Political Speech Classification

Individual Report

Minwoo Yoo (Group 3)

The George Washington University

December 8, 2025

Abstract

This individual report summarizes my contribution to the Group 3 project for the final course assignment. Because I was the sole member of Group 3, I completed all components of the project independently, including data construction, model design, implementation, experimentation, and documentation. This report outlines my responsibilities and describes the steps I carried out to produce the final empirical results and presentation.

Individual Contribution

As the only member of Group 3, I performed every part of the project. I managed the data pipeline, starting from loading the Congressional Record text, merging member metadata, and constructing the binary party labels needed for supervised learning. I created all filtered datasets, including the final subset restricted to 1981–1989 speeches with at least two sentences and 200 characters.

I wrote the full TF-IDF baseline pipeline using scikit-learn and implemented the Logistic Regression model with appropriate filtering and class-weight adjustments. I then built the RoBERTa fine-tuning system in PyTorch and HuggingFace Transformers, including tokenization, batching, GPU memory management, mixed-precision training, early stopping based on validation metrics, and the final evaluation code for accuracy, F1, and confusion matrices.

I designed and executed the experiments comparing filtered and unfiltered datasets, multiple epoch settings, and alternative preprocessing schemes. I also generated all tables, figures, and performance summaries reported in the group submission. The Beamer slides for the final presentation and all written documentation, including the group report, were also prepared solely by me.

I developed all code with the assistance of ChatGPT, which supported the drafting and debugging of scripts. All conceptual decisions, experimental designs, and model evaluations were carried out by me.