi Pandya, Tanvi Tyagi, Varun Tanna



THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL

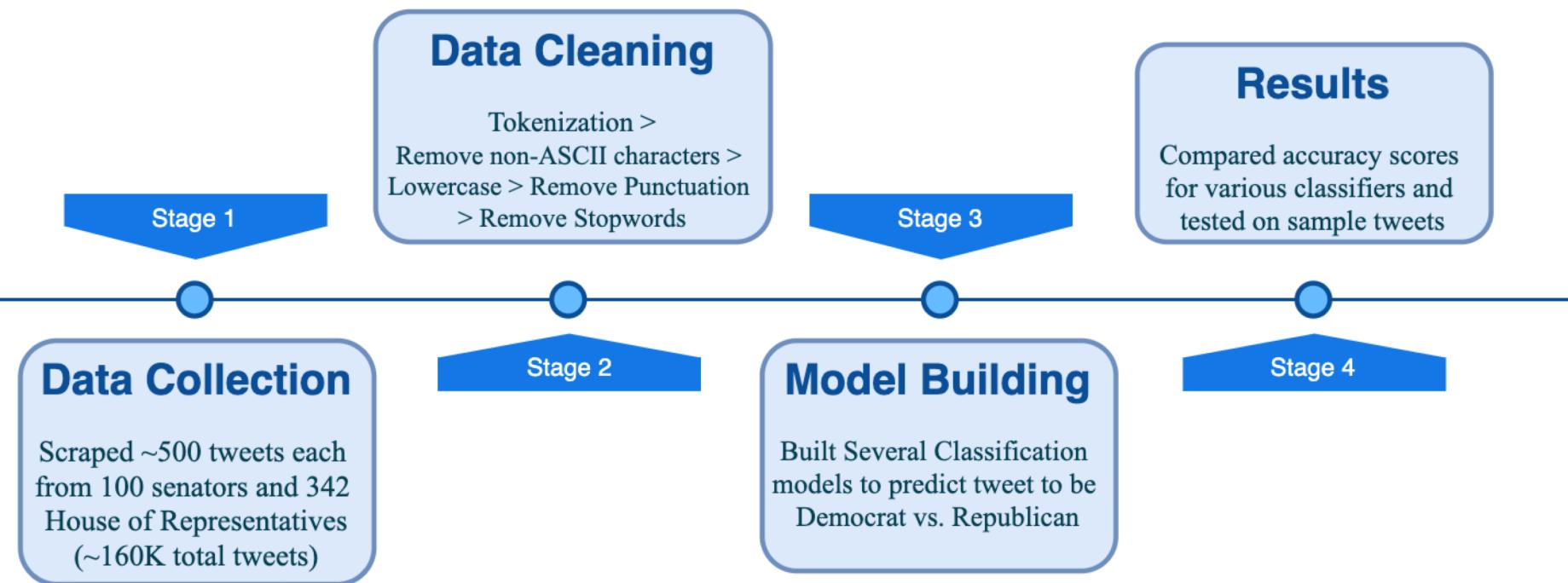
# Research Questions

---

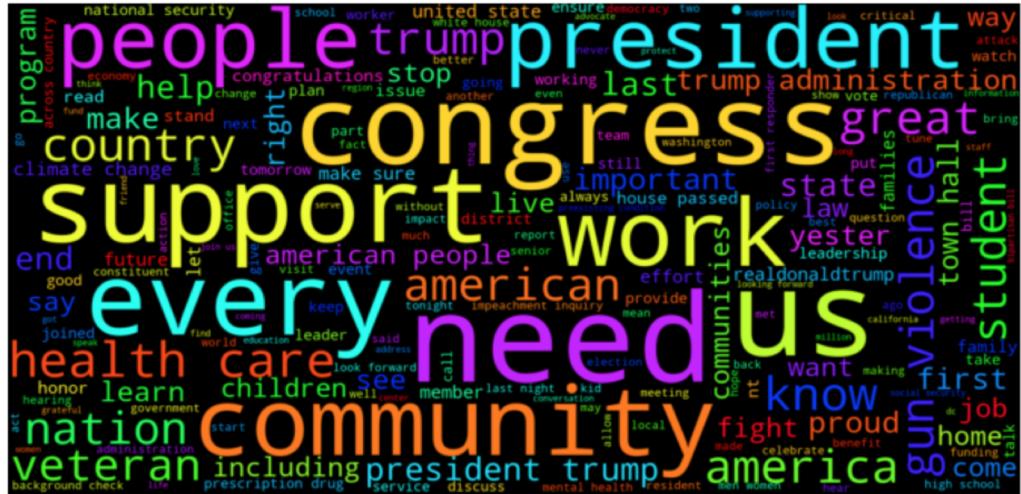
1. Identify individual's political party inclination (Democrat vs. Republican) based on their tweets.
2. Use semantic information within tweets to detect trends on candidate favorability before and after democratic debates.

