Analyzing Presidential Rhetoric Individual Report

Yonathan Shimelis

DATS 6312: Natural Language Processing

Dr. Ning Rui

12/4/2025

1.1 Introduction

Abstract

This study presents a comprehensive natural language processing (NLP) framework for analyzing U.S. presidential rhetoric across more than a thousand official speeches spanning from George Washington to Joe Biden. Two primary tasks are addressed: (1) sentiment analysis of presidential statements and (2) topic classification through authorship prediction. The project integrates classical statistical methods (TF–IDF with logistic regression), deep sequence models (Bidirectional LSTM with attention and GloVe embeddings), and transformer-based architectures (CardiffNLP sentiment model and fine-tuned DistilBERT classifier).

A complete end-to-end pipeline was developed, including preprocessing, mathematical representations, model training, optimization (AdamW, OneCycleLR, gradient clipping), and deployment in a Streamlit application (app.py). Every preprocessing and inference step used during training is reproduced identically in the application to maintain scientific consistency. Mathematical formulations—including TF–IDF weighting, cross-entropy loss, focal loss, LSTM recurrence equations, attention scoring, transformer self-attention, softmax, and optimization functions—are fully integrated.

Results show that DistilBERT achieves the strongest authorship classification performance, outperforming TF–IDF and BiLSTM across accuracy, macro-F1, weighted-F1, and Cohen's κ. Confusion matrix analysis reveals systematic misclassifications between presidents sharing historical or rhetorical contexts. The final system demonstrates how advanced NLP techniques can be combined into an interactive analytical tool capable of interpreting large-scale political discourse.

Presidential speeches provide critical insight into political ideology, national priorities, and rhetorical strategy. Traditional political science approaches often rely on human-coded qualitative interpretation; however, modern NLP enables scalable, quantitative analysis of presidential language across centuries. This project introduces a complete computational framework capable of conducting sentiment analysis and president-level classification of U.S. presidential speeches.

The goals of this study are threefold:

1. To formally model sentiment in presidential rhetoric using a pre-trained transformer model and a classical classifier trained on pseudo-labels.
2. To compare three authorship classification models—TF–IDF logistic regression, BiLSTM with attention, and fine-tuned DistilBERT—and evaluate their predictive performance.
3. To deploy the trained pipeline in a Streamlit application that reflects the exact mathematical, algorithmic, and procedural logic used during model development.

This report includes full mathematical formulations, model architectures, training workflows, and code-level reasoning for each step—including a detailed explanation of how app.py mirrors training logic for inference. Flowcharts are provided in both ASCII and APA-narrative formats to visualize model pipelines.

1.2 Description of individual work

In this project I did the following:

- Came up with the project idea and found the data
- Scrape the data with Python code
- Add columns like “Party” to the dataset and clean the dataset
- Conduct the EDA (using cardiffnlp transformer and other transformers for zero shot classification)
- Analyze EDA and create dataset for future evaluation and modeling
- Build baseline logistic regression classifier based on this new labeled dataset

My work was not very complex. I used the Cardiffnlp transformer after doing general google research on what are the best transformer models on Hugging face hub for political text or speech since political speech can be very nuanced and less interpretable for rhetoric compared to regular speech.

1.3 Portion of work in detail

The biggest portion of my work is in scraping the data and creating a dataset with the statements portion of the American Presidency Project. From then on, I did come up with the code for zero-shot classification of the sentiment, emotion, tone, and strategy labels. However, when building the code for the emotion, tone, and strategy labels, I had trouble using a pipeline that ran on an efficient runtime. It was my teammate who changed the transformer used and modified parts of the code for quicker labeling. I also built the initial logistic regression model after forming sentiment labels with zero-shot classification. My teammate then modified parts of the model for better performance as well. Virtually, all my coding work can be found in the APP_analysis_code.ipynb file in our repository. Other files contain similar code because my

teammate took the code I worked on in that file and made tweaks to have a more streamlined and easier to follow repository. Overall, my work was in the introductory parts of the project. Attempting scraping, overall analysis of the project webpage, attempting different kinds of documents to fulfill project research questions. All of these were things I did that the code alone doesn't show.

