

Language Analysis of SOTUs

Project Team 16

The problem

State of the Unions

The State of the Union Address is an annual message delivered by the President to the US Congress, addressing all topics in politics considered most important by the American public.

Voter Turnout

Voting during election season can be overwhelming. Many individuals do not know what topics to focus on when they are preparing to head to the voting booth for the next presidential election.

Problem statement

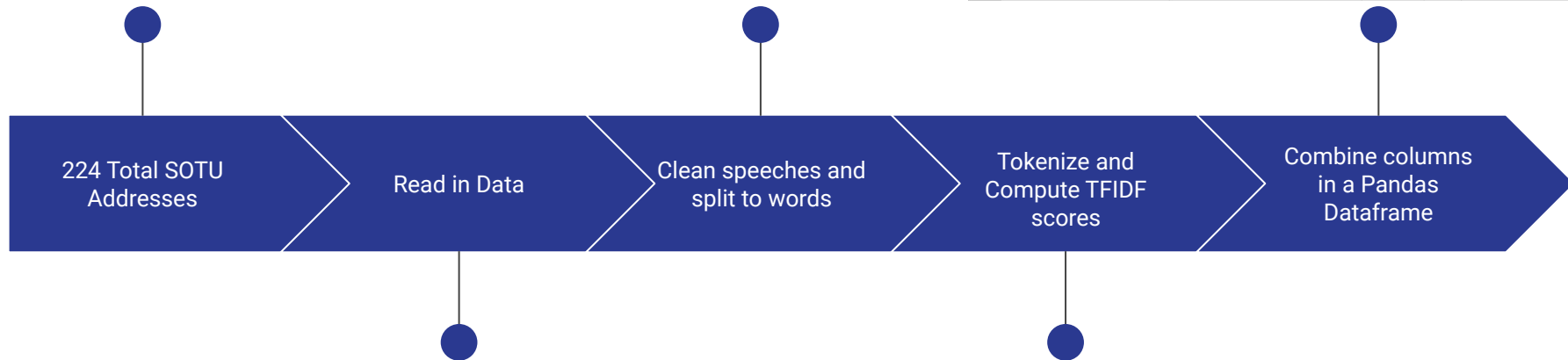
How do we combat low voter turnout? Is there a way to create a predictive formula to see what major topics will be discussed in the future?

Creating the Dataset

We downloaded all SOTU addresses from Kaggle.com.

We used regex and split each speech into lowercase tokens

	pres_sotu	speech	president	year	tokens
0	state_of_the_unions/Reagan_1985.txt	Mr. Speaker, Mr. President, distinguished Memb...	Reagan	1985	[mr. speaker, mr. president, distinguished, me...
1	state_of_the_unions/Nixon_1972.txt	Mr. Speaker, Mr. President, my colleagues in t...	Nixon	1972	[mr. speaker, mr. president, my, colleagues, i...
2	state_of_the_unions/Truman_1951.txt	Mr. President, Mr. Speaker, Members of the Con...	Truman	1951	[mr. president, mr. speaker, members, of, the, ...]
3	state_of_the_unions/Johnson_1964.txt	Mr. Speaker, Mr. President, Members of the Hou...	Johnson	1964	[mr. speaker, mr. president, members, of, the, ...]
4	state_of_the_unions/Buchanan_1859.txt	Fellow-Citizens of the Senate and House of Rep...	Buchanan	1859	[fellow, citizens, of, the, senate, and, house, ...]
5	state_of_the_unions/Bush_1991.txt	Mr. President, Mr. Speaker, members of the Uni...	Bush	1991	[mr. president, mr. speaker, members, of, the, ...]
6	state_of_the_unions/Adams_1799.txt	Gentlemen of the Senate and Gentlemen of the H...	Adams	1799	[gentlemen, of, the, senate, and, gentlemen, o, ...]



Using pandas and numpy we read in the data row by row.

We tokenized and computed the corresponding TFIDF scores for each word.

