How Trump Tweets Affect Daily Stock Prices

Xiaolin Cao, Ting-Wei Lin, Ming-Chen Lu, Suzy McTaggart

Winter 2020

Abstract

Predicting stock market has always been an interesting topic in various disciplines. Our group is interested in predicting the rise and fall of S&P 500 based on the Donald Trump's tweets data. We have implemented multiple machine learning algorithms to test and evaluate the corresponding model performances.

1 Introduction

The ability to accurately predict future changes in the stock market has wide appeal to businesses and investors. In fact, speculation on the stock market has become an industry unto itself.

Surprisingly, there are many reports proposing that President Trump's tweets have a significant effect on the stock market. Trump's activity on Twitter is known to be very frequent and highly contentious. According to Liu (2019)[2], President Trump's tweets do impact the stock market but not for a long time based on the analysis of 14,000 of his Twitter posts.

Our project elaborates on the analysis conducted by Liu (2019)[2] through increasing the time frame, adding a variable to indicate the presidential status, analyzing sentiment score for individual tweet, and including additional twitter features such as favorites and retweets.

If we could successfully classify our data and predict the market then it would be a good idea to start following Trump on Twitter!

2 Data Collection and Preprocessing

2.1 Data Collection

A total of 41,122 tweets over a period of May 4th, 2009 to Jan 20th, 2020 from Donald J. Trump account, along with variables like number of retweets and favorites, are extracted from Kaggle Trump tweets dataset[3]. Standard & Poors 500 (S&P 500) from May 4th, 2009 to Jan 20th, 2020 is also obtained from Kaggle[1].

In order to agree with the response variable (S&P 500), multiple tweets on a single calendar day were grouped together as one observation. New variables like sentiment score and tweet's topic were calculated to augment the two Kaggle datasets.

Trump tweets were summarized as the actual contents of tweets, retweets, and favorites per day. A 3-level categorical variable was also added to the dataset to indicate Donald Trump's presidency status (1=none, 2=candidate, 3=president).

The text contents of Trump's tweets were pre-processed and the corresponding sentiment scores were calculated in the form of compound, negative, neural and positive. The compound score was the normalized version of the rest. This process is described in detail in Section 2.2: Text Processing.

For S&P dataset[1], the "Adj. close" variable was used to dichotomize the change in the market after 1 day. We set this variable (S&P delta) equal to 1 if the "Adj. close" is higher than the previous day and set it to 0 otherwise.

The final dataset consisted of the following variables: Date, Tweet ID, Retweets, Favorites, Tweet text contents, Tweet sentiment score - compound, Tweet topic, S&P Open, S&P High, S&P Low, S&P Close, S&P Adj Close, S&P delta, and Presidency status.

2.2 Text Processing

The tweets were grouped by date to match with the daily S&P indicator. A total of 3151 rows of tweets were pre-processed through a sequence of filtering. Hyperlinks and special characters were removed from each tweet by regex matching. The tweets were tokenized into individual words. Each word was made to lowercase and was lemmatized to its base form. In addition, punctuation was removed. Figure 1 shows an example of such tweet:

```
['be','sure','to','tune','in','and','watch','donald','trump','on','late','night','with',
 'david','letterman','a','he','present','the','top','ten','list','tonight']
```