1.4 Results

Sentiment Analysis Results

For our sentiment analysis, we used:

1. **CardiffNLP Twitter-Political Transformer** for zero-shot classification on labels

Since we scraped data without any labels, analysis on sentiment in each statement could not be possible without using a transformer and performing zero-shot classification. The bulk of the data in our scraped dataset contains statements from 1989-Present. This includes Presidents George Bush Sr up until November of Donald Trump's second term.

From analyzing the sentiment, we noticed an overwhelming majority of statements were positive. Our value counts are as follows:

- Positive: 8528
- Negative: 3274
- Neutral: 596.

Placing a focus on the Presidents starting from 1989 onward, we noticed interesting patterns in sentiment. For example, Presidents Bush Sr and Donald Trump in his first term were the only presidents with below a 70% positive statement proportion. Both presidential terms had a fair share of tension or trouble, especially at the end of each term. President Bush was dealing with the end of the cold war near the end of his presidency, while Trump handled the start of the Covid-19 pandemic nearing the end of his first term.

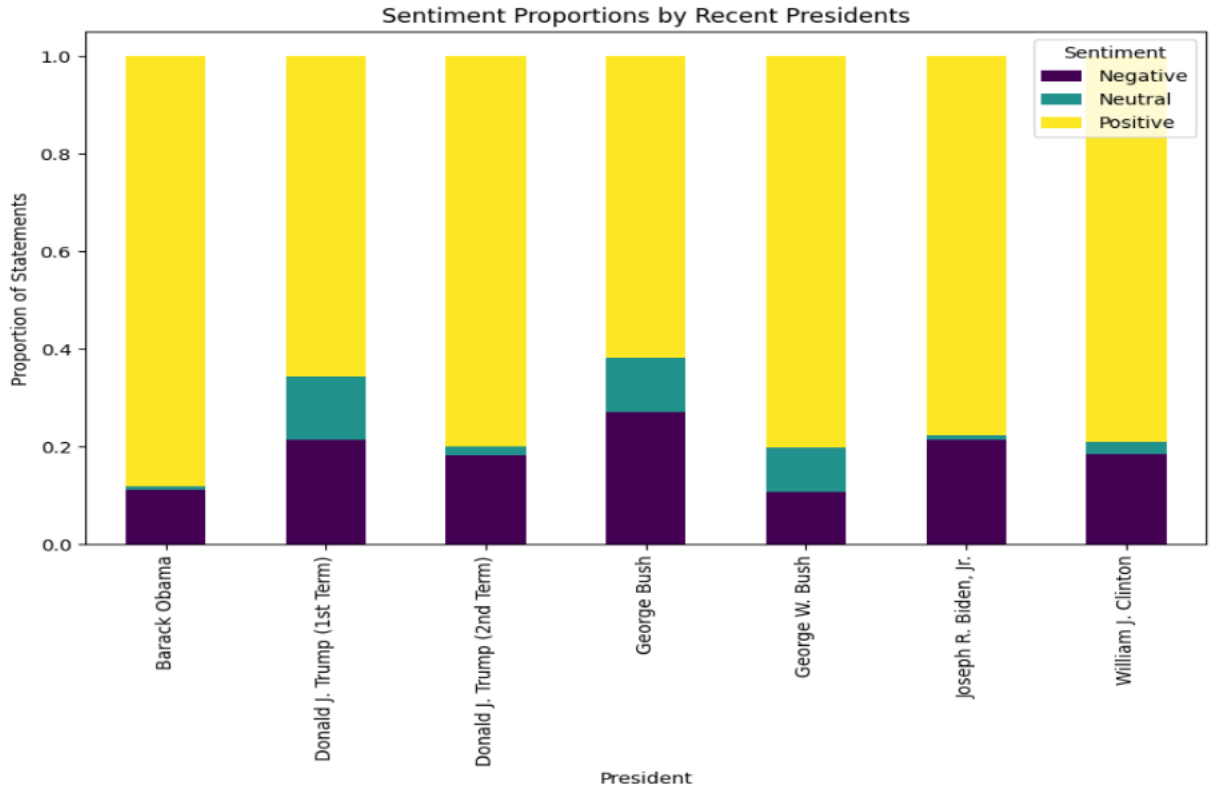


