

Rubric

Case Study by Julia Schopper

General Description: Submit a Github repository for this project.

Why am I doing this? The goal of this assignment is to use topic modeling to determine what policy priorities are most frequently referenced in State of the Union addresses which indicates the preferences of the American public.

What am I going to do? Clean the data, train the model using Latent Dirichlet Allocation, and revise the topics to interpret the top topics discussed in the SOTU addresses over time.

Tips for success:

- Remove “stop” words (words that occur frequently that have no real meaning) to improve the accuracy of your topics
- Be creative! There are many ways to go about answering the President’s question. Feel free to explore other options. One such example could be using sentiment analysis to analyze the public’s reaction to the President’s past SOTU addresses.

How will I know I have succeeded? You will know you have succeeded when you follow the criteria below

Spec Category	Spec Details
Formatting	
README.md	
LICENSE.md	
SRC Folder	
DATA Folder	
Figures Folder	
References	