Figure 1: An Example of Processed Tweet

After processing the tweets, we used CountVectorizer to extract three-letter tokens and to remove stopwords. Tokens that did not appear in at least 20 documents were removed. Tokens that appeared in more that 20% of the documents were also removed. Furthermore, a dictionary of words and their frequencies was created to fit Latent Dirichlet Allocation (LDA) model. We used LDA to construct topic models for tweets. Figure 2 displays the consequent topics:

```
[(0,
  '0.012*"fake" + 0.008*"medium" + 0.007*"hunt" + 0.007*"witch" + 0.007*"russia" + 0.006*"collusion" + 0.006*"election" + 0.006*"fbi" + 0.006*"did" +
ory"'),
 (1,
  '0.016*"border" + 0.007*"united" + 0.007*"tax" + 0.007*"trade" + 0.006*"wall" + 0.006*"china" + 0.006*"military" + 0.006*"north" + 0.006*"security"
*"mexico"'),
 (2,
  '0.013*"run" + 0.009*"true" + 0.007*"apprentice" + 0.007*"golf" + 0.006*"course" + 0.006*"2014" + 0.005*"apprenticenbc" + 0.005*"celebapprentice" +
n" + 0.005*"business"'),
 (3,
  '0.030*"barackobama" + 0.021*"cont" + 0.017*"china" + 0.013*"mittromney" + 0.013*"interview" + 0.009*"iran" + 0.009*"discussing" + 0.009*"tax" + 0.0
d" + 0.006*"oil"'),
 (4,
  '0.016*"hillary" + 0.015*"poll" + 0.011*"makeamericagreatagain" + 0.011*"clinton" + 0.010*"cnn" + 0.010*"foxnews" + 0.009*"debate" + 0.008*"tonight"
*"crowd" + 0.006*"cruz"')]
```

Figure 2: The Resulting Topics for Tweets

We also used pyLDAvis to visualize the model's ability to separate topics. The visualization is shown below:

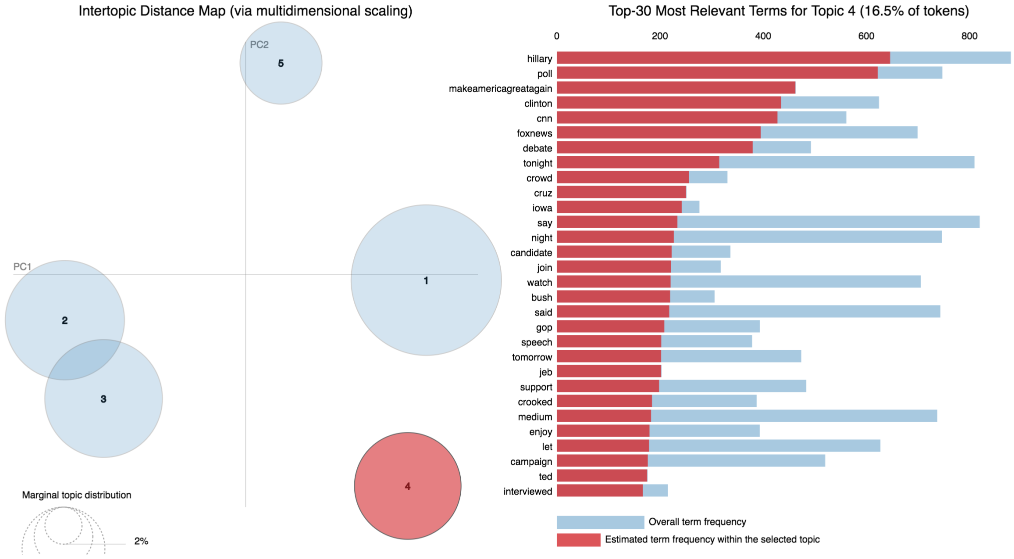

Figure 3: Election

As we can see from Figure 3, the five topics are well separated with small overlap between topic 2 and 3. After inspecting the content of these five topics, we decided to label them as the following and created a new variable ”topic” indicating which category a specific tweet would fall in:

- 0: personal
- 1: foreign
- 2: business
- 3: domestic affairs
- 4: election

In addition to the topic modeling, we also computed the sentiment score for each grouped tweet and created another variable ”ss_compound”. We used ”vader_lexicon” to facilitate the process. An example looks like the following:

```
china is threatening washington over the currency bill we should pas it immediately
washington is wasting over billion this year on solyndra type loan yet they want to cut
military spending in order to save medicare and stop record premium increase we must
repeal obamacare why is the un condemning israel and doing nothing about syria what a
disgrace
compound: -0.7783,
neg: 0.193,
neu: 0.71,
pos: 0.097,
```

Figure 4: Sentiment Score for One Tweet

We have adopted the compound score because it calculates the sum of all the lexicon ratings which have been normalized between -1(most extreme negative) and +1 (most extreme positive).

3 Exploratory Data Analysis

This section provides multiple plots to get an overview of the distribution of each predictor and the interaction between predictors and the response variable.