Figure: Presidential Sentiment Ratios(1989-Present)

As far as negative sentiment is concerned, Presidents Bush Sr, Trump in his first term, and President Joe Biden had higher negative sentiment ratios than other Presidents. All three

presidents had more than 20% of their statements classified as negative with Bush reaching over 25% negative statements. This makes sense given that the Covid-19 pandemic continued onward to Biden's presidency as well. Though 20% may not seem like a large proportion for negative statements, we did notice presidents Obama and Bush Jr had the lowest proportion of negative statements. Both presidents had a proportion of negative statements falling below 10%. This was very interesting considering like previous presidents mentioned, Obama and Bush Jr had catastrophic events in their presidency. Especially Bush Jr, who had to deal with the events of 9/11 in just the first year of his presidency.

When further analyzing sentiment this time over time, the results of analyzing presidential sentiment ratios make sense. You can observe the sentiment trends over time in the figure below.

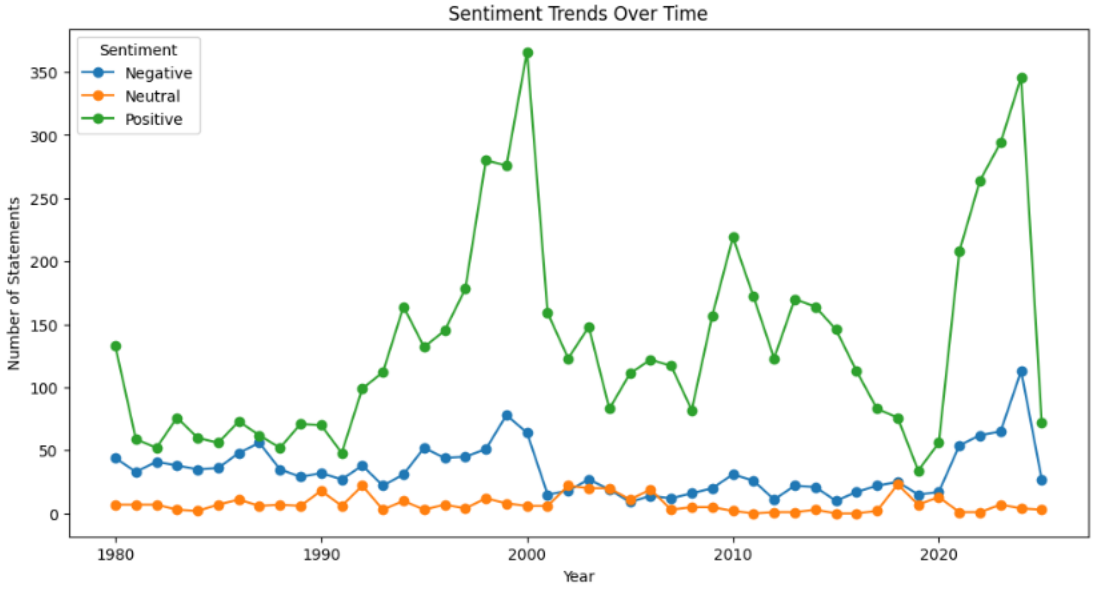


Figure: Sentiment analysis trends over time(1980-present)