# Research Question 1

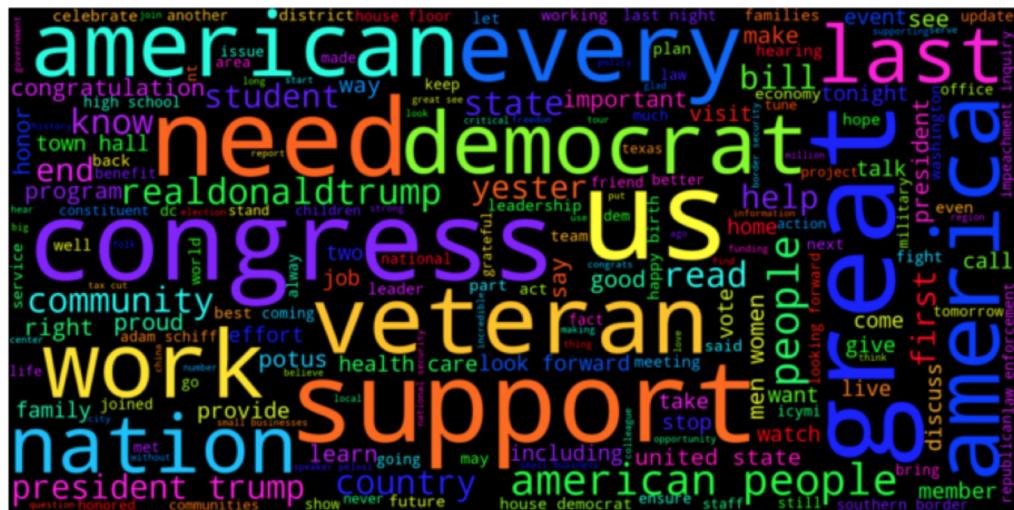
---



# Word Clouds



## Democrats



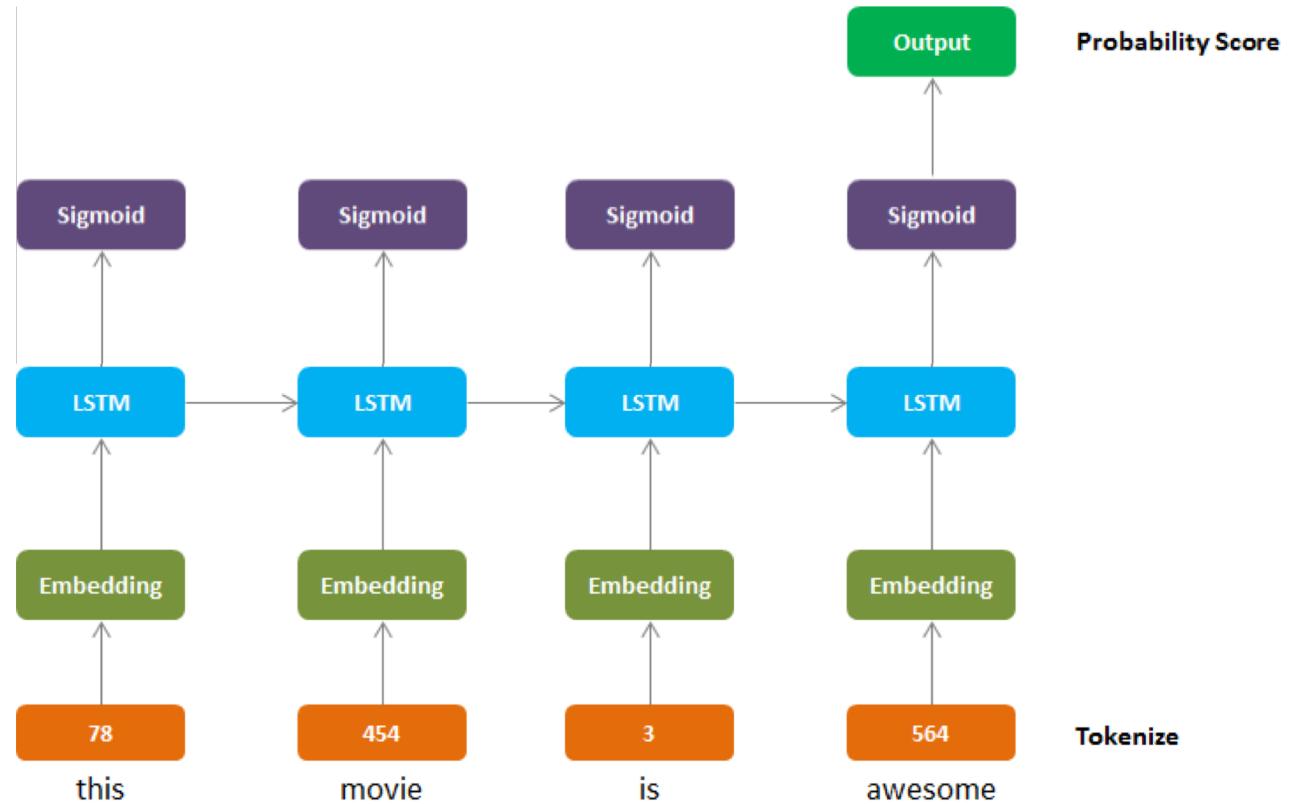
## Republicans

# Research Question 1 - Model Comparison

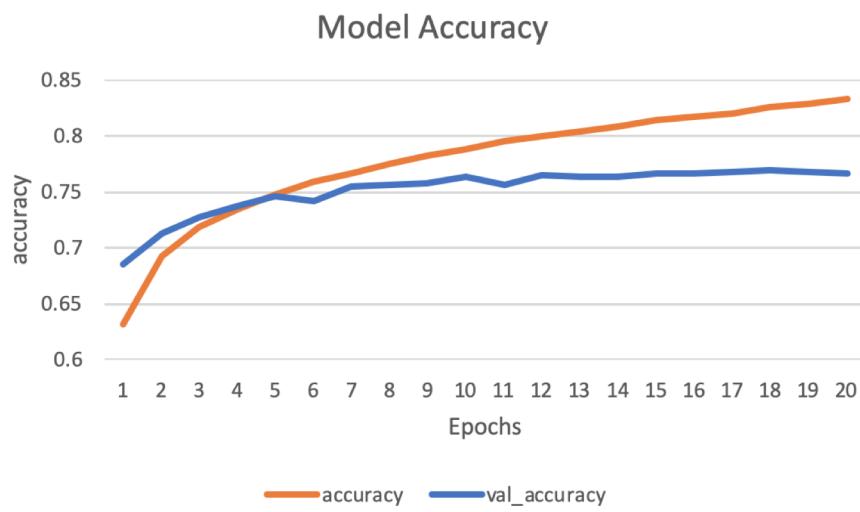
---

Model	Accuracy Scores (Senate)	Accuracy Scores (House of Representative)
Logistic Regression	0.35	0.81
Linear SVM	0.34	0.82
Multinomial Naive Bayes	0.37	0.82
Stochastic Gradient Descent	0.53	0.79

# Research Question 1 - RNN



# Research Question 1 - RNN



# Guess the Affiliation

---

Donald J. Trump @realDonaldTrump

The Republican Party has NEVER been so united! This Impeachment Scam is just a continuation of the 3 year Witch Hunt, but it is only bringing us even closer together!

12:56 PM · Dec 2, 2019 · Twitter for iPhone

---

**11.2K** Retweets   **45.4K** Likes

array([[0.7984898]], dtype=float32)

# Guess the Affiliation



Bernie Sanders 

@BernieSanders

Prisons should not be for profit.

Health insurance should not be for profit.

We need to move in the direction of protecting human lives and dignity, not defending the limitless greed of massive corporations.

