

Effect of Trump's tweets on oil price (732A92)

Pedram Kasebzadeh(pedka102)

2/11/2020

Abstract

Donald Trump, the current president of the united states, has always been very active on social media. His explicit words have been on the top of the news many times. His post on Twitter (known as tweets) are the subject of this project. The goal of this project is to study the correlation between his tweets and the oil price in the world market.

Sentiment analysis is the approach which is used in this project. First, a dataset of his tweets is collected. Then, by using sentiment analysis approaches points are assigned to his tweets per day. Finally, these points have been compared with the variation of oil price in world market. In order to compare different lexicons, nrc (from Saif Mohammad and Peter Turney), AFINN (from Finn Årup Nielsen), and Bing (from Bing Liu and collaborators) have been used in this project.

Contents

Abstract	2
Introduction	4
Section I: Theory	4
Section II: Data	5
Trump tweets	5
oil price	5
preprocessing	5
lexicons	7
Section III: Method	7
Evaluation	7
Section IV: Results	8
Evaluation results	8
Section V: Discussion	9
Section VI: Conclusion	9
References	11

Introduction

Number of social media users, more specifically Twitter, have been dramatically increased over the past decade (Clement 2019), it plays a huge role in different aspects of our lives on daily basis. Such as, rapid change on fundamental needs e.g. oil price, fluctuation on currencies' value specially on virtual currencies as studied in (Mai et al. 2015). Politics is another aspect in where from local to presidential elections social media has a major role.

In this project, the correlation between Trump's tweets and oil price in world's market is studied. The evidence from some countries such as Iran indicate that his tweets brought some difficulty in their economies, and people's life due to its currency became so vulnerable and fluctuated so much. Hence, in this study the actual effects of his tweets on oil price in world's market is evaluated.

The remainder of this report is organized as follows: Section I presents an introduction about the background and methods which are used in this project. Section II Describes how the dataset is generated in detail and the preprocessing step is presented in this section. The performance of introduced method have been evaluated in Section III. The results have shown in Section IV and discussed in section V. Finally, the work is concluded in Section VI.

Section I: Theory

Sentiment analysis refers to the use of natural language processing, text analysis, and machine learning to quantify, study emotion states and extract information from a given text data. It's a strategy to understand if a given text has positive, negative or neutral state. there are multiple levels of sentiment analysis, when it is done on a sentence it is called sentence level, document level, is when it is done over an entity. Aspect level and user level (connection between different users using graph theory) are two other levels which are not related to our work here. (Saini and Punhani 2019).

Natural Language Processing (NLP) is a tool that make a connection between computers and humans in their own language. For instance, by using NLP computers can read text, hear speech, evaluate sentiment and decide which part should be selected. Sentiment analysis uses different NLP algorithm in order to analysis the human data. Three main types of algorithms used are: *Rule based*, *Automatic* and *Hybrid*.

- *Rule based* is when systems that perform sentiment analysis based manually crafted rules, in other words a lexicon (i.e. lists of words and expressions) level sentiment analysis, which is mainly what is done in this report.
- *Automatic* is referred to when the system relies on machine learning techniques. Machine learning algorithms are divided in to 3 groups, Supervised, unsupervised and semi supervised learning. Algorithms such as Naïve Bayes, Support Vector Machine and Decision Tree are used in supervised learning. These algorithms use a dataset to train on and then are able to classify new content. The reason this approached was not pursued in this project was lack of a labeled dataset which is essential for the purpose of training these algorithms.
- *Hybrid* is when the system uses a mixture of those. Sentiment analysis could also be done in deep learning based, which could be considered as an automatic level with a number of layers in a Neural Network. ("Sentiment Analysis" 2020), (Ahmad and Nikita 2019)

In This project, sentence-level sentiment analysis is done over tweets. A rule-based algorithm which uses lexicons was used due to lack of a labeled dataset. Then a score was assigned to each tweet based on its sentiments. A *day score* for each day was obtained based on scores of that day tweets, then the correlation between daily world oil price and *day score* was calculated.

To find any relations between the tweets and oil price I used Pearson correlation.