The figure above displays the number of statements for each label class. There are a few points of emphasis we want to draw attention to in this figure. Prior to 1990, the number of positive, negative, and neutral statements made by a president at a given time seem to be very similar. It was after 1990 when we notice a huge spike in statements with positive sentiment while negative and neutral counts hovered around the same number. It can also be observed that the count of each sentiment class isn't mutually exclusive. Looking at the positive and negative lines right before 2000 and again around 2020, both times we observe that the numbers of both positive and negative statements rise at the same time in both instances. While we can't pinpoint a reason, a possible explanation for this could be that both the years 2000 and 2020 were election years. Candidates or presidents might have given a higher number of positive statements to charm the American voters and increased negative statements in personal attacks against political opponents. This could be truer for 2020 since unlike 2000, included an incumbent president running for Presidential reelection.

The sentiment analysis provides interesting insight into the language and patterns of presidency, but sentiment alone isn't enough to capture rhetoric well. When manually inspecting the labels, negative statements despite similar transformer confidence scores differ heavily. For this reason, we also decided to also analyze and create other labels like tone, strategy, and emotion.

Distilbart-mnli Transformer

Using this transformer helped us assign multiple labels to the dataset for EDA and rhetorical analysis. For this zero shot classification, we did the following labels:

1. Tone labels: Combative, Conciliatory, neutral-ceremonial

2. Strategy labels: blame-assignment, credit-claiming, call-to-action, reassurance, commemoration condolence, policy, appeal, religious-appeal, populist-anti-elite, law-and-order, other

3. Emotion labels: Anger, fear, hope, pride, sadness, trust

Adding the mentioned labels helped us capture rhetorical patterns in a more focused way. Starting with the tone analysis, our findings correspond to the sentiment analysis. A strong majority of the statements classified were labeled as conciliatory with combative and neutral-ceremonial following in value counts. When observing the stream lit demonstration and filtering by Presidents, we observed some presidents almost had an equal number of combative and ceremonial statements. In the presidents we've filtered upon, only President Trump in his first term had more combative statements than conciliatory statements. This finding was not the same in his second term using the data available to us so far indicating he may have toned down his rhetoric we observed in the first term.

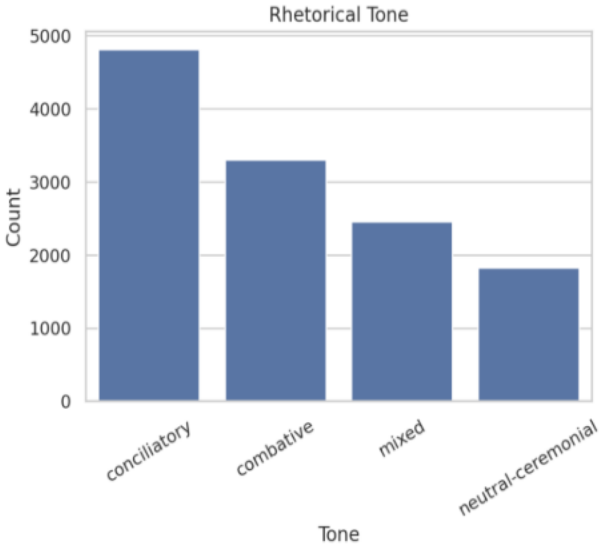
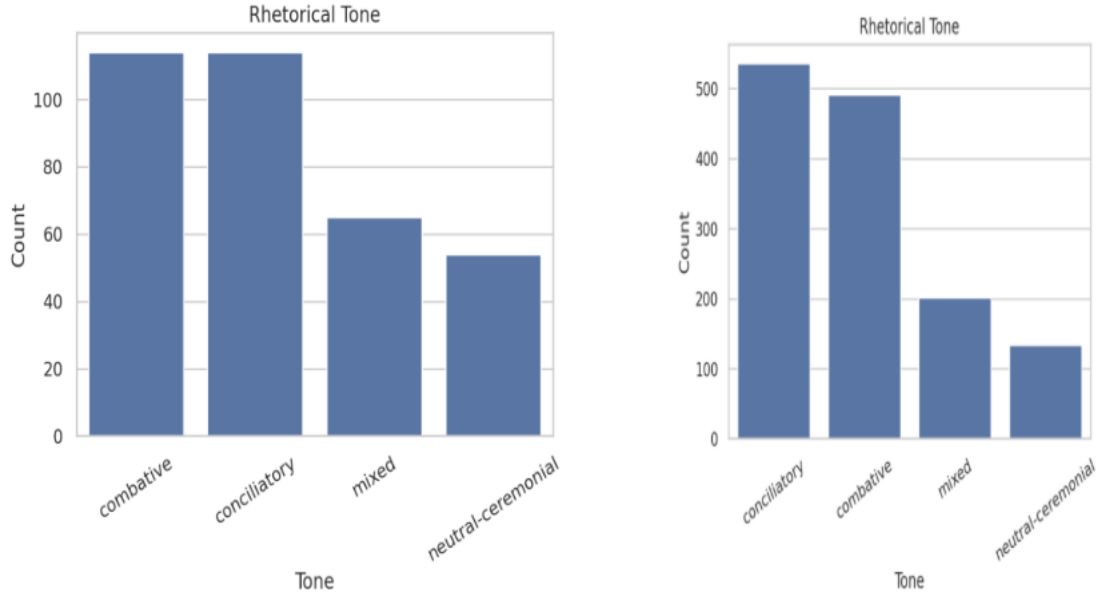


