CIS 522: Political Polarization in the United States

Sally Hu, Ken Schumacher and Aaron Yu

1 Abstract

With President Biden approaching his first 100 days in office, the United States is still very much divided politically. In this paper, deep learning models are used as the primary tools to study the root cause of this polarization. Specifically, the paper examines tweets of US politicians to study whether the divide is caused primarily by ideological differences or by rhetorical strategies implemented to appeal to their supporters. The exploratory data analysis (EDA) process shows the ideological differences between the two parties judged by the most common words used. Then, four non-deep learning models are implemented as baseline. A tf-idf feed forward network, LSTM and Bidirectional LSTM are implemented as deep learning baselines. BERT is used as an advanced deep learning model in our study. As an extension, we use a GPT-based model to generate "tweets" given certain political primers. Overall, we find that there are elements beyond ideologies that differentiate the two parties, which suggests that rhetoric has played a role in the increasing polarization of the United States.

2 Introduction

Over the past few decades, the United States has become progressively more and more politically polarized. Taking healthcare as an example, Republicans in the past four years used every opportunity to repeal Obamacare. Normally people might view bipartisanship as a good democratic process as there are more diversity of thought. However, this has become more than policy disagreement. Increasingly, we see politicians and voters on both sides using the rhetoric of "saving America," as if the other side is inherently evil. The underlying causes are complicated, but a major factor is attributed to the proliferation of social media usage over the years. People can quickly share their thoughts and opinions on a given topic, regardless of whether or not it is factual or complete, and it can quickly reach and influence thousands or millions of others. Most notably, those with clout, especially politicians, leverage this power to their advantage and to their opposition's detriment.

Our aim is to analyze tweets from U.S. politicians to gain insight into their use of polarizing rhetoric that isn't limited to the differences in their ideologies. That is, we want to study the features that the algorithms use to differentiate the two sides. The assumption we make here is that the tweets will consist of two features only: ideology and rhetoric. The analysis work will be a two step process. First, we will run our models using all the words in the tweets. This step checks how well the models can differentiate the parties through ideologies and rhetoric. Then, we will remove the common words used by one party but not the other. The goal is to eliminate the influence of ideology on the prediction process. So, the second step would be to develop models without the key ideological words from both parties. we want to check whether the models are able to differentiate the parties purely based on rhetoric. Comparing step one and step two, we want to make some inferences about how the rhetoric used can contribute to the increasing polarity in the nation. For example, if the step one and step two model have similar performances, then this implies that rhetoric is difference between the two parties. More detail will be provided in the Results section.

Our hypothesis is that politicians on both sides of the spectrum tend to rather focus on different topics when they approach their base. Beyond ideology (with ideological words removed), they tend to use similar rhetorical devices to unite their supporters, and they will both generally be guilty of using incendiary rhetoric. For example, Democrats might appeal more to one's identity and invoke some sort of morality within each person. They might use words that are more closely associated with one's race, ethnicity, gender, etc. Perhaps their tweets will be more based on evidence and facts. On the other side, Republicans might appeal to one's sense of patriotism and nationalism more. Perhaps they will use terms like "freedom" more often in their tweets. Their tweets might be shorter as they appeal to one's feelings more than facts. We'd like to find out how they are different from the Democrats besides the ideology. Besides those already discussed, other inductive biases might include popularity of the tweet (as measured by retweets) and the hashtags the politicians used. Overall, our hypothesis is that ideology differentiates parties, but their rhetoric remains the same. We expect the full tweet models (step one) to have reasonable predictive power. For the ideology-removed tweets, we expect the model to have very little predicative power as we believe both sides use incendiary rhetoric to attack the other side and attract their base. This in turn causes the increasing polarization we see in the nation currently.

3 Related Works

There have been many works that analyze rhetoric from politicians though not necessarily by utilizing machine learning models. There have been studies that analyze the way politicians present themselves on social media to appeal to their base, specifically by leveraging rhetorical ethos [1]. This form of communication and its frequent and instantaneous nature is revolutionizing campaigns and the political landscape in the United States. "As new media evolves, the rhetorical styles evolve [and] the connection gaps between rhetor and audience start to shrink" [2]. This "shrinking" of the gap was best leveraged by Donald Trump in his 2016 campaign in which he successfully riled up his supporters in a more personal manner [3].

There has also been work conducted related to political rhetoric on Twitter specifically. In her paper, Russell finds that Republican Senators "are more likely to name-call their Democratic opponents and to make expressions of intraparty loyalty" [4]. Other analyses focused on the sentiments of the tweets and the strategies leveraged to convince voters to vote a certain way: for example, it appeared that politicians in the 2018 Midterm election who were "in competitive races, losers, women and Democrats were more likely to use anxious, sad, and angry words in their tweets" [5].

Needless to say, the strategies employed by politicians via Twitter are numerous and varied. At one time, platforms such as Twitter was used to simply broadcast messages; now, they are more so used by politicians to personally appeal to voters[1]. For instance, we can see that more people are loyal to Donald Trump than to the Republican party; similarly, we can also see that more people are loyal to Bernie Sanders than they are to the Democratic party. As the use of social media as a communication platform to discuss politics continues to grow, the public needs to be better equipped so they can more properly interpret messages from politicians.

4 Dataset & Features

4.1 Dataset

The dataset we used consists of over 1.2 million tweets prior to June 2017 from U.S. politicians including the President, members of Congress and governors at the time. Each tweet is labeled with the corresponding politician's party and gender. Independents were relabeled as Democrats since they typically caucus with

[1]Madestam, J. and Falkman, L.L. (2017), "Rhetorical construction of political leadership in social media", Journal of Organizational Change Management, Vol. 30 No. 3, pp. 299-311. https://doi.org/10.1108/JOCM-10-2016-0204

[2]Johnson, Janet. (2020). Political Rhetoric, Social Media, and American Presidential Campaigns: Candidates' Use of New Media.