The Figure 5 shows the top 25 most common words mentioned in Trump tweets colored by the rise or fall of the stock market price. The high frequency of the words "realdonaldtrump" and "trump" is because the original data includes account name in their scripted text. It is also interesting to see that several words like country, time, vote, and watch were mentioned only when we had a rise in stock market, indicating that the prediction of rise/fall of the stock market may be possible through the Trump tweets.

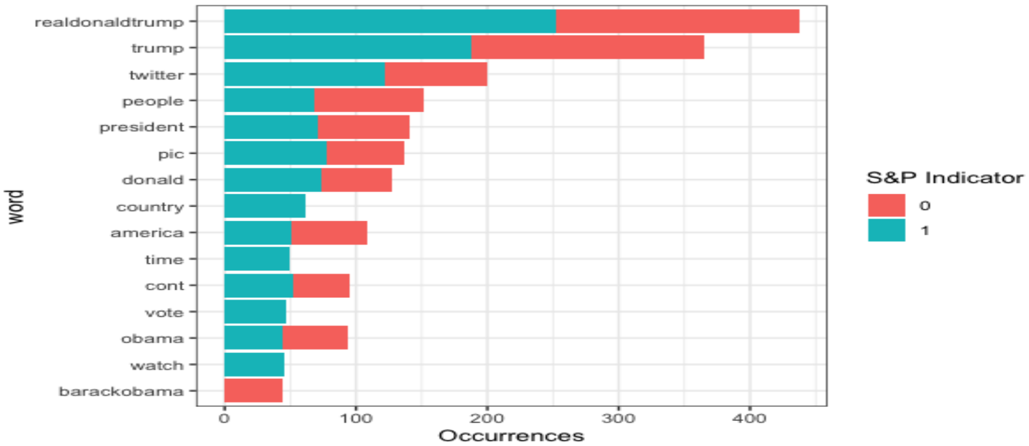

Figure 5: The most common words in Trump tweets

We then examined how S&P delta interacts with continuous predictors including favorites, retweets, and compound sentiment score. All three variables do not show significant difference between the rise or fall of the stock market price.

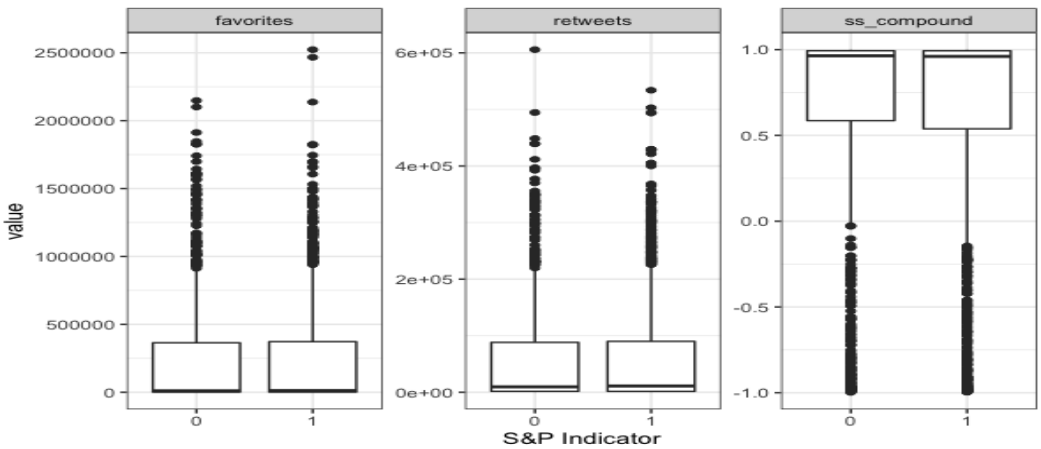

Figure 6: Boxplots of S&P Indicator by Favorites, Retweets, and Presidency

We looked at the relationship between sentiment score and two others predictors, topic and presidency, in the next figure. The Figure 7 shows that the overall tone of Trump's tweets is positive. When the tweets are categorized by topics, there are some variations across five topics, while the presidency status does not.