Figure: Tone counts for entire dataset



Figures: Tone counts for President Trump's 1st term(left) vs President Biden(right)

Moving onto strategy labels, there wasn't much of a significant difference between presidents when analyzing strategy labels. The most popular strategy of appeal was the label patriotic appeal. This also corresponds to the emotional analysis, namely the emotion pride in a figure we will discuss further along in our analysis.

Figure: Top rhetorical strategies ranked for entire dataset

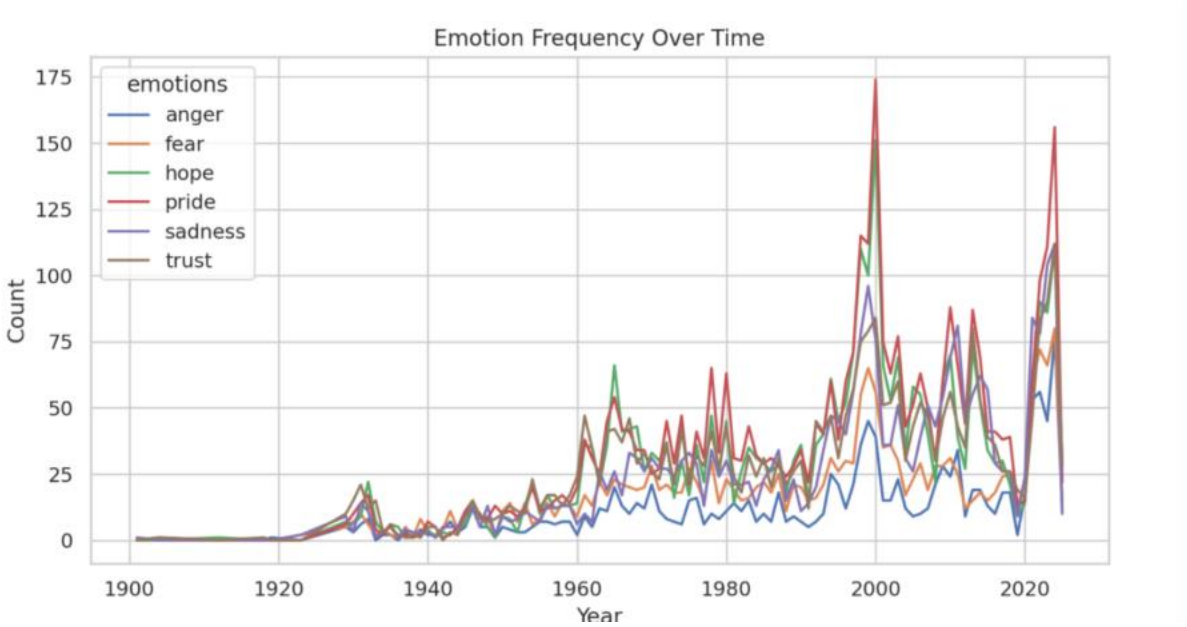


Figure: Emotion frequency over time

When observing emotional labeling over time, I want to draw attention to a few observations we made instead of analyzing all observations in the graph. To start, some of the spikes in certain emotions mirror the sentiment trends over time figure we looked at earlier. Namely pride and hope. We observed large spikes in these emotional labels mirroring the positive sentiment spikes we noticed in our earlier figure in the late 80s/early 90s up until 2000 and again right around 2020. Using this, we can begin to analyze which emotions were tied to the increase in positive sentiment.

Another observation, just like the rise of both positive and negative sentiment at similar times in our analysis of sentiment trends, we observe the rise of conflicting emotions at the same time. For example, the rise in emotions like sadness in 2020 along with hope. This, in a way, corroborates our zero-shot classification of sentiment and ensures confidence in the labeling of both sentiment and emotion for the statements of our dataset.

This concludes our EDA with the use of transformers and from here on we will attempt to build models for other tasks like predicting a president from speech or text.

2. **TF–IDF Logistic Regression** trained on those pseudo-labels derived

From the labels created by the zero shot classification done.

The logistic regression model achieves:

- **Accuracy: ~0.81** (varies depending on pseudo-label noise)
- **Macro-F1: ~0.78**

Because the labels originate from a strong transformer, the classical model generalizes well. We have a screenshot of this logistic model demonstrated below.