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

Where ρ is the correlation between X and Y , cov is the covariance, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y . In our case X is sentiment score for each day and Y is the oil price.

Since $\text{cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$, hence we can write equation 1 as:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (2)$$

Where μ_X is the mean of X , μ_Y mean of Y and E is the expectation.

Pearson correlation has an interval of $[1,-1]$ where 1 means a perfect positive correlation while -1 means a negative correlation(as in if one variable increases the other one would decrease).

Section II: Data

Trump tweets

Collecting data for this project was a bit challenging. First, twitter APIs was used to collect data, however, the limitations made it not the best approach, limitations such as the maximum number of tweets one can get this way. Second, a huge data set of tweets from many US politicians with 1.6 GB size was found online and Trump tweets were extracted. That left resulted in 7300 tweets from Trump in the period of 16th of July 2015 to the first of November 2016. which was not the best dataset since Trump became president as of January 20, 2017. The assumption was that there is a huge difference between a president tweet and a businessman tweet, which made this whole dataset useless.

Finally, the dataset used was from a website (archive, n.d.), this is a website mainly focused on Trump's tweets and gives a variety of options on how to filter tweets before extraction. The dataset extracted was a dataset of all of the tweets provided on the website, which was 46040 tweets from April 2009 to February 2020. However, furthermore It was filtered based on the date so it was just tweets in the time of Trumps presidency.

Figure1 shows a short illustration of the 10 most used words distribution in 2 categories, 'positive' and 'negative'. one interesting observation was how often Trump uses his name in his tweets and that Bing lexicon would categorize it as a Positive word! I did think about removing it from the tweets, however, since the goal is to find correlations a constant plus 1 in the scores would have no effect.

oil price

Now the oil price dataset is also required. The dataset is obtained from the internet.(Thomson, n.d.) This dataset consists of daily oil prices from May 20th, 1987 until March 6th, 2020. That being said, Figure 2 shows prices over the period of Trump Presidency.

preprocessing

To do this project, after collecting the data the most essential part was data cleaning. Data cleaning refers to identifying incomplete, duplicate, incorrect, inaccurate or irrelevant parts of the data and then replacing,