[3]Ross, AS, Rivers, DJ. Donald Trump, legitimisation and a new political rhetoric. World Englishes. 2020; 39: 623– 637. https://doi.org/10.1111/weng.12501

[4]Annelise Russell. 2018. US senators on Twitter: Asymmetric party rhetoric in 140 characters. Am. Pol. Res. 46, 4 (2018).

[5]Gervais B.T., Evans H.K., Russell A. (2020) Fear and Loathing on Twitter: Exploring Negative Rhetoric in Tweets During the 2018 Midterm Election. In: Foreman S., Godwin M., Wilson W. (eds) The Roads to Congress 2018. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-19819-0˙3

them and one Libertarian was removed for the sake of the analysis. The breakdown of the dataset by party and gender is as follows:

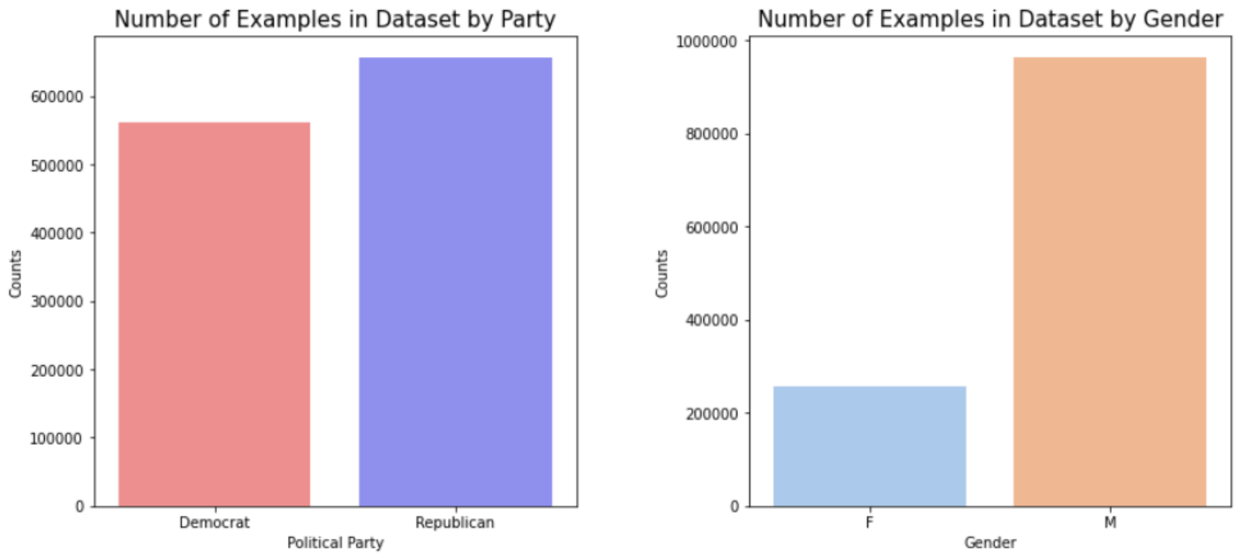


Figure 1: Breakdown of dataset by political party and gender

The dataset was fairly evenly split between Democrats and Republicans, but was highly skewed between females and males. For the scope of this analysis, we only focused on differentiating rhetoric between parties.

4.2 Exploratory Data Analysis

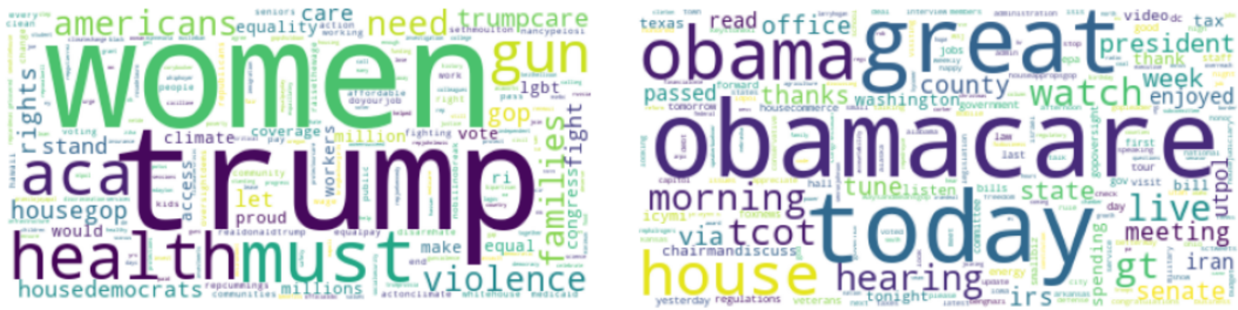
To get a sense of the differences between each political party, we first explored the tweets. The words used most frequently by each party show that they are similar at a high level.

