Classifying Political Statements by Party Platform

Michaela Eichel (mae97), Diego Aguirre (dta35)
Revised by Diego Aguirre

Using a medium size dataset, a pre-trained Distiblert Model was fine-tuned to perform 2 label text classification for political sentiment. Analysis was performed using a diverse test set, and the fine-tuned model outperformed its base counterpart. While this does illustrate the efficacy of the fine tuning methodology - a more robust analysis would be necessary before deployment.

I. Introduction

Our fine-tuned model is designed specifically to classify segments of written or spoken language as aligning with either the Democratic or Republican party platforms. If successful, this model could be used to help people better understand the United States political system in addition to their own beliefs. There are also various potential applications in bias detection, which could provide more clarity to users.

II. Training

For our training data, we needed a set of statements labeled as either Republican or Democrat, ensuring that each segment was appropriately categorized. This was achieved by using data collected by the Comparative Agendas Project[1]. The Democratic and Republican Party platform datasets compiled by Christina Wolbrecht at the University of Notre Dame provide quasi statements from every Democratic and Republican party platform from 1948 to 2020. The original datasets code the statements with multiple variables but we pruned everything except for the statement and each label during our data-engineering pipeline. For additional insight into the training data, please reference article 1 of the works cited.

For training, we started off with a pre-trained Distilbert (Uncased)[2] base that was trained on the Wikipedia Legacy Datasets[3] and BookCorpus[4].This model is a lightweight sentiment and text analysis tool with 67 million parameters that can compute class label probabilities. We chose this size in order to minimize our training time, as the larger full sized Bert was going to take an excess of 60 hours to train. Running Visual Studio with a 2.3GHz 8-Core Intel i9 in parallel (1357 Single Core Geekbench Score) - it took ~5 hours to train on the 4.82 million character long training set.

Fig 1 - Distribution of Character Lengths

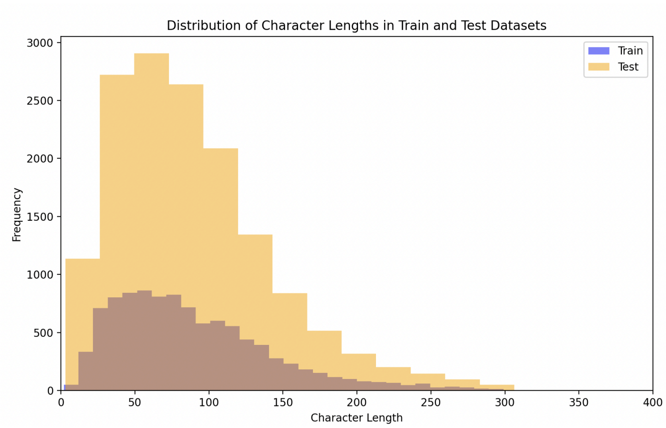


Fig 2 - Test Inputs and Outputs

```
Statement: "We need to lower taxes and reduce government spending."
Predicted Class: Republican

Statement: "Climate change is one of the biggest threats to our planet."
Predicted Class: Democrat

Statement: "We should invest more in renewable energy sources."
Predicted Class: Democrat
```

III. Testing

The test examples were manually collected from personally written statements aligning to common political beliefs, social media posts of politicians, excerpts from political speeches (mostly speeches from the 2024 DNC and RNC), a Danish US Embassy description of the political parties, as well as excerpts from the official 2024 Democratic and Republican party platforms. As a result, the testing data is slightly newer than the training data which is cut off at 2020. For more insight into the data, please reference (Fig 1) and (Fig 2).