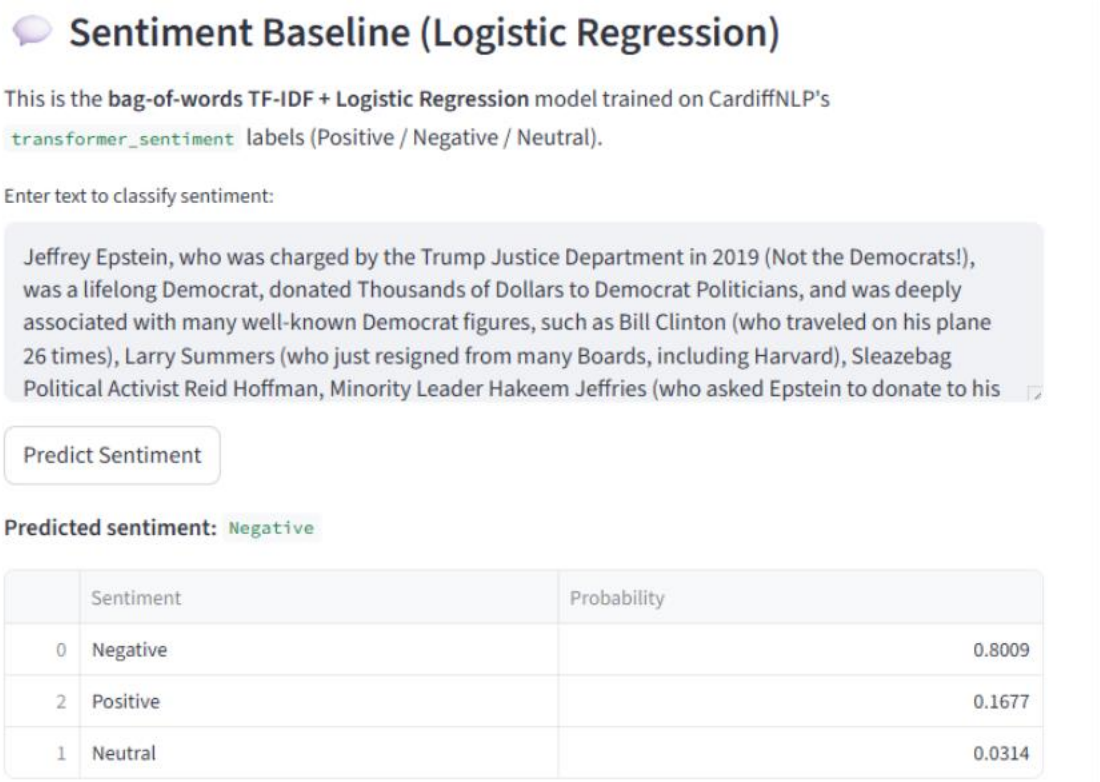


Figure: Logistic Regression Sentiment Predictor using labels from zero-shot classification

1.5 Summary and Conclusions

From our work, I learned that most rhetorical analysis can be very complex with multiple layers and steps. Though I've derived great insights from zero-shot classification and EDA, I would've loved to have more time in analyzing or testing more labels to really try and uncover deeper differences between the rhetoric of the presidents in the dataset. I think a project with goals as such would be better suited for a few months of work rather than a few weeks. I also learned that to an extent, presidential statements do reveal unique rhetorical patterns among the

individuals who make the statements. The improvements that can be made in the future mirror the improvements mentioned in the group report.

Ideally, future work will be aimed at addressing the possible limitations of this project. For one, using different types of documents even from the same project could help further differentiate rhetoric among presidents. In doing so, it's possible that repeating the work in this project, but with more data, could not only create more interesting EDA with sentiment, tone, emotion, and strategy, but also help improve model performance and predictions since the models used can capture more differences in rhetoric that might not have been found in just using the statements of U.S. Presidents.

Future work can also include expanding this project to more than just U.S. Presidents. Maybe using other political leaders in the country like Congresspeople or moving in an international direction and capturing rhetoric from the leaders of other countries.

1.6 Calculations