Democrat

Republican

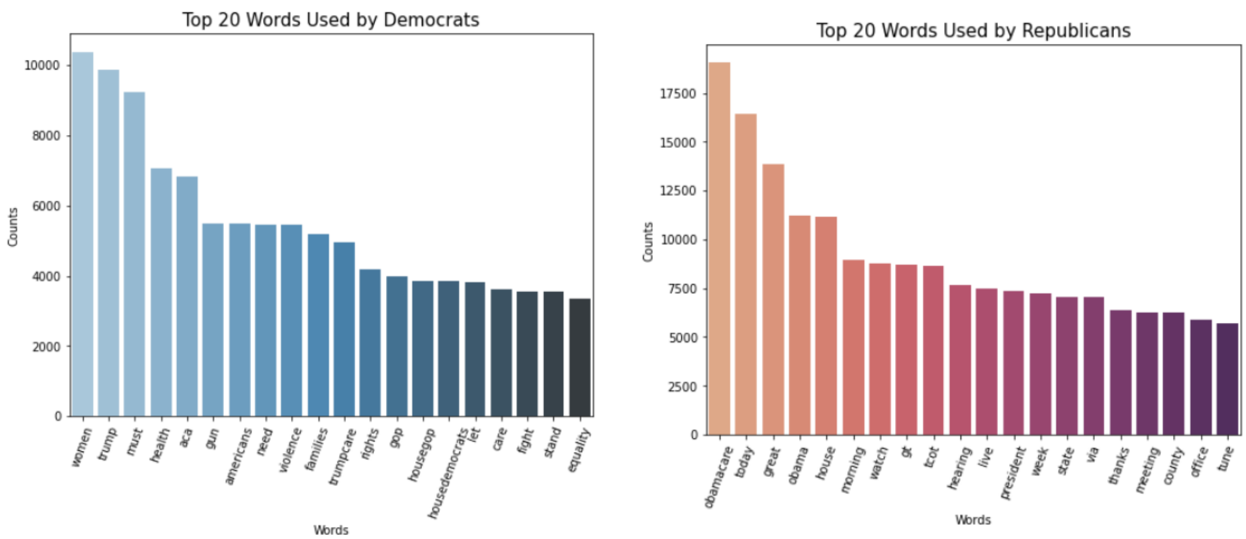
Figure 2: Top 20 most common words used in tweets by both parties

However, to get a better sense of the words that actually differentiate the parties, we also obtained the words that were most commonly used by each party relative to the other. The results are much more telling and it is more evident that the two parties generally focus on differing issues.



Democrat Republican

Figure 3: Top 20 most common words used by each party relative to each other



Democrat Republican

Figure 4: Frequency of top 20 words relative to the other party

4.3 Preprocessing and Feature Extraction

Based on the exploratory data analysis, we decided not to only use the tweets themselves but also the tweets with the 400 most common words (relative to each party) removed from them, 200 from each party.

We also preprocessed the tweets by lemmatizing it, removing stopwords, casting to lowercase, removing numbers, etc. We used the preprocessed tweets for all of our models except for BERT.

Aside from preprocessing the tweets, we also experimented with different features including:

- **TF-IDF**, which measures how important a word is within the collection of tweets.
- **Numerical representation of the text itself**, in which we truncated the length of each tweet to a maximum of 20 words and mapped each word to a number.
- **Word embeddings**, which we extracted from our corpus using Word2Vec.

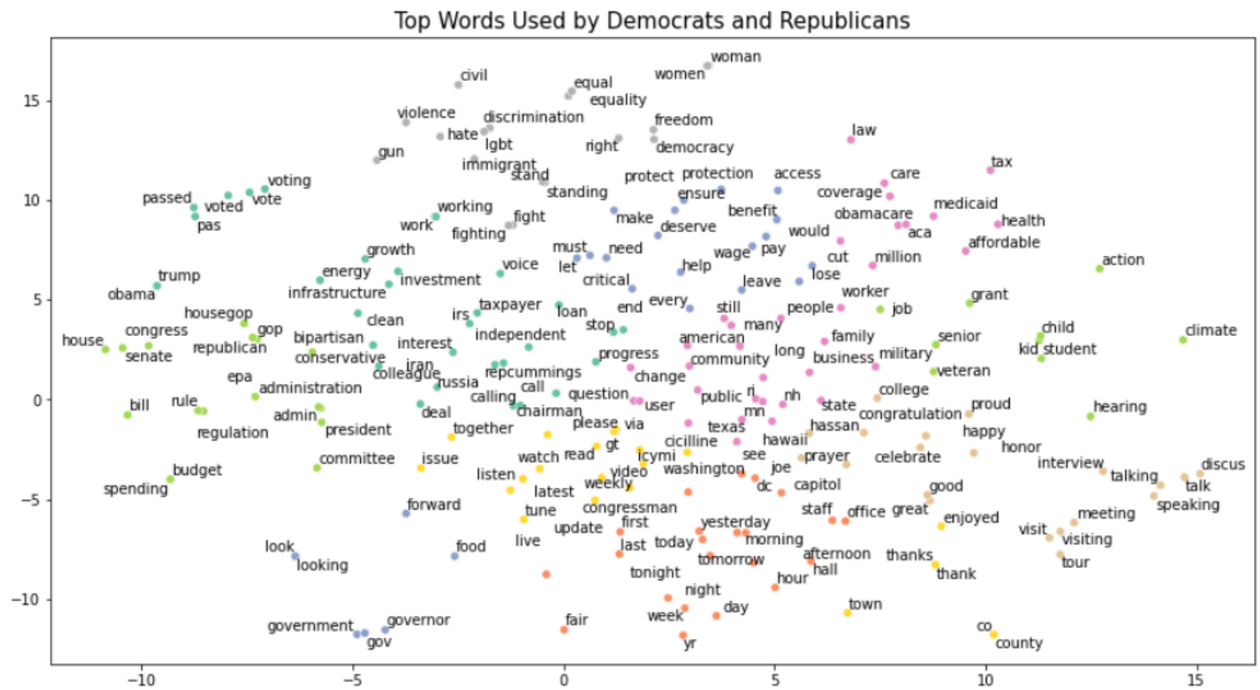


Figure 5: Word embeddings obtained from the tweets

5 Experiments

5.1 Methods

To explore our hypothesis, we wanted to see if there are differences in political rhetoric across party lines that transcend ideology. Consequently, we ran our models once on the preprocessed tweets with all words included and once on the preprocessed tweets with the top 200 most common words from each party removed. The idea is that the top most common words relative to the the other party will largely capture the words that reflect a party's stances and priorities, and removing them will provide us with a way to see if political rhetoric is still different between parties. Additionally, due to computational limitations, we randomly sampled 300k tweets from each party and worked with a dataset of 600k tweets instead.