4:05 PM · Dec 2, 2019 · Twitter for iPhone

**1.7K** Retweets    **7.9K** Likes

array([[0.02199409]], dtype=float32)

# Guess the Affiliation

---



Andrew Yang  

@AndrewYang

If I was America's doctor, my prescription would be money in people's hands.

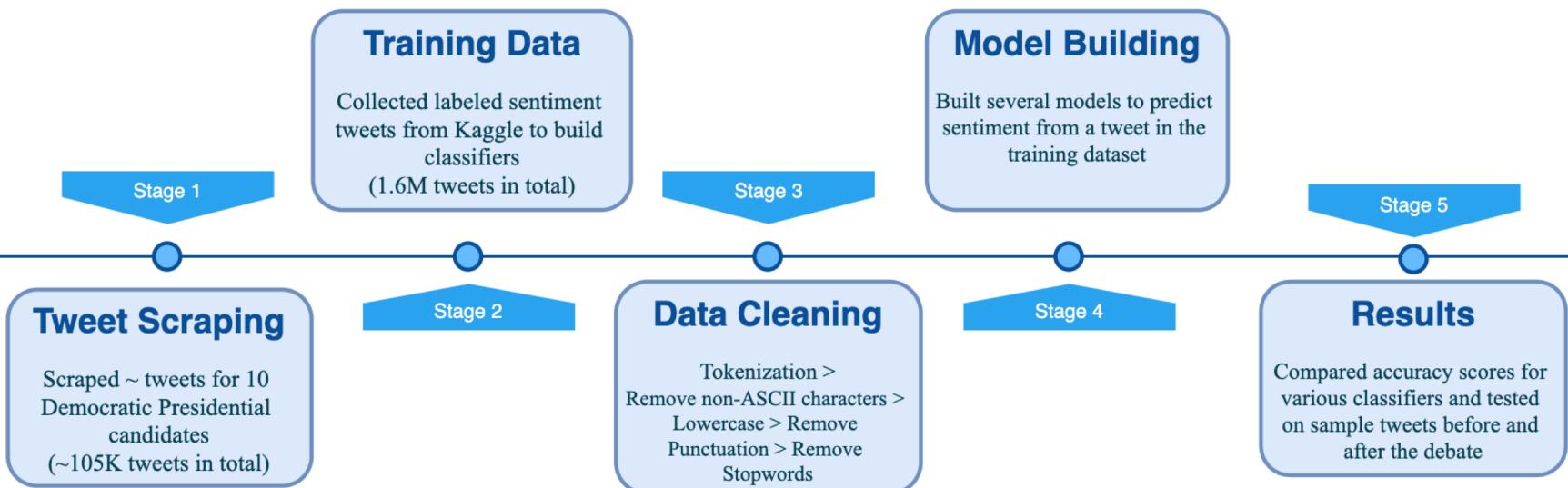
9:44 AM · Nov 25, 2019 · Twitter for iPhone

---

1.4K Retweets 13K Likes

array([[0.07099717]], dtype=float32)

# Research Question 2



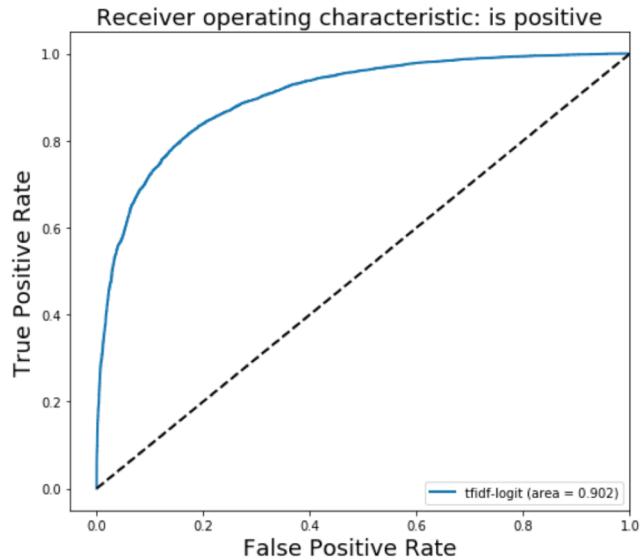
# Research Question 2 - Results

---

Model	Accuracy Scores
Logistic Regression	82.30
Linear SVM	81.96
Linear SVM with L1- Feature Selection	81.95
Multinomial Naive Bayes	80.33
Bernoulli Naive Bayes	79.34
Ridge Classifier	81.84
ADA Boost	60.76
Perceptron	75.69
Passive-Aggressive	78.82
Nearest Centroid	72.14

