

What's in a Speech?: Presidential Topic Modeling

A UVA Data Science Case Study by Avneet Chhabra, 2023

— BBC

[State of the Union 2023: Five takeaways from Biden's speech](#)

Joe Biden delivered an animated and at times combative State of the Union speech at a pivotal moment for his presidency.

Feb 8, 2023

The Guardian

[‘Democracy will wither’: Barack Obama outlines perils of unregulated big tech in sweeping speech](#)

In a keynote address at Stanford University, the former president made his most extensive remarks yet about the tech landscape.



Prompt: It's no doubt that being the president of the United States is a big job. Being elected for political reasons, with the promise of acting in ways that are in line with campaign ideas, presidents are not immune from the short-term, re-election way of thinking. Because of this, presidents are *incredibly* careful with what controversial issues they speak out on. Past research has confirmed that presidents will typically focus on giving public speeches about issues that can unite the country. As digital media makes it easier for presidential sound-bites to go viral and reach massive audiences, it's important to take note of what presidents deem significant and unifying enough to speak out on. Your task is to use data science tools that analyze text-based data to develop useful insights about the topics of presidential speeches.

Deliverable: You must produce a model that selects the most common topics from presidential speeches. You may elect to do so over certain time periods or for all of the recorded speeches available in the data. Most importantly, you must produce a model with coherent, translatable results that allow you to draw conclusions about what presidents choose to speak about. A final product should have a list of 3-5 of the most common themes/topics. After creating the model, produce a presentation for your peers that is also suitable for students unfamiliar with data science. Focus on the aspects of your model and results that really matter.

Case Study Contents:

- Rubric – what do you need to be successful?
- “The Politics of Presidential Speeches” & “Analyzing policy issues in presidential speeches and the media: An Agenda-setting study”
 - Feel free to skim these lengthy scholarly articles for an idea of why this is important.
- Topic Modeling in Python: Latent Dirichlet Allocation (LDA), Towards Data Science
 - Model overview with example code in python
- Beginners Guide to Topic Modeling in Python, Analytics Vidhya
 - What are the terms you need to know and understand?