Exploring the Effects of Presidential Elections on Healthcare Spending

Vivek Atmuri: vatmuri7@gatech.edu
Rhea Saravanan: **rsaravanan8@gatech.edu**
Saksham Purbey: **spurbey3@gatech.edu**
Nathaniel Petersen: **npetersen3@gatech.edu**

Abstract

Presidential elections have the potential to have a significant impact on spending amounts in certain sectors due to the uncertainty they generate, the surfacing of important issues, and the possible impact of policy changes. Government spending has been known to be sensitive to political events, which can have a direct impact on economic policies, regulations, and investor sentiment. To investigate how presidential election debates reflect government spending from 1976 to 2020, we used data from the UCSB Presidential Documents Archive. In this work, we created a novel sentimental analysis task and benchmarked it against various pre-trained language models on the proposed dataset. Using the regression analyses, we constructed a measure of how candidates' stances on critical issues sway government spending in the healthcare sector, offering valuable insights for businesses and investors to anticipate and adapt to changing political landscapes, enhancing their decision-making processes and overall market performance.

1 Introduction

Tracing back to data collected from 1976, it has been shown that there have been marked economic changes after presidential election cycles. While trends of government spending after an election cycle have been studied in the past, we will tackle this from a different angle, using natural language processing to help us achieve a more nuanced understanding. We are exploring this because it is important to understand why these economic changes occur and their impact on markets, businesses, job creation, and the economy as a whole. Throughout this project, our goal is to achieve a better understanding of exactly how and to what extent election cycles have an impact on government spending, and to create a model that can predict the total spending in the upcoming years.

2 Influence of Presidential Elections

Presidential elections can be quite a useful predictor of government spending related to healthcare and the markets it works with due to the amount of information and discussions regarding the administrations' future policies and the current policies that take place. These debates and campaign speeches influence businesses and investors by allowing them to change their investment strategies in certain sectors. For example, a candidate's speech on keeping the nation safe by conducting more COVID-19 tests could influence the healthcare market due to the predicted increase in funding. By looking at presidential elections and the news surrounding them, we will be able to see how certain policies or promises of policies can predict the government's spending of that year and in turn the financial market of the country.

3 Data Collection and Processing

For our data collection, we analyzed the discussion of financial topics in US presidential debates from 1976 to 2020, utilizing data from the UCSB Presidential Documents Archive along with the Commission on Presidential Debates. Focusing on debates between the two candidates to ensure consistency, we employed a two-step approach for preprocessing. First, each team member scraped three transcripts, compiling the relevant sentences into an Excel sheet. We then employed a filtering process to filter sentences related to healthcare using a Python script based on a list of predefined healthcare terms. The study's key target variable is the 'Change in Healthcare Spending', with the dataset comprising 1208 sentences, of which 700 were randomly sampled and labeled as -1 (indicating a decrease in spending), 0 (not relevant), or 1 (indicating an increase in spending). We then took the majority vote on the scores to ensure accuracy throughout our annotations. To enhance our analysis, we also incorporate spending data from the Federal Reserve Economic Data (FRED), examining correlations between the candidates' discussions on financial spending and actual changes in national spending. This approach allows us to provide an understanding of the evolution of financial discourse in presidential campaigns.

4 Model

The research methodology in this study employs the RoBERTa variant of the BERT base model to perform sentiment analysis on presidential debate speech transcripts. Its primary goal is to categorize sentences within these speeches as negative, neutral, or positive, offering insights into future federal healthcare expenditures under the leadership of elected individuals. RoBERTa is the Robustly Optimized BERT model, and its main differences from BERT are dynamic masking, larger training batches, and a larger pre-training dataset. RoBERTa improved on BERT's undertraining with a comprehensive pre-training process on additional corpuses such as CC-News and OpenWebText. Additionally, RoBERTa utilized dynamic masking in its pre-training process by duplication of training data ten times, randomly masking each sentence in various ways, and inputting the different masked variations of the sentences during each of the 40 epochs of training. Dynamic masking resulted in higher GLUE and SQuAD scores relative to static masking, which was used in the BERT pre-training process. The technical implementation involves key libraries, including Transformers and PyTorch. The RoBERTa model is loaded, and its tokenizer is applied to tokenize sentences, converting textual data into tokens. Then, the tokens are initially converted into numerical vectors and then converted into PyTorch tensors, allowing for seamless integration with the PyTorch framework. To ensure robust model performance, the dataset is strategically split into training 60%, fine-tuning 20%, and testing set 20%. The training phase spans five epochs, during which the model dynamically adjusts its parameters through back-propagation, progressively reducing the cross-entropy loss function. After each epoch, the model undergoes evaluation on the validation set, and accuracy and F-1 scores are recorded. Finally, testing is performed to evaluate the model's performance on the testing dataset, the subset of the dataset that was not utilized during training and fine-tuning. Addressing the risk of over-fitting, a common concern in machine learning, is achieved through the use of regularization techniques. In this implementation, dropout is embedded in the BERT architecture, ensuring adaptability and effectiveness by preventing excessive dependence on specific features during training. The implementation of this exploration integrates essential elements for model implementation, including the installation of the Transformers library, while utilizing Pandas for data handling, Torch for fundamental deep learning operations, and AdamW for optimization, highlighting the comprehensiveness and sophistication of the approach.