# Research Question 2 - Results

## Logistic Regression with Tf Idf Vectorizer



Train set has total 1547039 entries with 49.93% negative, 50.07% positive

Validation set has total 15786 entries with 50.75% negative, 49.25% positive

Test set has total 15787 entries with 50.22% negative, 49.78% positive

---

	negative	positive
precision:	[0.82698251	0.81513863]
recall:	[0.81120194	0.83065884]
f1 score:	[0.81901622	0.82282555]

---

	negative	positive
precision:	[0.82944874	0.81730656]
recall:	[0.81335811	0.83312663]
f1 score:	[0.82132462	0.82514077]

---

	negative	positive
precision:	[0.82814961	0.81473749]
recall:	[0.81034581	0.83221438]
f1 score:	[0.81915098	0.82338321]

---

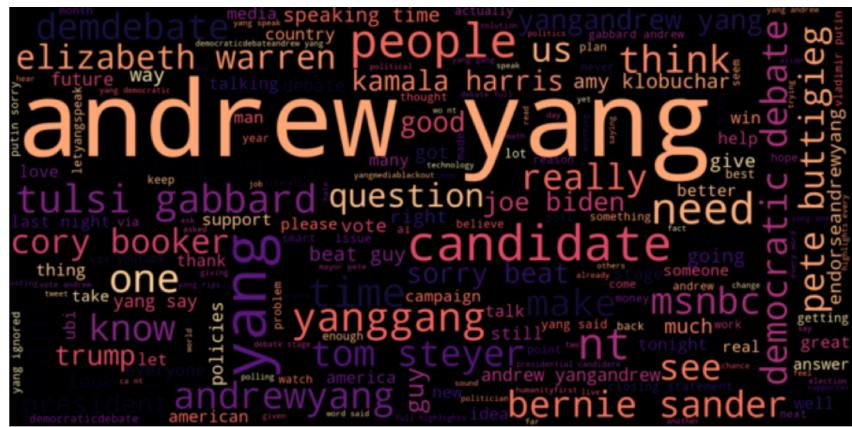
	negative	positive
precision:	[0.82746305	0.8140137 ]
recall:	[0.80958481	0.83156263]
f1 score:	[0.81842631	0.82269459]

---

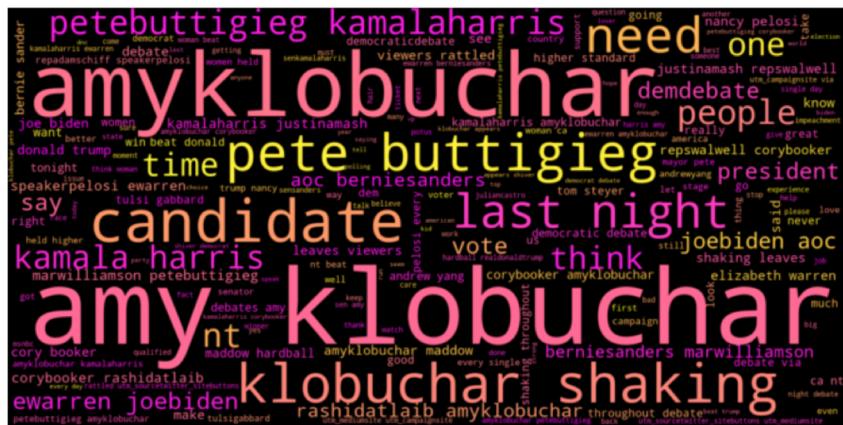
	negative	positive
precision:	[0.82906586	0.81674548]
recall:	[0.81273028	0.83280286]
f1 score:	[0.8208168	0.82469602]

