

# Predicting 2020 Democratic Candidates Based on Question Responses

Aashish Ghimire & Chase Mortensen

# Who is most likely to have said it?

"Then let us have - the women have an opportunity to speak. The Bloomberg corporations and Mayor Bloomberg himself have been accused of discrimination. They are bound by non-disclosures so that they cannot speak. If he says there is nothing to hide here"

One purpose of this project was to train a model that would take a response at one of the DNC debates and predict which candidate gave the response.

- Removing common words like “the,” “and,” and even “medicare” improved the accuracy
- Short responses, like “Doing nothing is what will happen-” weren’t especially interesting - removing short responses improved the accuracy

# Who is most likely to have said it?

"Then let us have - the women have an opportunity to speak. The Bloomberg corporations and Mayor Bloomberg himself have been accused of discrimination. They are bound by non-disclosures so that they cannot speak. If he says there is nothing to hide here"

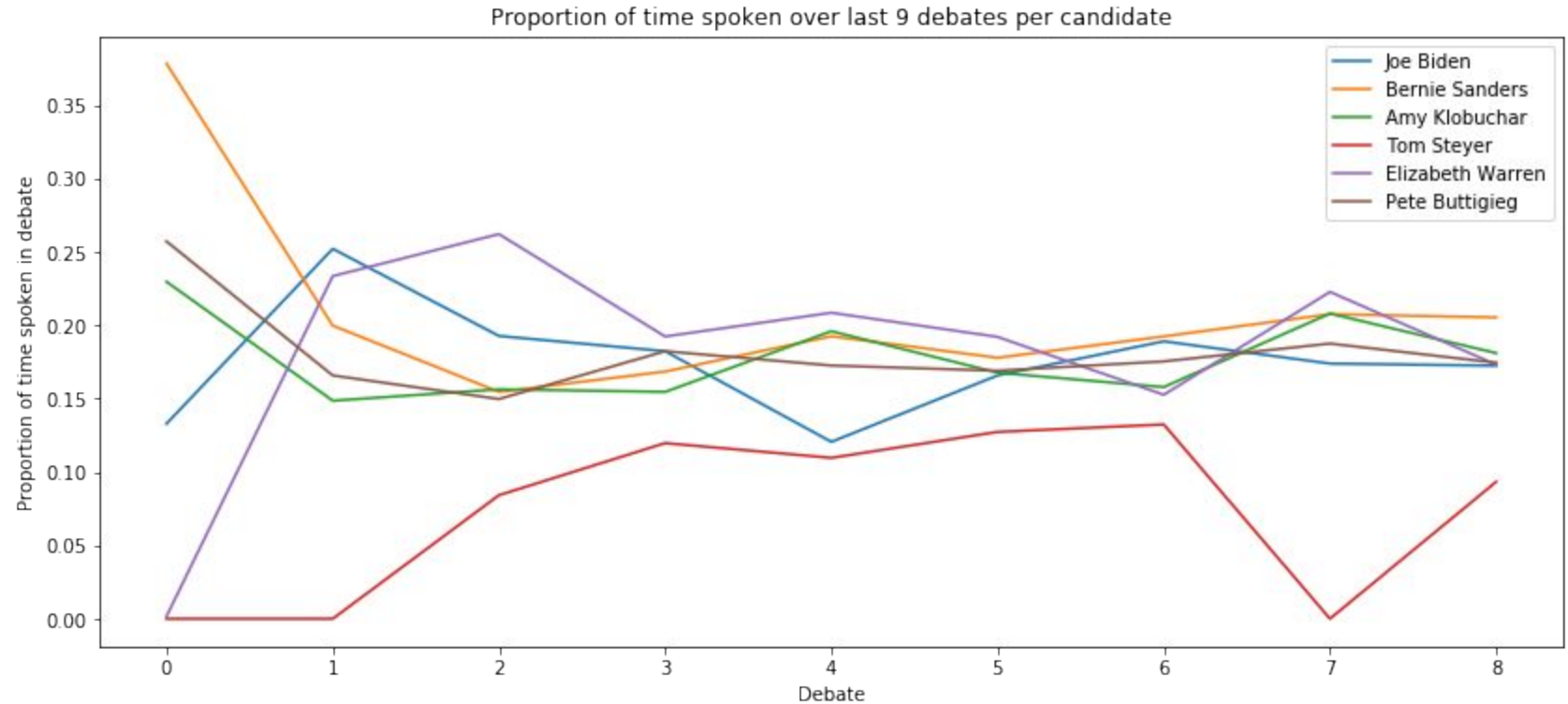
## Approach:

Take a response like the one above and predict candidate based on each word.  
Predict most commonly occurring candidate for the response.