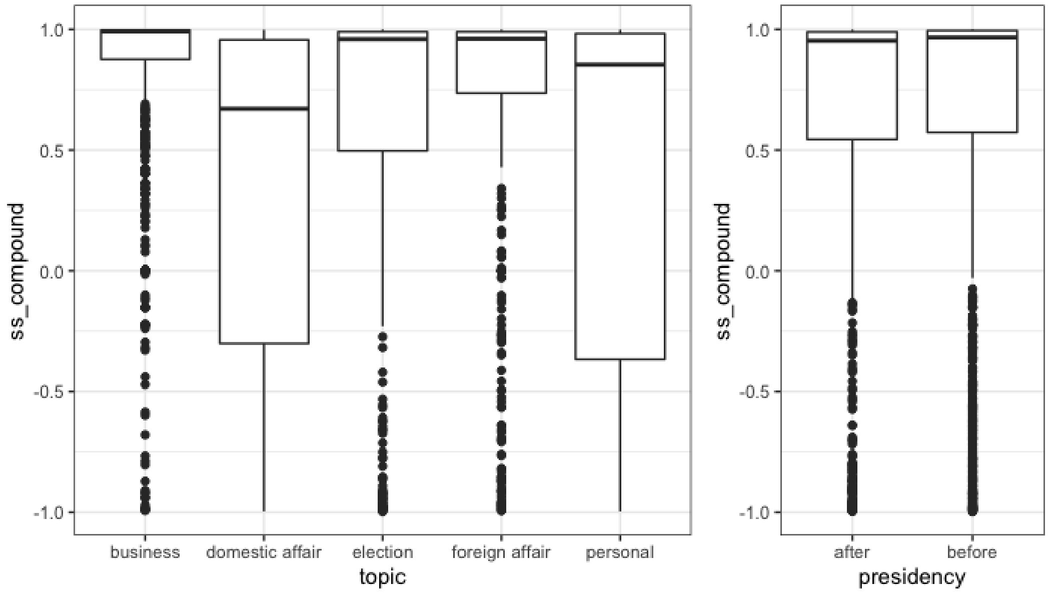

Figure 7: Boxplots of Sentiment Score by Topic and Presidency Status

The Figure 8 indicates that favorites and retweets are highly correlated ans we don't see an obvious boundary between the red and green points which are labelling as the increase or decrease of the S&P stock indicator.

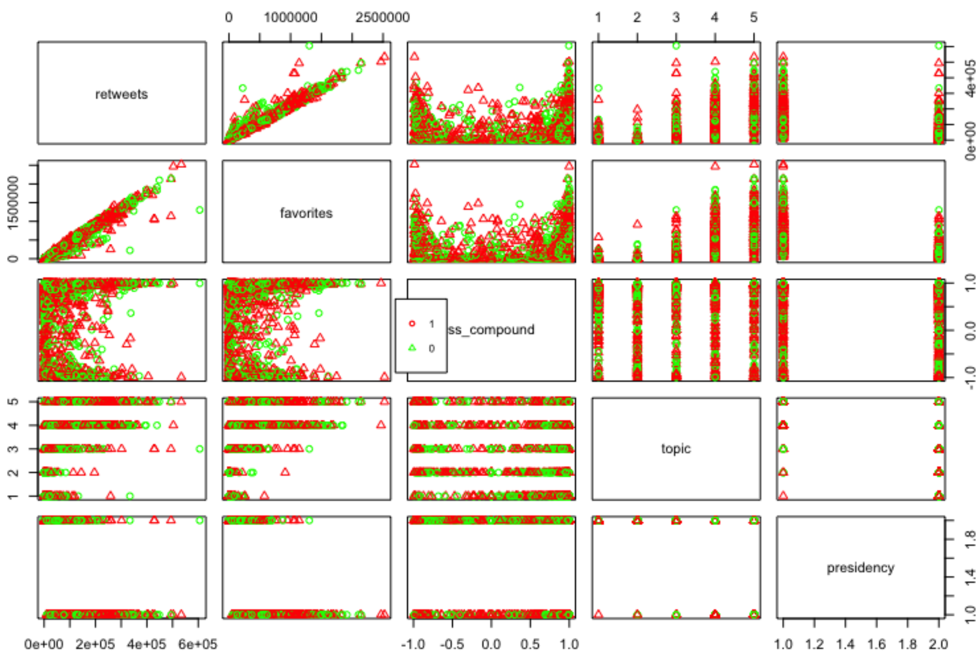

Figure 8: Paired plots for all predictors

4 Model Selection and Results

Since the rise and fall ($c_k = 0$ or 1) of the S&P indicator is our response variable, this analysis is treated as a classification problem. The features extracted above including the number of retweets, the number of favorites, the compound sentiment score, the topics, and the status of presidency were fed to the classifier and trained using multiple supervised methods. The data was randomized and split into two sets: 80 percent for training the model and 20 percent for testing the model performance.