---

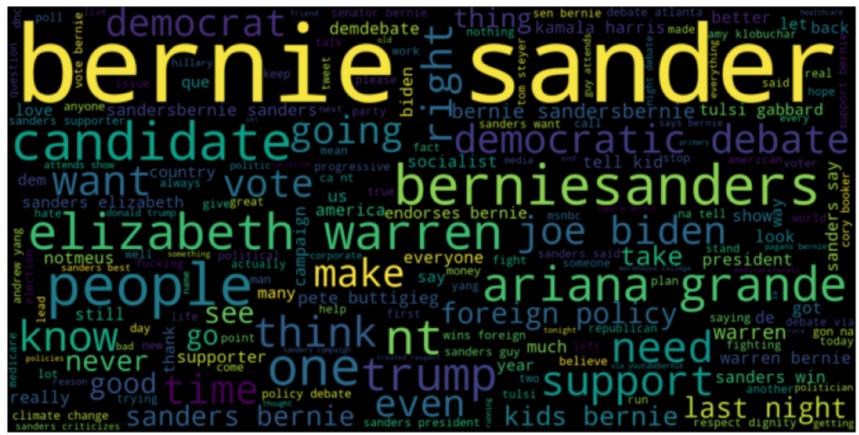
	accuracy	precision	recall	f1 score
accuracy:	82.18% (+/- 0.11%)			
precision:		82.19% (+/- 0.10%)		
recall:			82.18% (+/- 0.11%)	
f1 score:				82.17% (+/- 0.11%)



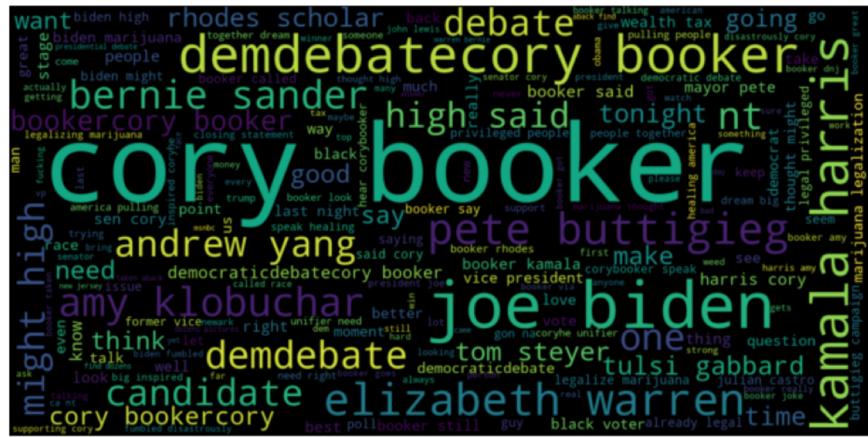
	Before Debate	After Debate
<b>Andrew Yang</b>		
Number of Tweets Collected	3928	6882
Number of Tweets Classified as Positive	3031	5088
Number of Tweets Classified as Negative	897	1794
Sentiment Score	0.7716	0.7393



Amy Klobuchar	Before Debate	After Debate
Number of Tweets Collected	1376	3249
Number of Tweets Classified as Positive	1052	2455
Number of Tweets Classified as Negative	324	794
Sentiment Score	0.7645	0.7556