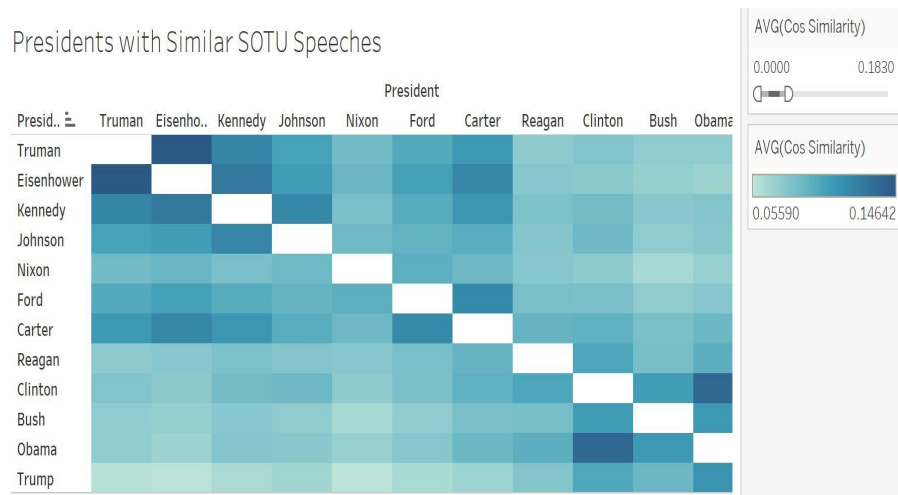
Visualizations

Figure 1

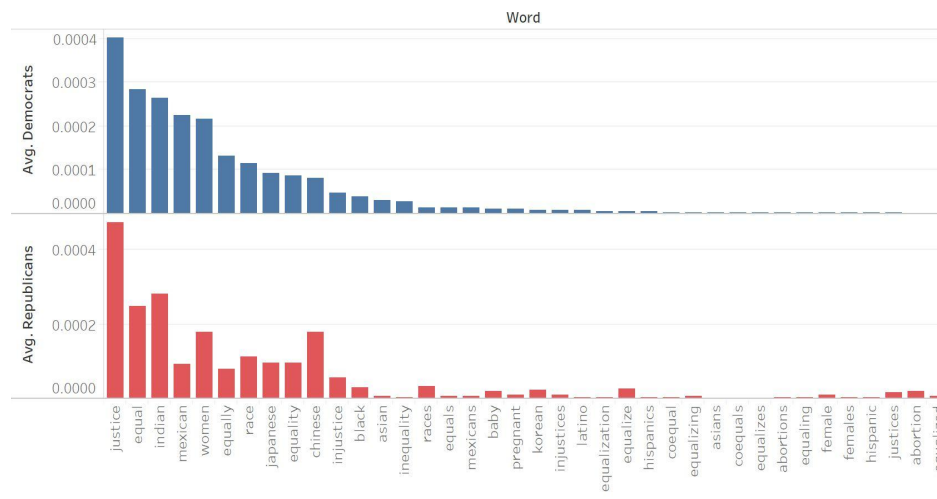
TFIDF & Cosine Similarities

Figure 2

Presidents with Similar SOTU Speeches

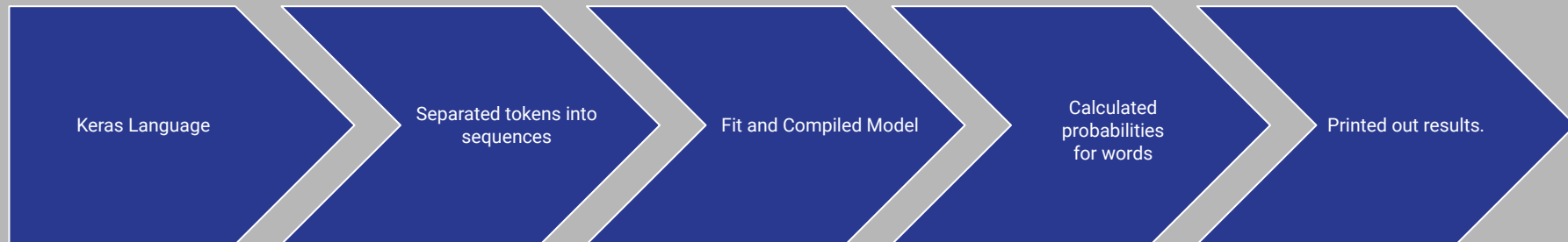


Women's Rights, Social Justice, Race Equality



Language Models and Markov Chain

1. We created a language model that was built on the 2017 Trump State of the Union Address.
2. We also used the predictive power of a Markov Chain using the same speech.



Results of Predicted Text: Language Model

Starting Phrase 1

"companies and workers be taken advantage of us any longer they have taken advantage of our country no longer i am going to bring back millions of jobs protecting our workers also means reforming our system of legal immigration the current outdated system depresses wages for our poorest workers and puts"

"allow a justice to the united states supreme court from my list of 20 judges who are republicans to be able to break the cycle of violence the department of disunity overseas and unite to form i moment out and border re to lay who mistreated the construction and flexibility up in the u the resources that trivial dame you global project after another but ignored we have blessed its series of tragic foreign policy disasters solving these and so many other pressing problems are not jamiel II announced at the united states is much in the muslim great wall"

Accuracy ~ 0.52

Starting Phrase 2

"is not compassionate but reckless to allow uncontrolled entry from places where proper vetting cannot occur those given the high honor of admission to the united states should support this country and love its people and its values we cannot allow a beachhead of terrorism to form inside america and we"

"allow our nation to improve policy and is such are beginning to do us in the gallery a dream millions for vandalism of our current immigration drug became a new citizens who wear our independence it is placing its message of israel the principles no to need the world my current anniversary of our founding 250 years was such you also with this that was 15 where so shot from new chapter of a highly successful trudeau will grow up in a nation community safely on the progress i ve made with keeping a 6 trillion we could have rebuilt"

Accuracy ~ 0.95

Results of Predicted Text: Markov Chain