5 Current Results

The BERT base-uncased model was fine-tuned in this project using the data set built as described in the previous sections. The data set was split into 60 % for training, 20 % for validation, and 20 % for testing. After fine-tuning for 5 epochs, the model was tested and it was found to have an F-1 score of 0.76.

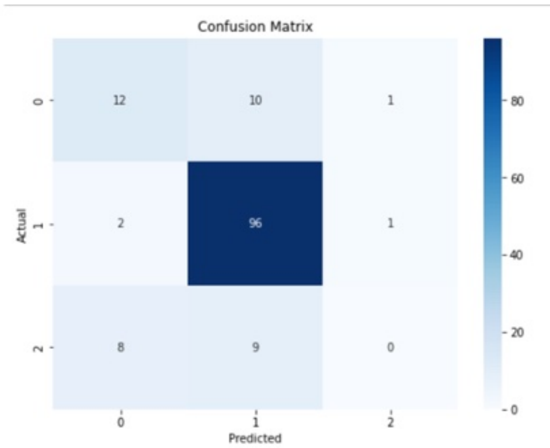

Figure 1: Confusion Matrix Evaluation of Model

For each general election year, a score was computed using the winning candidate's presidential debate speech. First, the sentences of the speech were fed into the model to obtain a classification for each sentence as hinting towards increase in spending, decrease in spending or no relation to spending. The final score for the candidate was computed as

$$\text{score} = \frac{\text{Number of increase} - \text{Number of decrease}}{\text{Number of increase} + \text{Number of decrease}}$$

The score obtained for each presidential general election from 1976 till 2020 was computed for the winning candidate. Using the data obtained about the healthcare spending, the percentage change in outlays by the agency from the year of election to the next year when the elected president signs the new budget is computed for every general election year from 1976 till 2020.

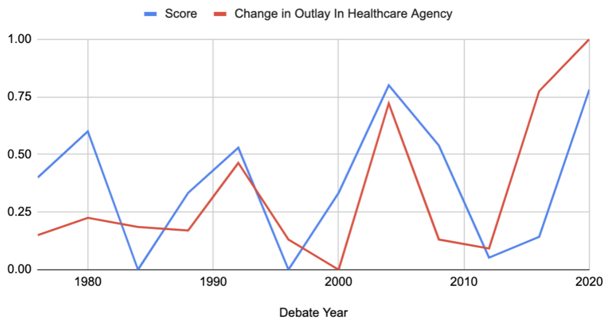

Figure 2: Relation between Deficit Change and Campaign Speech

The correlation coefficient between the percentage change in deficit and score computed was found to be 0.5184

6 Current/Next Steps

With our current results, we can see a potential expansion of this project in several ways. Firstly, in certain debates where there were a small amount of sentences labeled increase or decrease, the score was quite volatile, giving scores that may not accurately represent the healthcare spending sentiment for that president's term. In the future, by weighting the score differently based on how many sentences there were, we may be able to better represent the spending sentiment. Furthermore, we can consider analyzing a larger dataset that includes campaign speeches and interviews to capture a broader spectrum of intentions regarding government spending in the United States. Moreover, we can look into expanding this project by analyzing other languages like Spanish to study the sentiment of voters who don't consider English as their first language. Apart from these potential outcomes, we can look into enhancing the model by collaborating with political scientists and economists to address potential bias in our model and ensure fairness across all demographic groups.

References

[1] Oehler, A., Walker, T.J., & Wendt, S. (2013). Effects of election results on stock price performance: evidence from 1980 to 2008. *Managerial Finance*, *39*(8), 714-736. https://doi.org/10.1108/MF-May-2012-0126

[2] Chava, Sudheer and Du, Wendi and Shah, Agam and Zeng, Linghang, Measuring Firm-Level Inflation Exposure: A Deep Learning Approach (September 23, 2022). http://dx.doi.org/10.2139/ssrn.4228332

[3] Presidential candidates debates (1960-2024). 6.2 Presidential Candidates Debates (1960-2024) | The American Presidency Project. (n.d.). https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/presidential-campaigns-debates-and-endorsements-0

[4] The White House. (n.d.). Historical tables. https://www.whitehouse.gov/omb/budget/historical-tables/

[5] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv.org, Jul. 26, 2019. https://arxiv.org/abs/1907.11692

[6] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv.org, Jul. 26, 2019. https://arxiv.org/abs/1907.11692