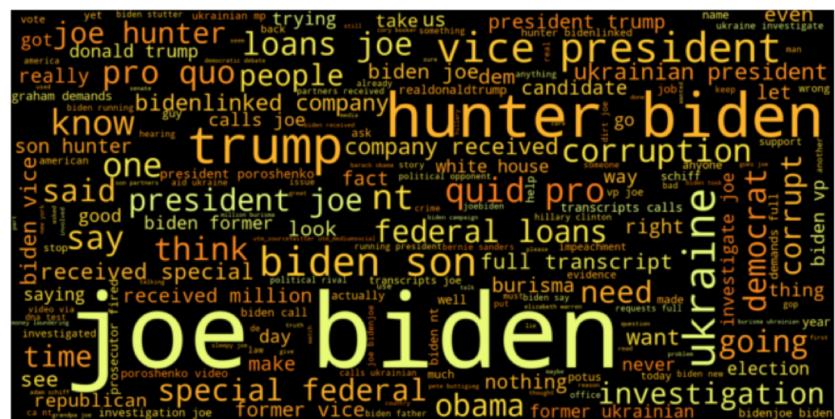
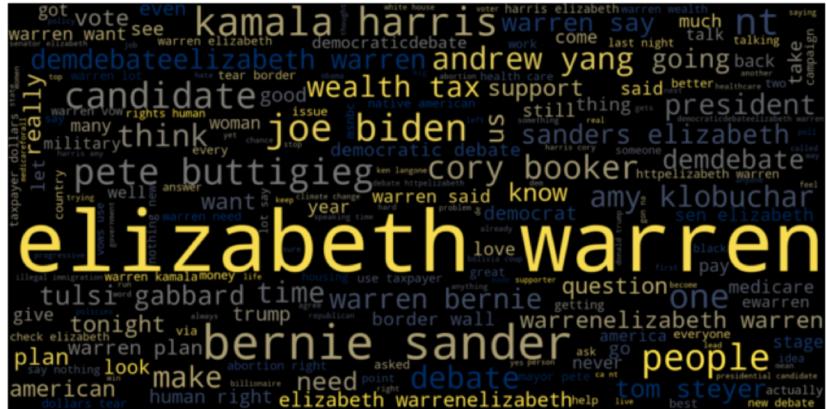


	Before Debate	After Debate
<b>Bernie Sander</b>		
Number of Tweets Collected	4201	8324
Number of Tweets Classified as Positive	1395	3102
Number of Tweets Classified as Negative	2806	5222
Sentiment Score	0.3320	0.3727



	Before Debate	After Debate
<b>Cory Booker</b>		
Number of Tweets Collected	1252	6800
Number of Tweets Classified as Positive	1034	5390
Number of Tweets Classified as Negative	218	1410
Sentiment Score	0.8259	0.7926

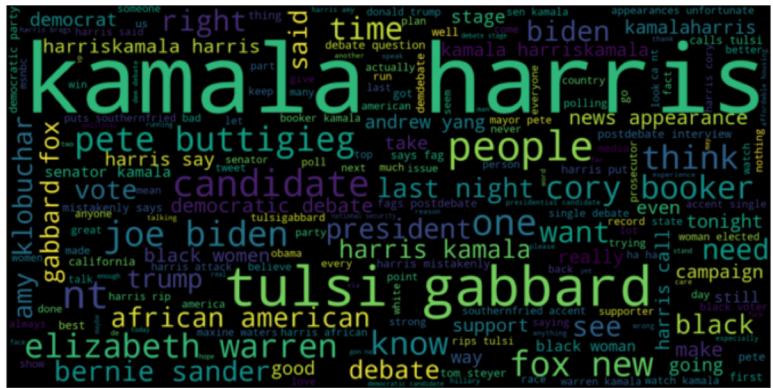
# Research Question 2



	Before Debate	After Debate
<b>Elizabeth Warren</b>		
Number of Tweets Collected	5544	7024
Number of Tweets Classified as Positive	4216	5214
Number of Tweets Classified as Negative	1328	1810
Sentiment Score	0.7605	0.7423

	Before Debate	After Debate
<b>Joe Biden</b>		
Number of Tweets Collected	8860	11176
Number of Tweets Classified as Positive	5972	7711
Number of Tweets Classified as Negative	2888	3465
Sentiment Score	0.6740	0.6870

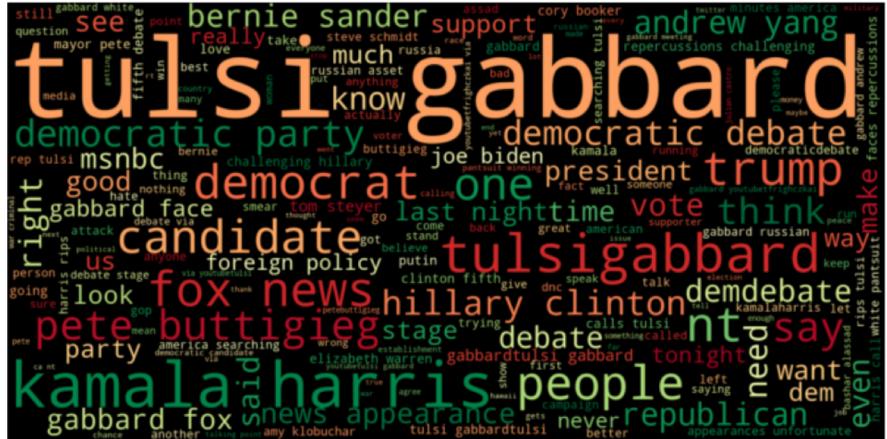
## Research Question 2



Kamala Harris	Before Debate	After Debate
Number of Tweets Collected	2520	6828
Number of Tweets Classified as Positive	1986	5198
Number of Tweets Classified as Negative	534	1630
Sentiment Score	0.7880	0.7613