Fig 3 - Distribution of Predicted Probabilities

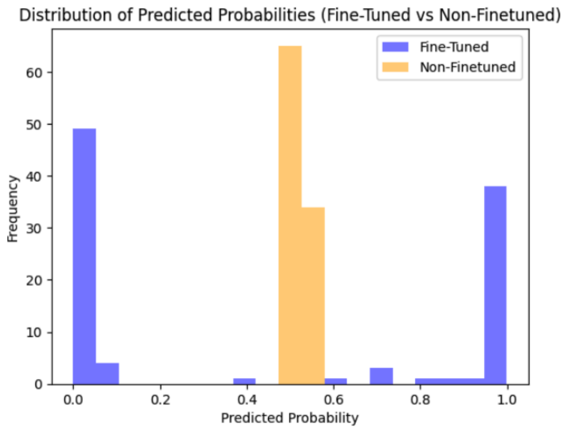


Fig 4 - ROC Curve Comparison

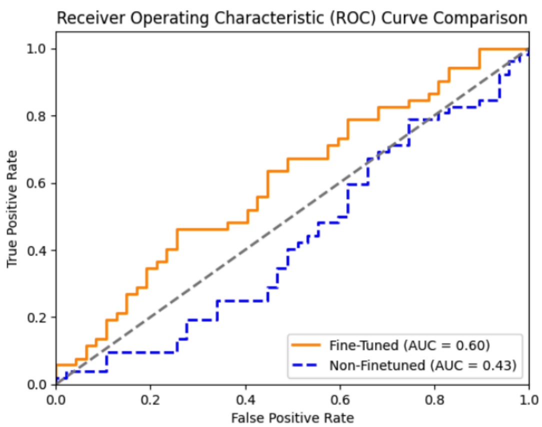
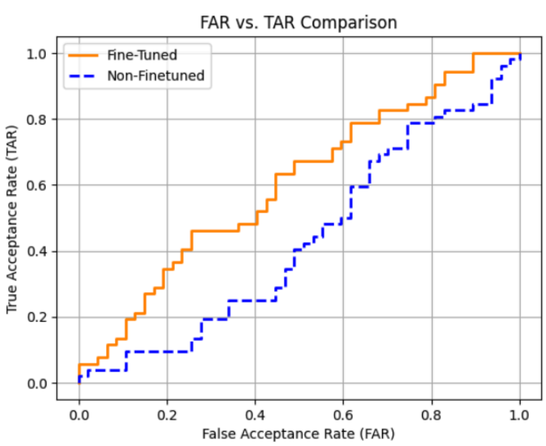


Fig 5 - FAR vs TAR Comparison



IV. Results

Overall, the fine-tuned model performs better than the non-fine tuned model on predicting the political alignment of text segments but there is still room for improvement. As seen in the distribution (Fig 3) of predicted probabilities (Democrat = 0, Republican = 1), the non-fine tuned model predicts a neutral political sentiment on the segments of text, with a slight skew toward leaning Republican. The AUC of 0.433 (Fig 4) on the regular base model confirms that this model provides no additional value - if not worse predictions than randomly guessing. The fine-tuned model outperforms the regular model by a good amount, but is still far from perfect as highlighted in the FAR vs TAR comparison (Fig 5). Furthermore, the fine-tuned ROC of 0.6043 (Fig 5) is better than a random guess, but still far below the ideal threshold of 0.9. It is relatively unbiased, but has less confidence when making predictions about Republican content.

V. Conclusion

While the performance improvement indicative of a successful training set - it failed to reach a level of performance necessary for deployment. In its current state, the model has a non-zero propensity of steering users towards a biased decision. If one wanted to improve our model, they could use a larger training set that had additional training examples that were more ambiguous and less direct. Furthermore, with additional hardware one could train a larger model architecture with higher parameter count - allowing for a more complete encoding of this high dimension space. If issues continue to persist, one could try switching to an information retrieval system. Rather than classifying user input with a specific political party, this retrieval system could pull up relevant documents and articles related to the prompt, similar to an ai agent.

Works Cited

1. Jones, Bryan. “United States.” *Comparative Agendas*, www.comparativeagendas.net/us. Accessed 20 Dec. 2024.
2. “Google-Bert/Bert-Base-Uncased · Hugging Face.” *Google-Bert/Bert-Base-Uncased · Hugging Face*, huggingface.co/google-bert/bert-base-uncased. Accessed 20 Dec. 2024.
3. “Distilbert/Distilbert-Base-Uncased · Hugging Face.” *Distilbert/Distilbert-Base-Uncased · Hugging Face*, huggingface.co/distilbert/distilbert-base-uncased. Accessed 20 Dec. 2024.
4. Datasets, Legacy. “Legacy-Datasets/Wikipedia · Datasets at Hugging Face.” *Legacy-Datasets/Wikipedia · Datasets at Hugging Face*, huggingface.co/datasets/legacy-datasets/wikipedia. Accessed 20 Dec. 2024.
5. Bookcorpus. “Bookcorpus/Bookcorpus · Datasets at Hugging Face.” *Bookcorpus/Bookcorpus · Datasets at Hugging Face*, huggingface.co/datasets/bookcorpus/bookcorpus. Accessed 20 Dec. 2024.
6. “MacBook Pro (16-Inch Late 2019) Benchmarks.” *Geekbench Browser*, browser.geekbench.com/macs/macbook-pro-16-inch-late-2019-intel-core-i9-9880h-2-3-ghz-8-cores. Accessed 20 Dec. 2024.