# Which debates were atypical?

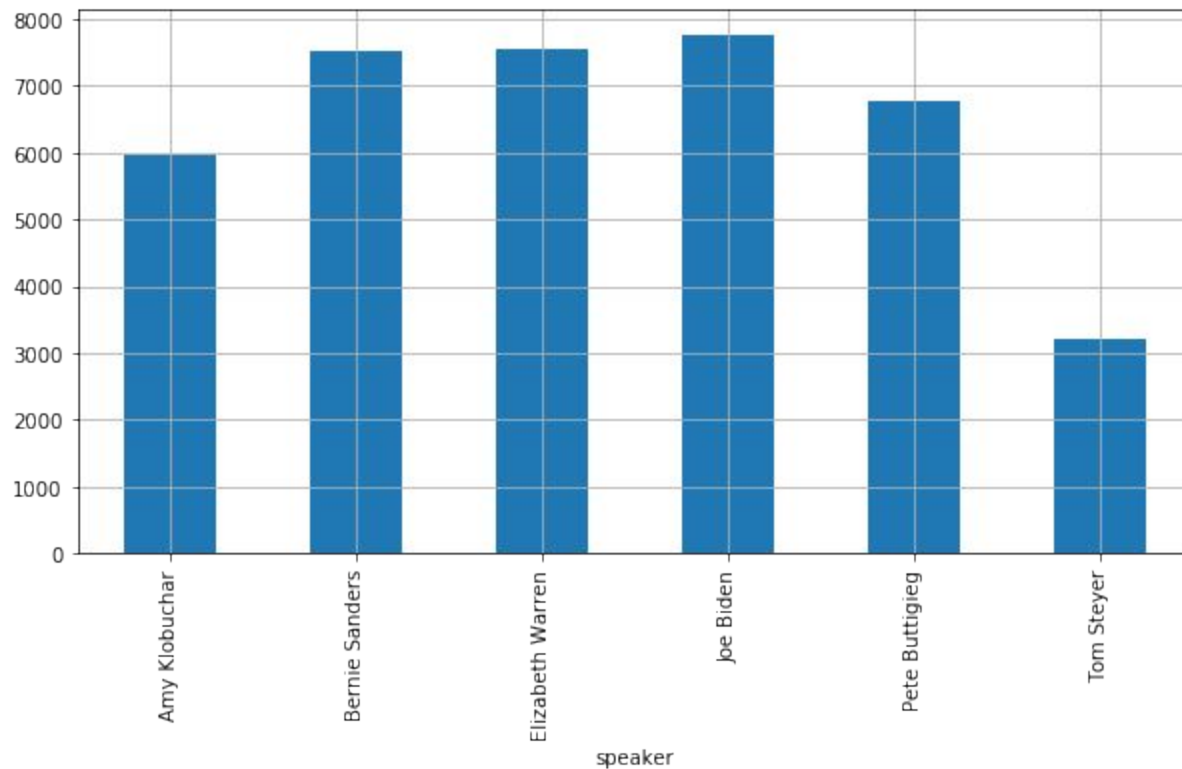
There have been ~12 DNC debates (two were divided into two nights) - which debates were most like the others? Which were most different?

- We are defining atypical debates as ones where it is most difficult to predict which candidate gave which response



Our Naive Bayes Classifier tends to predict Joe Biden and not predict Tom Steyer - it's affected by the proportion of total words spoken.

# Total speaking times over 9 debates\*



# Results

It works! How well? Depends on...

- Do you include Bloomberg?
- How long are acceptable responses?

# Results

Accuracy per  
debate after  
discarding answers  
less than 30 words  
in length.