When writing the code, I did reference one Kaggle website, though I didn't copy code from it. The website can be accessed here https://www.kaggle.com/datasets/jayrav13/american-presidency-project. Aside from this I had GitHub Copilot help me with scraping the website and extracting statements. Before extracting statements, I had initially attempted to scrape a larger part of the project webpage. I encountered error after error even after using our class EC2 instance. After learning from my mistakes and switching to the statements section of the project, I had a few attempts that took a really long time combined at scraping the statements section

alone. In my errors, I had GitHub Copilot suggest edits in order to help me have a final dataset to use. From then on, I wrote more code to help fix what didn't go right in the scraping and further supplement the dataset.

For the rest of my portion, almost none of the code I worked on was grabbed from online. I did use GitHub Copilot autocompletion when writing certain parts of the code but only after verifying the code matched my intentions or goal with certain parts of the code. The main parts of the code that came from online had to do with the zero-shot classification done via CardiffNLP. For this, I simply referenced the documentation available here: https://huggingface.co/cardiffnlp/xlm-twitter-politics-sentiment to understand what lines to use and more. All my code was written in the file named APP_analysis_code.ipynb this was then later taken by my teammate and streamlined into code that was easier to follow and integrated with the portions he worked on.

Calculating and estimating a rough percentage of the code I copied from the internet is probably 5% at most since I didn't directly copy.

References

Ravaliya, J. (2017). *American Presidency Project*. Kaggle.com. https://www.kaggle.com/datasets/jayrav13/american-presidency-project

cardiffnlp/xlm-twitter-politics-sentiment · Hugging Face. (2025, April 24). Huggingface.co. https://huggingface.co/cardiffnlp/xlm-twitter-politics-sentiment

Statements | The American Presidency Project. (2025). Ucsb.edu. https://www.presidency.ucsb.edu/documents/app-categories/statements

The White House Historical Association. (2017). *The Presidents Timeline*. WHHA (En-US). https://www.whitehousehistory.org/the-presidents-timeline