Intuitively, we tried logistic regression model to classify the binary dependent variable. The classifier result had an error rate of 41.41%. Due to the high error rate, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) classifiers were used to train the model. Based on the resulting error rates, neither LDA nor QDA gave a lower test error than logistic regression.

We then considered the tree-based methods since the classifiers would correspond to our research goal: the correlation analysis of the stock price and sentiment of tweets. Table 1 shows that among bagging, random forest and boosting, random forest has the lowest error rate of 40.56%. The results also show that random forest does have a better predicting performance than bagging, agreeing with what we have learned in class. The support vector machine classifier was also used to train the model with the linear kernel and the cost of 0.1 was chosen by 10-fold cross validation.

Table 1: Classifiers Comparison.

Method	Error Rate
Random Forest	0.4056
Boosting (AdaBoost)	0.4141
Logistic Regression	0.4141
SVM	0.4141
Bagging	0.4823
LDA	0.4845
QDA	0.4933

To improve the classifiers performance, we removed the highly correlated variable, favorites, and trained the data using the above classifiers. Among those classifiers, the error rate of bagging reduced to 47.35% though others did not result in a notaable difference. Our other attempts included converting the sentiment score into a categorical variable (positive, neutral, and negative) as well as adding the term frequency–inverse document frequency (tf-idf) matrix into our feature space in hopes of boosting the predicting performance.

Concluded on all the attempts, random forest classifier gives the best accuracy rate of 59.44%. While this result can't provide us a strong confidence in showing the correlation between stock market movements and the sentiments of Trump tweets, we explored various classifiers and concluded that random forest works the best on this particular dataset.

5 Discussion and Conclusions

While the testing errors of the modeling suggest limited ability to accurately predict changes in the stock market related to Trump tweets, there are some interesting practical conclusions to these analyses.

The analysis in the Barron article primarily focuses on (1) the correlation of tweet frequencies on changes in the stock market and (2) specific reference to tariffs and the Fed. Our analysis suggests that these foci are not adequate to accurately predict changes in the market and that some results of the article may be overstated.

Our modeling supports the relationship proposed in the Barron article between tweet frequency (via number of tweets and number of favorites) and reaction in the market. However, it is not possible to suggest Trump tweets caused the change in the market. Trump's tweet frequency would likely increase as world events occur that would impact the market regardless of those tweets (i.e. Trump will tweet more when there is something big to tweet about!).

Our expanded scope of overall sentiment (rather than the limited focus on Barron article on tariffs and the Fed) allow us to speak more broadly on the impact of Trump tweets on the market overall. The relatively low performance of our modeling suggests that Trump tweets are not the primary driver of stock market changes.

Based on this initial analysis, we conclude that President Trump's tweets may be slightly amplifying the impact of other predictors on the market but that they are not primary drivers of change.

Future research could include a review of other influential twitter feeds on the market as compared with President Trump's, inclusion of known market predictors in the modeling, or review of daily news headline sentiments on the market as compared with Trump tweets.

References

[1] Ben Caunt. *sp 500 index past 69 years*, 2019. URL `https://www.kaggle.com/bencaunt1232/sp-500-index-past-69-years`.

[2] Evie Liu. *Yes, Trump's Tweets Move the Stock Market. But Not for Long.*, 2019. URL `https://www.barrons.com/articles/donald-trump-twitter-stock-market-51567803655`.

[3] Austin Reese. *Trump Tweets: Tweets from @realdonaldtrump scraped January 20th, 2020*, 2020. URL `https://www.kaggle.com/austinreese/trump-tweets`.