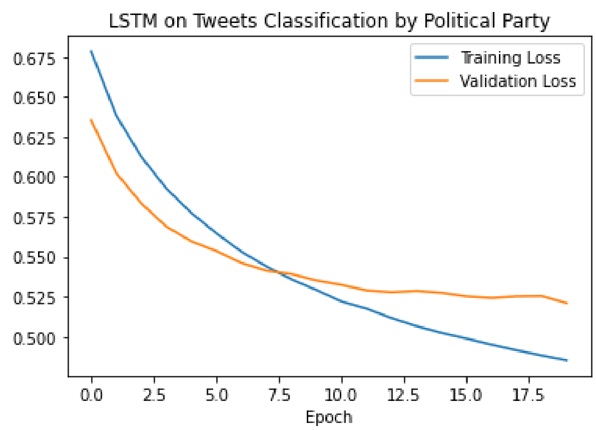
The architectures that we believe were most appropriate for this task were recurrent neural networks (RNN), namely Long Short Term Memory (LSTM) networks since they can capture long-term temporal dependencies and context. We also leveraged a bidirectional LSTM and hypothesized that it would perform even better than the normal LSTM because it would capture complete, sequential information from the words before a given word and also those after it. Finally, we also hypothesized that using BERT would elicit better results since it provides a contextualized embedding for a given word depending on the context in a given sentence. It is also heavily pre-trained to handle many NLP tasks and we figured it could be easily fine-tuned to accommodate our objective.

5.1.1 Non-Deep Learning Benchmarks

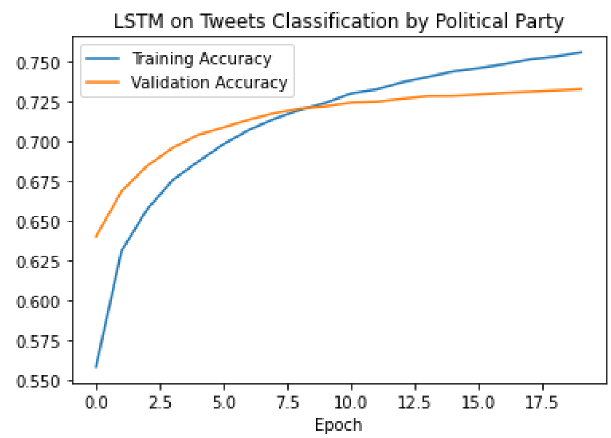
To begin, we implemented four non-deep learning models: Random Forest, Logistic Regression, Gradient Boosting and Naive Bayes. Each model was run on the tweets with all words and tweets with no common words with different features extracted: tf-idf, the numerical representation of the tweet, and averaged word embeddings.

5.1.2 Base Models: FNN, LSTM & Bidirectional LSTM

Our deep learning baseline models include a feed forward neural network (FNN), LSTM and bidirectional LSTM. We started with a FNN to test the effect of tf-idf embedding. The best performing FNN we tested has 3 fully-connected hidden layers, each with size 16. ReLU is used as the activation function. For LSTM, we started with the basic model in PyTorch and then added additional method for improvement. The best model we constructed has the following architecture. The input to the LSTM is the word embeddings of length 300. There is first a dropout layer with parameter 0.2 for regularization purposes. Then, the inputs are fed into the LSTM. The 2 hidden layers of the LSTM are of size 16 each. For the final prediction step, we used a linear layer to convert the LSTM results to the political parties label with LeakyReLU as our activation function. The model parameters are initialized with Xavier initialization and we used cross entropy as our loss function. We will address the model building process shortly. For the last baseline model, we used a bidirectional LSTM. The intent is to improve the model's memory by considering context in both directions. This can be useful for long tweets as the standard LSTM tends to forget previous content entirely after about eight steps. The structure, incorporated regularization and initialization process of the bidirectional LSTM is very similar to the LSTM.

