Examining Executive Orders using NLP

Laura Friese

Do Executive Orders Vary?

This analysis will examine the relationship between various factors and the word choice of the executive orders . We will also examine the use of these strategies over time, or for each president individually. Knowledge of common rhetoric is a vital tool in the dismantlement of misinformation, which is so important in our current political climate of “fake news.”

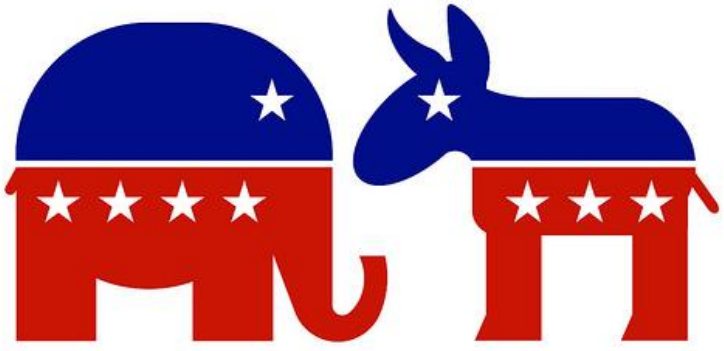
About Our Data

This dataset is from the national register, via the Office of the Federal Register. It records every executive order given since 1994, spanning the presidential terms of Trump, Obama, Bush, and Clinton. The total number of executive orders in this time period is 928

Intro

Political Parties

Presidents from different political parties could write executive orders differently



Term

The term (first or second) that a president writes a given executive order could affect the word choice or contents of that order

Political Party

Understanding the linguistic and word choice differences between political parties can provide a lot of information of interest. This could provide insights about the values of each party, as well as the typical persuasive strategies and rhetoric each party uses.

Term

It’s also possible that the content of executive orders could vary by term. Perhaps a “newer” president chooses different words or topics to issue executive orders on than a more experienced president.

Methodology

- Cleaning the Data
- Creating Additional Variables
- Analysis on Length
- Creating the Party and Term Models

Data Cleaning

Pulling the text from these URLs and attaching it to the DataFrame was the bulk of the data cleaning process. I extracted the HTML text of the executive orders from the HTML links using Beautiful Soup and urllib. To extract just the text of the order, I identified phrases at the end and beginning of the orders and used them to slice the text. There was an issue with the splitting, however: many of the executive orders were actually not in HTML format .To fix this, I removed all pages that did not have an HTML copy of the executive order, removing 223 of our 928 observances.

Creating Additional Variables

I created three additional variables after the data cleaning process. They were:

- President (using the executive order numbers)
- Political Party (using the name of the president)
- Term (using the date of the order)

Analysis on Length

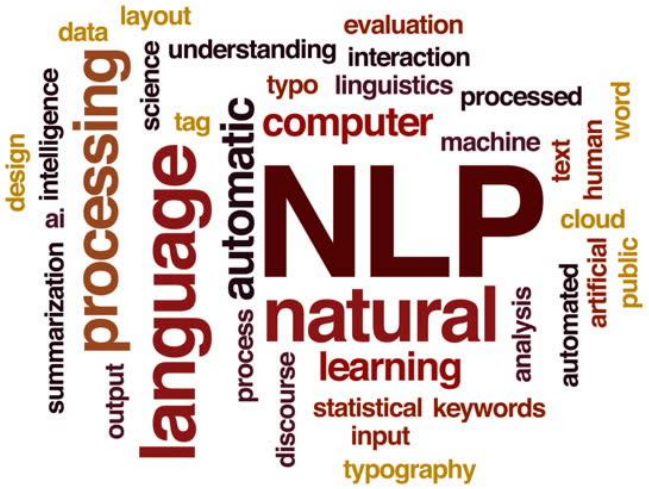
Average Republican executive orders were 6545.6 characters long, and Democrat's executive orders were 7998.9 characters long. However, Trump's executive orders were 8724.5 characters long on average, Obama's were 8187.8, Bush's were 5894.1 and Clinton's were 7064.6. The high number of Trump's was likely balanced out by Bush's low order length. Initial statistical observations also indicated an increase of executive order length over time, but this increase was so small that it is likely clinically insignificant, even if it is statistically significant

Creating the Party and Term Models

To create the models, we use train test split, from the sklearn package, to split the data into a training set, with which to train the model, and a test set, which we will use to test the accuracy of the trained model. The model we are using will be TfidfVectorizer, as this model works well with very sparse datasets. The classifier chosen for this project is the MultinomialNB (Naive Bayes) classifier, which is suitable for discrete variables

Results

- The Party Model
- Party Model Features
- The Term Model
- Term Model Features



The Party Model

We found the model to be able to predict the party of the test executive orders with 65% accuracy. it’s still much higher than chance, indicating that there are some features or words that can be used as predictors for each party. Democrat words start with A. This is because many words had the same weight, and thus the model sorts them alphabetically.

Party Model Features

Democrats	Republicans
aapi	secretary
aapis	act
abandoning	amended
abate	national
abduction	federal
abetted	agency
aboriginal	security
abortion	inserting
abortions	council
absentee	property

The Term Model

The best model resulted from an alpha value of 0, and created a slightly better model with an accuracy score of about 60.1%. This, however, is not much higher than the natural division of first and second term executive orders. The blank lines in the first term were clerical pieces that were not properly removed from the corpus. The fact that they're the first terms in this list, as well as all the words starting with a, indicates that like with the previous words, this list was done in alphabetical order

Term Model Features

First Term	Second Term
____	secretary
_____	national
_____	act
_____	federal
aapi	property
aapis	agency
abbreviations	amended
abedinico	security
abetted	government
abiding	director

Note: The lines were clerical errors, indicating a lack of highly weighted words

Conclusions

- Analysis
- Limitations
- Future Projects

Analysis

It does not appear that we are able to create a good predictive model to determine the party or the term of the president based on the text of the executive order. We found which words were more common, and we were able to garner some predictive power, but the highest we could achieve came from the party model with an accuracy of 66%. This is higher than chance, but is still not an incredibly reliable model. Additionally, the repeated high weight words in the term model indicate that our findings may have more to do with frequency used words and the distribution of the executive orders

Limitations

This analysis could be further improved by using a more robust dataset. We had to remove a large number of the executive orders that hadn’t been digitized in the same format as the others, and that further limited the scope and applicability of this dataset. With more historical data perhaps the model would have been much more accurate. Additionally, the examination over time would have been much more fruitful and the variation between individual presidents would have mattered less.

Future Projects

Given our findings, there are many ways that we could create better models or do more research in the future, such as:

- Repeat this model building with more historical data
- Examine factors such as age or political climate
- Sentiment analysis could better show the differences between political parties' executive orders

Thank You!