**MSDS 692 Project Proposal: Presidential Sentiment Analysis**

Jeremy A. Beard

jbeard@regis.edu

Regis University: Anderson College of Business and Computing

MSDS 692: Data Science Practicum

Christy Pearson

August 27, 2023

**MSDS 692 Project Proposal: Presidential Sentiment Analysis**

In this project, there were a few different ideas I had, a few different avenues for which I wanted to find data. I wanted to either perform some political analysis, perform some medical imaging prediction (such as pneumonia analysis on lung scans), or perform some analysis on recent Data Science university graduates and predict my own post-graduate performance, such as predicting salary or other details about employment. I really wanted to perform the analysis on Data Science graduates and see if they’re working as a data scientist currently and what their salary is. Unfortunately I wasn’t able to find a lot of great data on this topic. I did find a lot more data surrounding politics and imaging prediction. There were some useful datasets I found on Kaggle in the form of presidential speeches throughout the years. We will be utilizing either a dataset about generic presidential speeches (*United States Presidential Speeches*, 2020) or a dataset about the SOTU addresses through the years (*State of the Union Corpus (1790 - 2018)*, 2018). This should show us how different political topics have shifted throughout the years, and what the priorities of different presidents have been. Sentiment analysis on each of the speeches will show the different sentiments gleaned from the various speeches and how that has changed.

**Description**

As stated before, this project will explore some speeches given by presidents through the years. I expect to choose the dataset which had SOTU speeches from presidents from Washington all the way to Trump. This SOTU speech is indicative of the atmosphere surrounding the presidency at the time and it will be interesting to explore how the atmosphere has changed over the years. It will be especially interesting to have data reaching all the way back to the first president of the United States of America.

**Methods**

This project will largely utilize natural language processing techniques in order to achieve its purpose. Some objectives which come to mind are finding the top 5-10 subjects of each SOTU speech, analyzing the general sentiment of each SOTU speech, creating word clouds as an accessory activity. It would be especially useful to try to quantify the sentiment of each of the speeches and create a visualization of this.

**Project Timeline**

|  |  |  |
| --- | --- | --- |
| **Week** | **Deliverables** | **Date** |
| **1** | * Project Proposal | August 27 2023 |
| **2** | * Datasets selected * Week 2 Status Report | September 3 2023 |
| **3** | * Initial data loading and data cleaning performed * Week 3 Status Report | September 10 2023 |
| **4** | * Main analysis on single speech finished * Week 4 Status Report | September 17 2023 |
| **5** | * Main analysis extended to all speeches * Week 5 Status Report | September 24 2023 |
| **6** | * Analysis on all speeches congregated, summary visualizations * Final presentation started * Week 6 Status Report | October 1 2023 |
| **7** | * Presentation completed, initial dry runs made * Week 7 Status Report | October 8 2023 |
| **8** | * Final Project Presentation | October 15 2023 |

**References**

*State of the Union Corpus (1790 - 2018)*. (2018, October 19). Kaggle. https://www.kaggle.com/datasets/rtatman/state-of-the-union-corpus-1989-2017

*United States presidential speeches*. (2020, May 31). Kaggle. https://www.kaggle.com/datasets/littleotter/united-states-presidential-speeches