Debate	Accuracy
2-25-2020	0.851
2-19-2020	0.920
2-7-2020 (150)	0.840
1-14-2020	0.775
12-29-2019	0.840
11-20-2019	0.865
10-15-2019	0.915
9-12-2019 (116)	0.966
7-31-2019 (41*)	1.0
7-30-2019	0.942
6-27-2019	0.921
6-26-2019 (34)	0.853
<b>ALL</b>	<b>0.879</b>
<b>ALL (without discarding)</b>	<b>0.716</b>



## Alternative Approach - What about using a phrase?

- Relatively small dataset means not many repeated phrases.
- Does really good in avoiding false positive because of ambiguity or common words.
- Cost a lot in accuracy because of not existing in training dataset
- We tied using 2 words and three words phrases:

len(phrase) \ test	by phrase	by answer
two words	0.63	0.26
three words	0.83	0.24

**Bonus:** What about other past presidential candidates?

# Who is most likely to have said it?

"Well, thank you, Lester, and thanks to Hofstra for hosting us. The central question in this election is really what kind of country we want to be and what kind of future we'll build together. Today is my granddaughter's second birthday, so I think about this a lot. First, we have to build an economy that works for everyone, not just those at the top. That means we need new jobs, good jobs, with rising incomes. I want us to invest in you. I want us to invest in your future. That means jobs in infrastructure, in advanced manufacturing, innovation and technology, clean, renewable energy, and small business, because most of the new jobs will come from small business. We also have to make the economy fairer. That starts with raising the national minimum wage and also guarantee, finally, equal pay for women's work."

Hillary Clinton said this in a 2016 debate. If this response had been given in a 2020 debate, who is most likely to have said it?

# Elizabeth Warren

"Well, thank you, Lester, and thanks to Hofstra for hosting us. The central question in this election is really what kind of country we want to be and what kind of future we'll build together. Today is my granddaughter's second birthday, so I think about this a lot. First, we have to build an economy that works for everyone, not just those at the top. That means we need new jobs, good jobs, with rising incomes. I want us to invest in you. I want us to invest in your future. That means jobs in infrastructure, in advanced manufacturing, innovation and technology, clean, renewable energy, and small business, because most of the new jobs will come from small business. We also have to make the economy fairer. That starts with raising the national minimum wage and also guarantee, finally, equal pay for women's work."

Sanders Klobuchar Biden Buttigieg Warren Steyer

According to our single-word model, this response would most likely have been given by Elizabeth Warren

# Who is most likely to have said it?

"I don't mind releasing -- I'm under a routine audit. And it'll be released. And -- as soon as the audit's finished, it will be released. But you will learn more about Donald Trump by going down to the federal elections, where I filed a 104-page essentially financial statement of sorts, the forms that they have. It shows income -- in fact, the income -- I just looked today -- the income is filed at \$694 million for this past year, \$694 million. If you would have told me I was going to make that 15 or 20 years ago, I would have been very surprised. But that's the kind of thinking that our country needs. When we have a country that's doing so badly, that's being ripped off by every single country in the world, it's the kind of thinking that our country needs, because everybody -- Lester, we have a trade deficit with all of the countries that we do business with, of almost \$800 billion a year. You know what that is? That means, who's negotiating these trade deals? We have people that are political hacks negotiating our trade deals."

**Donald Trump** said this in a 2016 debate. If this response had been given in a 2020 debate, who is most likely to have said it?

# Joe Biden

"I don't mind releasing -- I'm under a routine audit. And it'll be released. And -- as soon as the audit's finished, it will be released. But you will learn more about Donald Trump by going down to the federal elections, where I filed a 104-page essentially financial statement of sorts, the forms that they have. It shows income -- in fact, the income -- I just looked today -- the income is filed at \$694 million for this past year, \$694 million. If you would have told me I was going to make that 15 or 20 years ago, I would have been very surprised. But that's the kind of thinking that our country needs. When we have a country that's doing so badly, that's being ripped off by every single country in the world, it's the kind of thinking that our country needs, because everybody -- Lester, we have a trade deficit with all of the countries that we do business with, of almost \$800 billion a year. You know what that is? That means, who's negotiating these trade deals? We have people that are political hacks negotiating our trade deals."

Sanders Klobuchar Biden Buttigieg Warren Steyer

According to our model, this response would most likely have been given by Joe Biden (unfortunately, we counted "--". New words, like \$694 are also predicted to be Biden since he had the highest proportion of words spoken)

# Who is the heir to the Obama legacy?

As a note, this model was biased toward Joe Biden. We tested 10 responses each from Trump, Obama, Clinton, and Romney. Our model predicted Joe Biden on all responses but five: four Clinton responses and one Obama response were attributed to Elizabeth Warren.

More (and balanced) data would improve these predictions. (Also removing the "--" from the new data)

Additionally, having weights for words could improve accuracy.