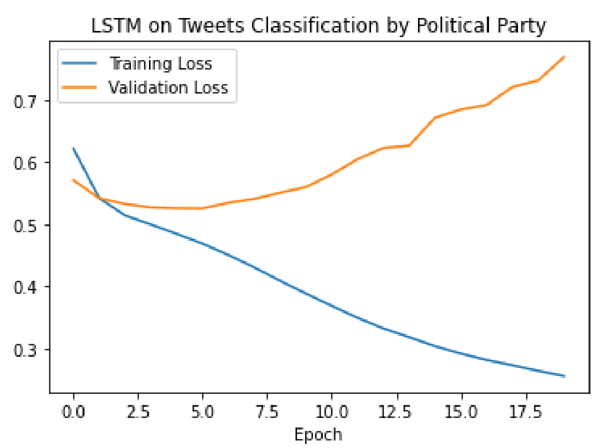


Loss

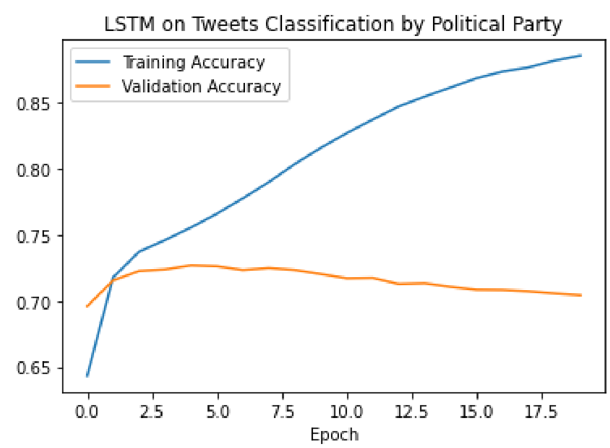


Accuracy

Figure 6: Loss and Accuracy Plots using Dropout



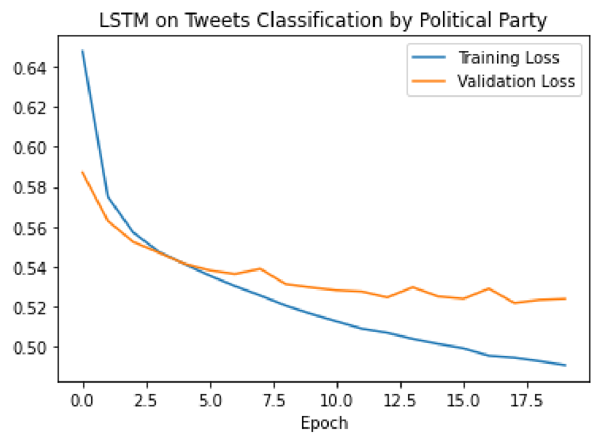
Loss



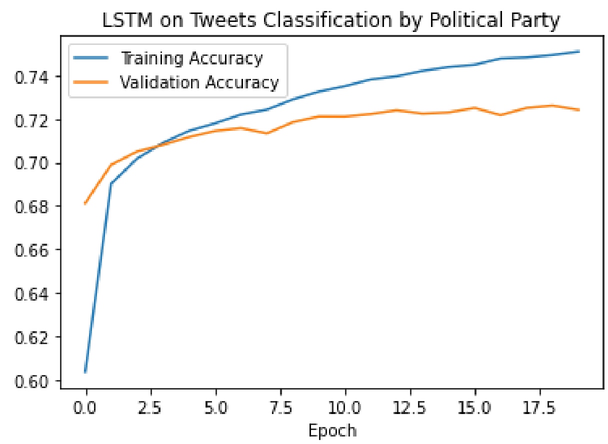
Accuracy

Figure 7: Loss and Accuracy Plots using L2 Weight Decay

We will use LSTM with text to illustrate the model building process. There are 6 areas we considered: regularization, initialization, optimizer, criterion, architecture, and hyperparameters. First, we considered regularization. This is an important consideration for us because it is easy for the models to overfit on some words/features and thus not generalize well. We compared 2 techniques: Dropout and L2. Dropout forces the model to work with less weights while L2 kept the weights small. After testing, the best dropout rate is found to be at 0.2 (Figure 6). We compared that to the L2 regularization with various weight decay parameters (Figure 7 has weight decay of 0.001). We found that L2 regularization still tend to overfit regardless of weight decay size. Our training accuracy kept going up to above 85 percent while the validation accuracy decreased back to 70 percent, which is achieved very early in training. We tried to using a combination of both but found no improvement over the dropout only models. For parsimony, we chose dropout as our regularization method.



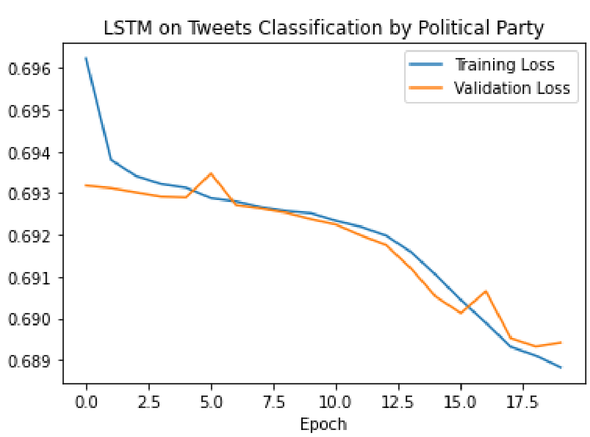
Loss



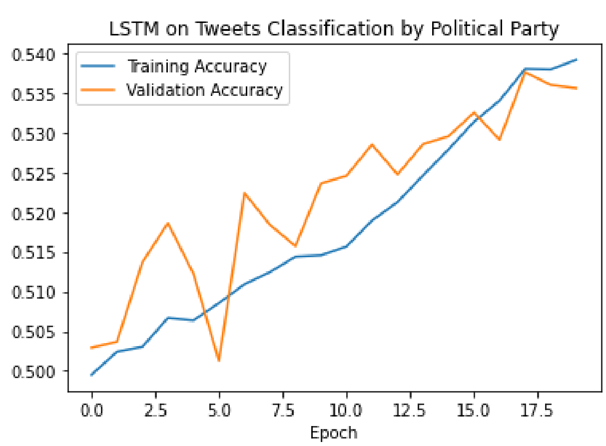
Accuracy

Figure 8: Loss and Accuracy Plots using Xavier Initialization

After that, we considered initialization. In particular, we considered Xavier initialization on a Gaussian distribution (Figure 8). After experimentation, we found the best gain parameter to be 1.4, close to the theoretical value. This is an important step as it can prevent the model from being stuck in local minimum to some degree. It can avoid numerical instabilities such as exploding gradient problem or the vanishing gradient problem. Compared to the standard normal initialization, Xavier tend to converge faster for large scale applications. In our study, we found that Xavier initialization is substantially better. The training accuracy and validation accuracy goes up faster. Perhaps this is because the model started on a steeper path in gradient descent. Xavier normal also had a better performance compared to Xavier uniform. So, we chose Xavier normal as our initialization technique.

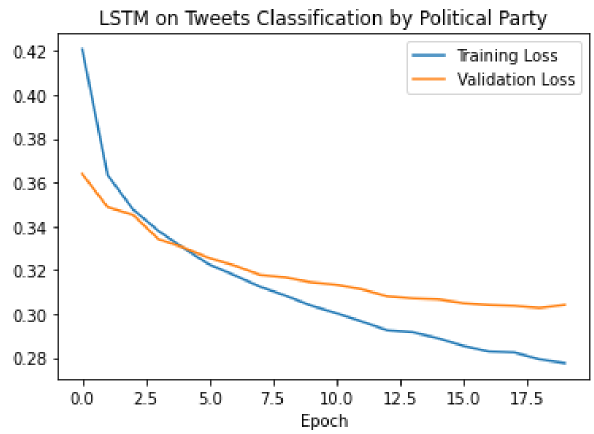


Loss

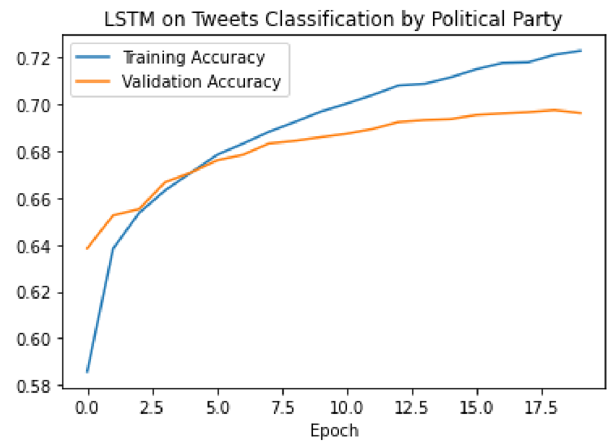


Accuracy

Figure 9: Loss and Accuracy Plots using SGD



Loss

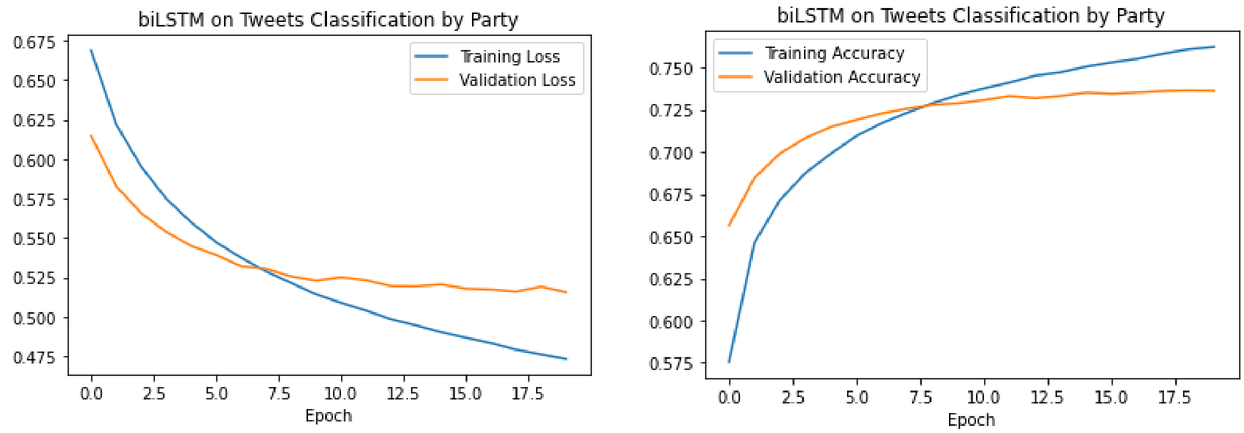


Accuracy

Figure 10: Loss and Accuracy Plots using Multi Margin Loss

For the optimizer, we experimented with SGD and Adam. One noticeable difference between the two is that SGD tend to have slower loss decline with more fluctuations (Figure 9). This is expected as Adam has the momentum combined with the exponentially decaying average of gradient squared, similar to RMSprop. This can help Adam to go down the loss curve as if it is a heavy ball with friction. Together, momentum and friction make Adam prefer flat minima in the loss surface, increasing the efficient of training. After training for 20 epochs, Adam is significantly better in terms of loss and accuracy. For the criterion, we tried Cross Entropy Loss and Multi Margin Loss with a margin of 1 (Figure 10). Cross Entropy Loss has significantly better results. After the accuracy reaches a steady state, the Cross Entropy Loss is able to achieve around 5 percent more compared to Multi Margin Loss. So, we will use Cross Entropy Loss moving forward.

For model architecture, we experimented with the number of hidden layers in LSTM as well as the activation function. Higher number of layers in the hidden units caused the training and validation accuracy to start higher. However, the rate tended to not reach as high if we had used a simpler model. We found that two hidden layers seemed to work best. For the activation function, LeakyReLU with a negative slope of 0.1 seems to outperform other activation functions. Compared to ReLU, LeakyReLU converges faster. Validation accuracy fluctuated a little more in training compare to ReLU, but the end test accuracy was higher. Lastly, we tuned the hyperparameters. For example, we experimented with learning rate and input embedding size. Overall, we found the best learning rate to be 0.0003 and the best embedding size to be 300.



Loss

Accuracy

Figure 11: Loss and Accuracy Plots for the Best DL Baseline Model

To summarize, we first found that dropout performs better than L2 weight decay. Then, we improved out initialization with Xavier initialization. Afterwards, we tested various optimizers, activation functions, loss functions and hyperparameters. We repeated roughly the tuning process for other models. The experimentation process for other models are shorter as we have an understand of the impact of six aspects of deep learning. Figure 11 shows the training output for the best deep learning baseline model, as bidirectional LSTM in the beginning of the section. The training process is smooth and there is not too much overfitting. For more detail about the tuning process, please refer to the code notebook. The best model outputs will be shown in the Results section.

5.1.3 Advanced Model: BERT

We utilized two pretrained BERT models as advanced models to complete the objective detailed earlier. Research has repeatedly shown that using a pretrained BERT model on a downstream task, such as our goal of predicting political party given a tweet, often can result in high performance[6]. Due to the constrained computational resources available in Google Colab, we opted to use the pretrained BERT-Mini and BERT-Small model provided by Google through HuggingFace. We had no reason to expect the BERT-Mini model to outperform the results of BERT-Small - it would be odd for a smaller model with less parameters to do so - we wanted to run the results in parallel to visualize any performance increase in using a larger BERT model. Doing so may give insight to how performance is constrained by the size of the BERT model used.

The BERT models used were trained with $L = 4$ stacked encoders. The BERT-mini model used a hidden size of $H = 256$ while the BERT-Small model used a hidden size of $H = 512$. While we were satisfied with the results from these models, we recognize slightly higher performance metrics may be attainable using larger BERT models in more resource-rich environments. Tuning the BERT-Mini and BERT-Small models was a computationally expensive and long process (training a single epoch for BERT-Mini took about 19 minutes). We decided to set the number of training epochs at 3 to align with our constraints, a value that is consistent with that often used with BERT models in similar online contexts. Further, we used a standard training batch size of 16 and a learning rate of $1e - 5$. Unlike with the previous models, we ran our BERT models directly on the original tweets (uncleaned version) with and without common words removed, and compare the performance of the two in the Results section.

5.1.4 Extension: Tweet Generation

As an extension to our discriminator models, we attempted to "generate tweets" using the DistillGPT-2 model. As the name implies, the model is a distilled version of GPT-2, a large-scale transformer-based

[6] Turk, Iulia. "Well-Read Students Learn Better: On the Importance of Pre-training Compact Models" https://arxiv.org/abs/1908.08962

language released by OpenAI. The GPT-2 model was built to predict the next word given a set of preceding text and was trained on an immense dataset of 8 million webpages. In order to produce better and more relevant results, we decided to fine-tune the distillGPT-2 model.

Which dataset to use in fine-tuning was a difficult decision; at first, we looked for a dataset trained on opinions (e.g. news articles or another dataset using tweets). While this strategy would produce "tweets" that were more opinionated, we found the DistillGPT-2 model to lack political context necessary to make the produced statements coherent. We then utilized the "wikitext-2-raw-v1" dataset from HuggingFace that was sourced from Wikipedia articles, and that method produced superior results. While this strategy results in more narrative-style statements than those generally found on Twitter, we opted to accept this shortcoming in return for the increased context and coherence the Wikipedia fine-tuning offered.

After tokenizing, training and fine-tuning, we provided various primers for our DistillGPT-2 model to complete. Some of these primers were meant to represent popular political issues (e.g. "Climate change is") while others were centered around a certain individual or nation (e.g. "Donald Trump was a"). Interesting cases and examples are covered in the following section. We limit the length of GPT-generated statements to 280 characters, consistent with Twitter's current word limit.

5.2 Results

The overall results are as follows. Only the best performing non-deep learning model, baseline deep learning model and BERT models are presented here.

Model	**Tweets: All Words Dev Accuracy**	**Tweets: No Common Dev Accuracy**
Random Forest - Embeddings	0.64	0.61
Bidirectional LSTM - Text	0.735	0.69
BERT-Small	0.795	0.779

Table 1: Overall Results

5.2.1 Non-Deep Learning Benchmarks

No matter the features used, Random Forest and Logistic Regression consistently outperformed the Gradient Boosting and Naive Bayes models. We will therefore analyze the results generated by these two models on the development set.

			Tweets: All Words		**Tweets: No Common**	
#	**Model**	**Features**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
1	Random Forest	tf-idf	0.630	0.686	0.570	0.602
2	Logistic Regression	tf-idf	0.620	0.682	0.580	0.609
3	Random Forest	Text	0.630	0.686	**0.630**	**0.683**
4	Logistic Regression	Text	0.500	0.574	0.500	0.570
5	Random Forest	Word embeddings	**0.640**	**0.703**	0.610	0.662
6	Logistic Regression	Word embeddings	0.630	0.685	0.600	0.642

Table 2: Results from Random Forest and Logistic Regression models

Overall, the Random Forest model performed the best. Interestingly, when using just the text as features, the

Random Forest model found basically no difference between whether or not common words were removed. However, this is not the case in every other scenario. Removing the common words negatively impacted the accuracy and AUC scores, but not by as much as we had expected. Preliminary results suggest that there might be actual rhetorical differences in the way Democratic and Republican politicians tweet without accounting for ideology.

5.2.2 Base Models: FNN, LSTM & Bidirectional LSTM

#	Model	Features	Tweets: All Words		Tweets: No Common	
			Training Accuracy	Test Accuracy	Training Accuracy	Test Accuracy
1	FNN	tf-idf	0.662	0.639	0.563	0.558
2	LSTM	Text	0.732	0.731	0.574	0.562
3	Bidirectional LSTM	Text	**0.762**	**0.735**	**0.692**	**0.690**
4	LSTM	Word embeddings	0.676	0.650	0.523	0.506
5	Bidirectional LSTM	Word embeddings	0.751	0.690	0.615	0.553

Table 3: Results from Baseline Deep Learning Models

The bidirectional LSTM produced both the best training and testing accuracy for tweets with all words as well as with word embeddings from Word2Vec. However, the performance improvement (for all words) in test accuracy for tweets with all words is not very significant compared to LSTM. One potential reason is the tweets are relatively short. Some tweets just have a couple words left after we removed the ideological differences. This limits the advantages of bidirectional LSTM has the regular model can retain the earlier content as well. When we removed the common words, the bidirectional LSTM works considerably better compared to others. Our conjecture is that rhetoric is better captured by processing the text from both directions. Overall, we will select the bidirectional LSTM as our best baseline deep learning model.

5.2.3 Advanced Model: BERT

We found it unsurprising that the BERT model performed the best given our classification task. BERT models are differentiated in their ability to contextually understand input sentences; we believe this advantage may have allowed the model to better understand rhetoric, leading to higher marks. Extensive pretraining on a huge corpus of data also likely provided a boost in accuracy. Interestingly, the BERT models performed far better on the uncleaned data instead of the cleaned data; perhaps this is because the model is able to better use contextual clues originating from features such as punctuation in a way that other models are not. Interestingly, the performance differences between BERT-Mini and BERT-Small were minimal, suggesting that performance is not hindered by the size of the pretrained BERT model utilized. The high performance of both models after removing common words in tweets further suggests validity in our hypothesis that rhetoric provides power in predicting the party behind an individual tweet.

#	Model	Tweets: All Words		Tweets: No Common	
		Training Accuracy	Test Accuracy	Training Accuracy	Test Accuracy
1	BERT-Mini	0.808	0.795	0.754	0.744
2	BERT-Small	0.807	0.795	0.792	0.779

Table 4: Results from BERT Models

5.2.4 Extension: Tweet Generation

"**The United States of America i**s attempting to build a wall along the Mexican border, despite a petition from some supporters claiming the wall is necessary to keep the country secure"

"**President Barrack Obama was a** brave man and dedicated patriot of the Republic. He did what he was willing to do to preserve his national security. He was prepared to do just that."

"**President Donald Trump was a** leader of the Republican Party. The President's views about women's rights and reproductive rights are varied, and have changed with the presidential election."

"**The US government is** preparing a comprehensive strategy to fight climate change, which will include a national and multilateral strategy. The Joint Typhoon Warning Management and Control (JMT) is a multi-coordinated effort aimed at managing a global warming related threat."

Figure 12: Generated "tweets" from GPT model

Above is a small sample of "tweets" generated by the DistillGPT model described in the previous section. Note that a small minority of tweets produced by the model were coherent; many either quickly digressed or included factually incorrect statements. However, as seen above, the GPT model was able to produce statements that sound reasonable and could be imagined on Twitter. As noted previously, one limitation of our political-tweet generation model is that it was fine-tuned on Wikipedia articles, most of which are written in an objective and informational form. As a result, the produced tweets are more descriptive and less opinionated than those that would likely be found on Twitter. It would be more difficult to produce an "opinionated" GPT model, say one that sided with Democrats or Republicans. One may be able to do so by fine-tuning the model on tweets only by a certain party, which may provide context on party-specific views around certain issues.

5.2.5 Discussion of Results

It is evident that using more complex models elicited better results (differentiating Democrats and Republicans using tweets). Interestingly, the features that produced the best results across all models typically included word embeddings as well as the text itself.

Although our results might indicate that there are differences between political parties that transcend ideology, further analysis needs to be conducted. It is very possible that removing the top 200 most common words from both sides was insufficient and that words were left in the tweets still reflected a given party's beliefs to some degree. A minor concern is that we removed too much common words and perhaps some important rhetorical elements were removed. This is mitigated as we checked in EDA on the common words removed. Also, this does not affect our hypothesis testing as we initially believed rhetorical usages are the same from both sides.

On the other hand, it is perfectly possible that there are indeed rhetorical strategies that differ between parties. Given the "black box" nature of our models, we would further investigate the differences and pinpoint whether there are indeed true differences in rhetoric between Democrats and Republicans or if we didn't remove enough common words from the models. However, these results initially imply that there are rhetorical differences between parties.

6 Social Impact Analysis

Twitter and other social media platforms are continually playing a more prominent role in United States politics. Donald Trump's frequent use of Twitter is largely considered to have played a significant role in his success during the 2016 presidential election, and even the then-president has noted: "I doubt I would be here if it weren't for social media, to be honest with you." [7]. During the same election, millions of Facebook

[7] Baynes, Chris. "Donald Trump says he would not be President without Twitter", https://www.independent.co.uk/news/world/americas/us-politics/donald-trump-tweets-twitter-social-media-facebook-instagram-fox-business-network-would-not-be-president-a8013491.html

users had their personal information utilized without consent that was used to created targeted political advertisements [8]. Following the 2016 election the role of social media in politics only increased. Most recently, President Donald Trump was permanently banned from Twitter on January 8, 2021 for tweets that incited violence and supported rioters who broke into the U.S capital [9].

Social media (and specifically Twitter) still plays an important role in American politics and elections. Many American citizens use social media networks like Twitter as their primary news source and millions swipe through their Twitter feeds each day. Due to this high amount of influence, we find it important to help inform Twitter users how political leaders use their platform to influence the public.

Specifically, we take an interest in how underlying rhetoric of politicians is used to capture an audience. Clearly, Democrats and Republicans discuss different subjects and take opposing stances on many of those issues. Influence through this form (ideology) is obvious and easy to identify; politicians tweet attempting to rally further support and convince viewers to side with their views (and eventually vote for them, etc.). However, the work in this paper suggests that a separate vessel of influence is at work (rhetoric). This is important to identify - users are less aware of how they are attracted to different tweets and politicians through this method. Is your feed largely made up of shorter tweets with consistent calls to patriotism? Are your criteria for whether you follow someone on Twitter partly based on the rhetoric of their typical tweet, rather than the topic of the tweet itself? We hope our work reminds people to be acute for this bias while interacting and following users on Twitter. In an age where Twitter still has a significant effect on politics, we believe this is important information to uncover.

7 Conclusion

Prior to the analysis, we hypothesized that Republican and Democratic politicians mainly focus on different issues when they approach their base on social media platforms such as Twitter. Beyond their ideology, their rhetorical strategies would generally align with one another and make it difficult to differentiate between parties. However, we found that this does not appear to be the case. By identifying that rhetoric between the two parties was differentiated, we hoped to inform Twitter users in their future use of social media.

The results found support the idea that politicians across parties tweet differently. Initially, the bidirectional LSTM and BERT models were able to achieve a high accuracy when predicting party based on a full tweet from our data set. After the removal of ideology between the two parties, these models were still able to predict party with high accuracy, suggesting that rhetoric from the tweets had predictive power in terms of political party. We recognize limitations from our study; it is possible that we removed too few "common" words from the two sides, and ideology is still primarily used by the model to make predictions on party.

As an extension, we trained a GPT-2 model and provided political primers to create artificially generated "tweets". We find in the best cases, these tweets are coherent and somewhat representative of what one might find on social media. We believe that the political effects of social media usage will only become more relevant in the future; thus, it is critical to better understand how we use these technologies and how they affect us and our world views.

[8] Confessore, Nicholas. "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far" https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

[9] "Permanent suspension of @realDonaldTrump", https://blog.twitter.com/en˙us/topics/company/2020/suspension.html