	Before Debate	After Debate
<b>Pete Buttigieg</b>		
Number of Tweets Collected	5961	6619
Number of Tweets Classified as Positive	4798	5144
Number of Tweets Classified as Negative	1163	1475
Sentiment Score	0.8049	0.7771

# Research Question 2

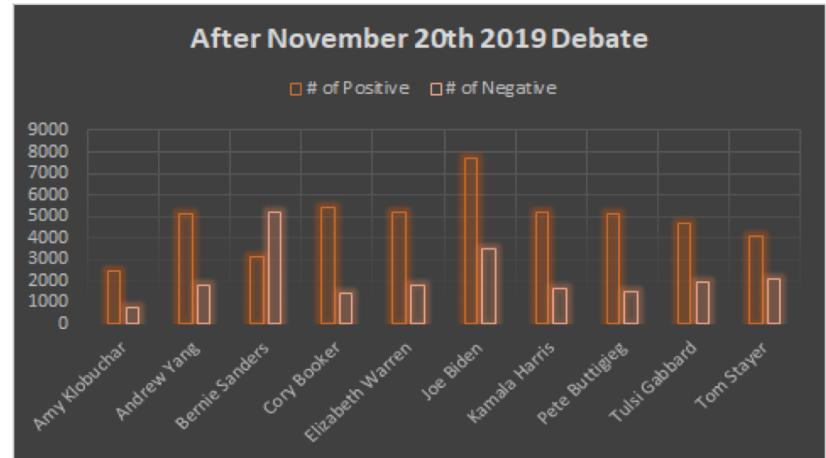
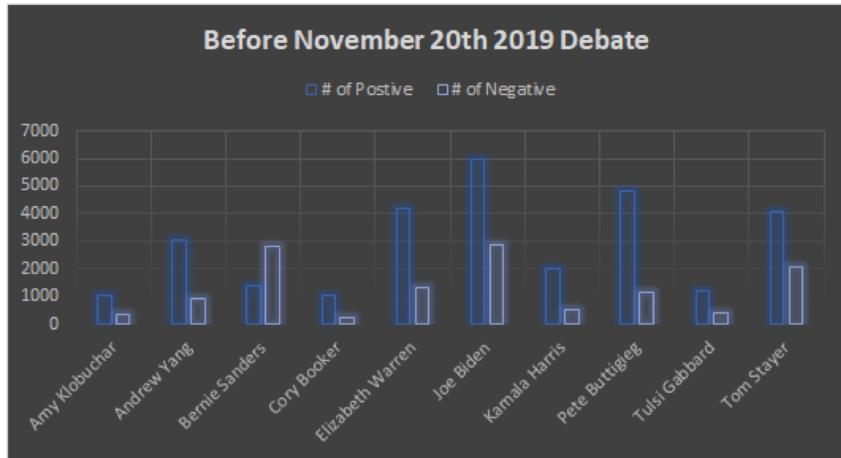


	Before Debate	After Debate
<b>Tulsi Gabbard</b>		
Number of Tweets Collected	1635	6617
Number of Tweets Classified as Positive	1213	4684
Number of Tweets Classified as Negative	422	1933
Sentiment Score	0.7419	0.7079



	Before Debate	After Debate
<b>Tom Steyer</b>		
Number of Tweets Collected	537	6156
Number of Tweets Classified as Positive	396	4084
Number of Tweets Classified as Negative	141	2072
Sentiment Score	0.7374	0.6634

# Research Question 2



# Future Directions

---

- We intend to use party classifier to get sample tweets before 2020 Presidential Elections to see state-wise party popularity.
- Incorporating additional parameters (Likes, Retweets, Replies)
- Usage



THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL