Trump's Twitter Ticks

ECE324 Final Report

Claire Dalkie and Calvin Herd

December 3, 2019

Word Count: 1924

Permission to Post Video: Yes
Permission to Post Final Report: Yes
Permission to Post Source Code: Yes

Introduction

The project aims to predict the change in the S&P 500 due to tweets posted by Donald Trump. The stock market is very dependent on people's confidence in their nation's economy and can be extremely volatile due to changing confidences and outlooks. Donald Trump, the President of the United States, possibly holds the most powerful role in directing the economy. Throughout his presidency, he has consistently tweeted his general thoughts, intentions, and policy standpoints. We make use of machine learning's capabilities of understanding text and connections between words to create a neural network. With this, we will hopefully be able to identify whether a given tweet has a positive, negative, or neutral impact on the existing trends of the S&P 500, a stock market index that measures the performance of the 500 largest companies in America, across several industries.

Background and Related Work

There are two key pieces of work related to our project. One research paper out of Stanford University describes the correlation between "public sentiment" and "market sentiment" with regards to a public feed of tweets and the daily Dow Jones index [1]. The paper describes their process of passing the tweets through a sentiment detector, and then using a Self Organizing Fuzzy Neural Network to predict the Dow Jones. The paper concludes that public mood can be captured by Twitter feeds, and that the model could accurately predict the Dow Jones value after three days of calm and happy tweets. In addition, an article from Vox describes major banks' research into a potential a correlation between Trump's tweets and the stock market [2]. JPMorgan identified the most prominent words that move the markets, while the Bank of America determined that the stock market falls when Trump tweets more than 35 times and goes up when he tweets fewer than five times [2]. These related works indicate that Trump's tweets have an effect on the stock market, and that neural networks can be used to predict the correlation between tweets and stock market values.

Data and Data Processing

Label	# of Tweets before Augmentation	# Tweets after Augmentation
Negative Change	3001	8398
No Change	3739	10235
Positive Change	2615	7241
Total	9355	20874

Figure 1: Distribution of Entries in Dataset

The two data sources used for this project have been Trump's tweets and the S&P 500 index. The tweets were exported as a JSON from "Trump Twitter Archive," converted to a Pandas DataFrame, and then cleaned. This involved using the preprocessor and re libraries to remove hashtags, usernames, links, and emoticons, and to tokenize the tweets. The tweets were then vectorized, using a pre-trained GloVe model based on 2.7 billion tweets. The dataset was then augmented, using techniques such as synonym replacement and word deletion, to expand the dataset size from 9355 to 25874 tweets. In addition, the hourly values and changes for the S&P 500 were collected from barchart.com. In order to label our dataset of tweets, we found the second derivative at each hour, using the three hours prior to the tweet and three hours after the tweet. This was then converted to a one hot encoded label, with three options – positive change, no change, negative change. Finally, the tweets were labelled by searching for the label corresponding to the hour after the tweet.

Architecture

Our project uses a Recurrent Neural Network, with three layers. A gated recurrent unit that processes the batch of tweets, along with the hidden layer and two linear layers, compresses the output into an output of three labels: [negative change, no change, positive change]. We used a learning rate of 0.015, batch size of 16, and an output activation function of SoftMax. The decision to use an RNN and the SoftMax function was based on the sentiment analysis nature of our problem, that analyzed the text and produced a multi-class classification output, with probabilities that summed to 1. The other hyper parameters such as number of layers, dimensions of layers, and batch size were determined through several rounds of testing.

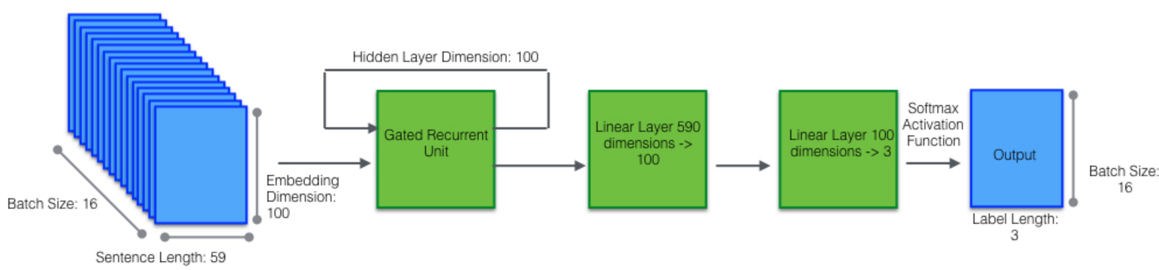

Figure 2: Structure of the Model

Baseline Model

The Baseline Model has been created classically. This has been done by finding articles that discuss the effect of Donald Trump's tweets on the stock market and identifying the key words that are most relevant to the tweet. Frequency analysis was also conducted, where we identified the words Trump used most often, to improve the baseline model's breadth. The classical model searches through the tweets for the identified key words and makes predictions based on the label associated with the phrase. This has had moderate success; however, it is limited by its inability to identify the context of a tweet, and the fact that it can only produce effective labels for tweets that contain at least one key phrase. In addition, there aren't many tweets that contain these obvious key words, so this model has a heavy bias towards predicting no effect on the market.

Quantitative Results

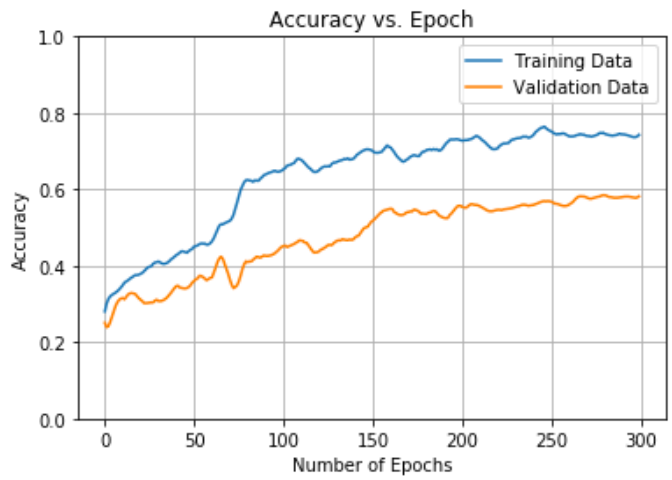

Figure 3: Accuracy vs. Epoch

	Max Value Reached
Training Accuracy	74%
Validation Accuracy	58%
Validation Accuracy	54%

Figure 4: Accuracy Results

Given the multi-class nature of our problem, a random guesser would be able to achieve 33% accuracy. Thus, all the accuracies achieved by the model demonstrate relative success. After running for 300 epochs we overfit our model slightly, as we achieved a higher training accuracy than a validation or testing accuracy.

Metric	Result
Accuracy	47%
Specificity (Test Set, Up vs. Down, Final Epoch)	73.5%
Sensitivity (Test Set, Up vs. Down, Final Epoch)	60.5%
Total Precision for Up	44.3%
Total Precision for Down	40.6%

Figure 5: Summary of Metrics

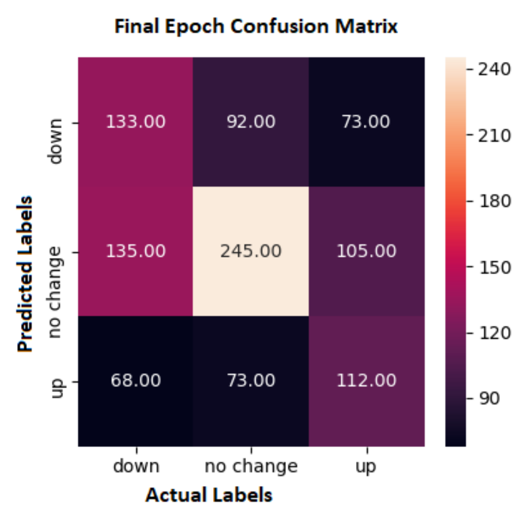

Figure 6: Confusion Matrix

We also calculated some additional metrics to further analyze the results of our model, summarized in the table in Figure 5. The precision for "up" and "down" is counted by the number of times "up" or "down" was identified correctly versus the number of times the model guessed "up" or "down" and it wasn't the case. Specificity equals the number of correct guesses at down versus incorrect guesses of up when the actual was down. Sensitivity was the natural opposite of that, in the correct guesses of down versus the incorrect guesses of down when the actual was up. Furthermore, the confusion matrix is very useful in identifying where the model makes the most errors, and what general predictive trends the model adopts.

Qualitative Results

In Figure 7, we display two sample inputs to our model with their associated outputs. The model was able to pick up on a few key trends, such as vocabulary used to determine output. This led to some successes, as shown in the first tweet. This tweet features some key terms that typically would indicate a downward movement in the stock market, such as “trade war.” The model correctly identifies with 84% confidence that the stock market went down as a result of this tweet. The second tweet, however, showcases where the model failed. In this tweet, the sentiment seems positive, using terms such as “money” and “technology.” Unfortunately, the actual result of this tweet was negative, due to the fact that this tweet was a complete lie. The “new” factory Trump describes was actually built in 2013. This exemplifies an issue the model ran into where it was tasked with properly guessing people’s reaction to a tweet, rather than predicting general sentiment, as even a positively worded tweet might cause very negative reactions.

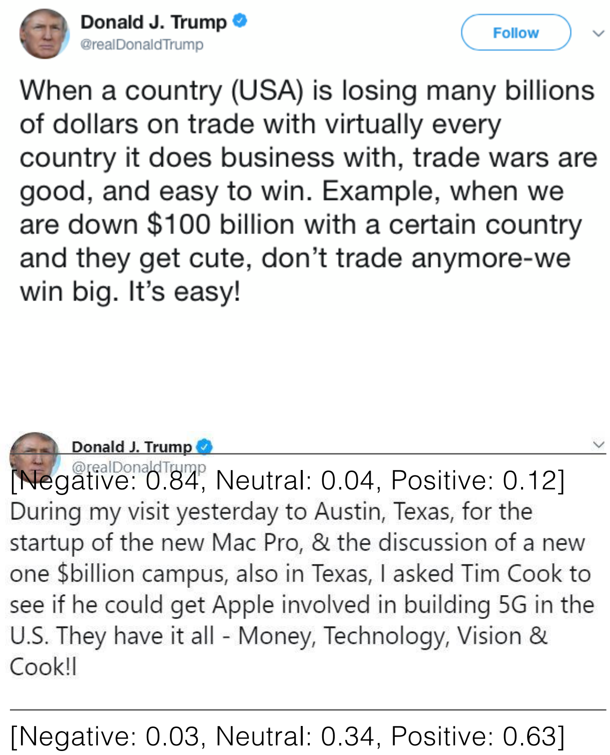

Figure 7: Sample Results

Discussion and Learnings

Ultimately, the model did not perform as well as we would have liked. However, we believe it’s current set-up of hyperparameters and structure were optimized to the best of their ability. We found that the model over-emphasized predicting “no change.” We see this as a positive, as we would much rather make an incorrect prediction which would not be actively detrimental to a person choosing to make a trade. This pattern is demonstrated by the difference in precision versus specificity/sensitivity. The net precision was low, but the model, when making a claim of up or down, was correct a net total of 63.5% of the time. This showcases some progress in its predictive capabilities that would not be inherently obvious when viewing the raw accuracy.

As discussed in the Qualitative Results section, the largest difficulty the model faced was properly identifying the reaction people will have to a tweet. The issue remains that the exact same tweet, stated in a different political climate or at a different time of year, could invoke completely opposite reactions from the market. The model was tasked with finding these patterns and making connections outside of the context of the tweets. It was unable to do so effectively, beyond keyword identification and potentially some sentiment analysis. After more analysis into the effect of Trump's tweets on certain stocks, we realized the model essentially needed the qualities of a lie detector. Furthermore, we faced the issue of distinguishing between a genuine shift in the stock market and a typical market fluctuation. We attempted to select the best metric of the stock market to capture the immediate change caused by a tweet, but we don't think it was completely accurate. Thus, this distinguishing task was passed along to the model, which further complicated the problem and confused the results.

We have learned a lot throughout the project, and would make some changes if we were to create another similar project. Our biggest change would be to the data labelling. We would do far more research on the actual effect of the tweets on the market, in terms of which stocks were impacted, at what time, and in what way. In doing so, we would be able to establish a clear correlation between our input and labels, and avoid our prevalent issue of discerning whether the model was not functioning, or that the dataset was simply not good enough for proper training. Furthermore, we would consider what additional tweets or data could be added to supplement the dataset that could potentially help the model understand the context of the situation.

Overall, considering the high level of difficulty in the assignment, we were glad to make some progress, but we were disappointed in the fact we were not able to make a workable, publishable model.

Ethical Framework

We consider the ethical context of our project with respect to the Reflexive Principlism.

Theoretically, Trump could use our model to adjust his tweets, either in content, tone, or timing, so that they impact the economy in a less negative way, applying nonmaleficence to his actions. Furthermore, he could act in a way that results in beneficence, as his actions could positively affect the stock market, and the millions of people dependent on how it does. However, given his track record of writing whatever he wants, he would most likely maintain his right to autonomy and write with disregard to our model. Politicians close to Trump could use the model to try to sway Trump's actions in either direction of nonmaleficence, by provoking his emotions, suggesting policies, or delivering the advice of the model. Furthermore, if they have insider knowledge of what he will tweet, they could use the model to see how the market will react, and potentially engage in maleficent insider trading.

Anyone could use our model to guide decisions when trading or automatically instigate trades. This presents the potential ability to improve the beneficence through additional profits for themselves, or their respective company and investors. However, if our model makes a mistake and facilitates a trade that loses stakeholders thousands of dollars, this would diminish the bank's or individuals' nonmaleficence. Regardless of how this is used to dictate stock decisions, the results of the model could also affect public opinion of Donald Trump. The model could ensure that everyone has all the knowledge about the effects of Trump's tweets, including on the economy, highlighting the "justice" principle of ensuring equal access of all information available.

References

[1] A. Mittal and A. Goel, "Stock Prediction Using Twitter Sentiment Analysis," Stanford University.

[2] E. Stewart, "The Volfefe Index, Wall Street's new way to measure the effects of Trump tweets, explained," Vox, 09-Sep-2019. [Online]. Available: https://www.vox.com/policy-and-politics/2019/9/9/20857451/trump-stock-market-tweet-volfefe-jpmorgan-twitter?fbclid=IwAR28OTb6ZDKNrOSd2EPyX1GmAkTIanTAf0_Fnsivn_FmH3cpWLiB7QR_Er8. [Accessed: 29-Nov-2019].