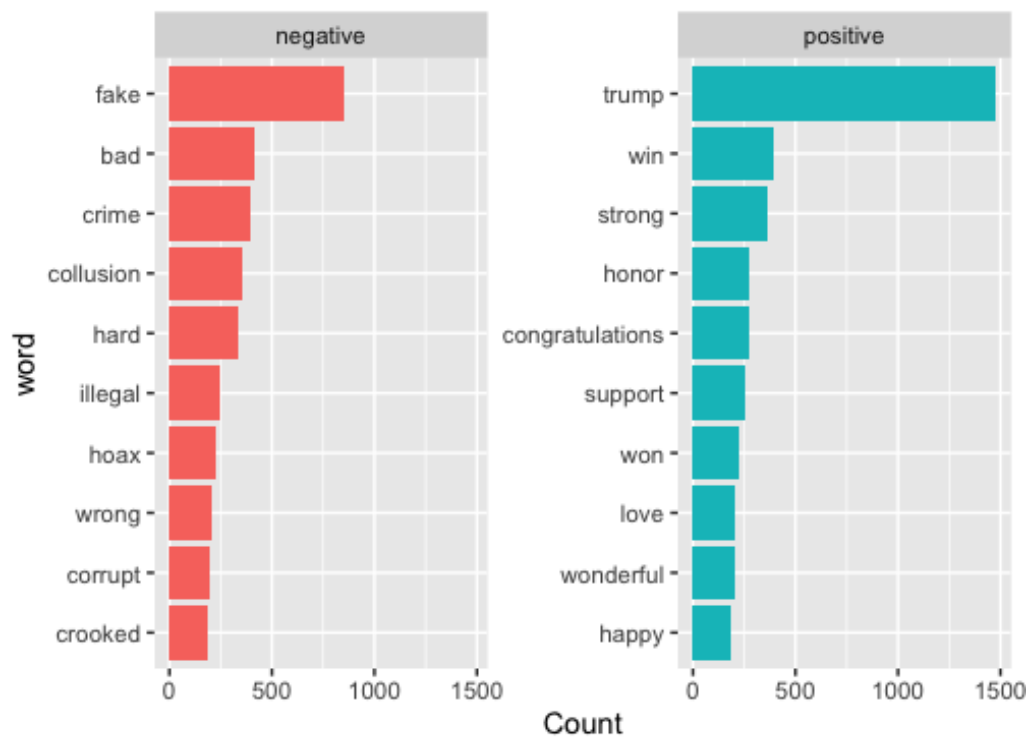


Figure 1: Top 10 Negative and Positive words



Figure 2: Oil Price fluctuations

modifying, or deleting the dirty or coarse data. (Saini and Punhani 2019) to do so I had to use different libraries and multiple functions.

In survey of related works, a sarcasm detection system came across that could be helpful on the data cleaning phase, However, was not used, as its reliability on detecting sarcasm over Trump’s tweets could not be guaranteed as even a human sometimes find it challenging.

lexicons

Three manually created lexicons were used, the NRC Emotion Lexicon AFINN (Nielsen 2011) (about 2477 words and phrases), and the Bing Liu Lexicon(Kohavi and Computing Machinery 2004) (about 6,800 words).

- The *NRC* Emotion Lexicon provides a list of English words which is annotated manually; containing 8 basic emotions such as anger, fear, anticipation, trust, surprise, sadness, joy, and disgust and two sentiment negative and positive. The NRC Emotion list contains about 14,000 words. (Saif M Mohammad and Turney 2013)
- The *Bing* Lexicon categorizes words into positive and negative, As is one of the most popular lexicons. (Paracchini 2016)
- The *Afinn* Lexicon was originally generated for Twitter sentiment analysis; hence it could be a good fit for this project. It has more than 3000 words and it uses a scoring range of [5, -5] where 5 is very positive and -5 is very negative. (Nielsen 2011)

Section III: Method

My method consists of 3 parts:

1. Find reliable list of words with strongly positive or negative sentiment. (lexicons)
2. Count the number of positive and negative words in each tweet.
3. Summing up the scores based on number of positive and negative words.

Figure 3 shows an overview of the steps done in the process of sentiment analysis in this project.

After cleaning the data, since 3 different lexicons had to be used, a function was created with all the necessary functions within (such as tokenization), so it was more convenient to work with. For creating the functions and performing sentiment analysis, R language was chosen, it has a lot of text mining packages and is a great tool to manipulate huge datasets. (Tatman, n.d.)

Evaluation

As the dataset was not labeled, an unsupervised approach had to be chosen for the purpose of evaluation.

Grouping similar tweets seemed like a good approach to get the possible sentiments of the tweets, a popular algorithm to do so is Kmeans (Wojcik 2019), however tweets are short and noisy entities hence the cause high sparseness, this challenge could overcome by the help of term frequency-inverse document frequency (tf-idf) technique, which was the chosen approach.(Orkphol 2019)

Kmeans would categorize the tweets in to two clusters (“Positive” and “negative”) based on how similar they are considering the words used in them. Evaluation is based on the sentiment analysis scores compared to Kmeans clusters.

The evaluation part was done in Python since the word2vec function was malfunctioning on the used operating system (Mac OS).

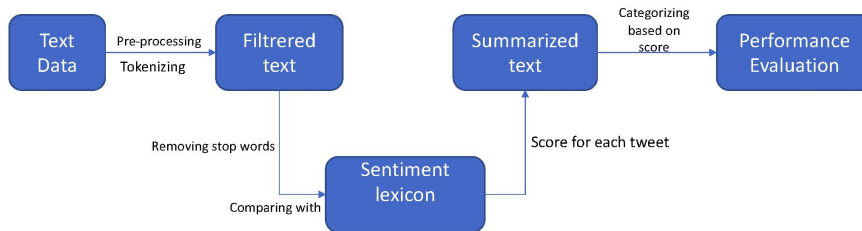


Figure 3: A diagram of the process

Section IV: Results

This section is devoted to presenting the results in form of graphs and tables.

Figure 4 reveals scores for 3 different Lexicons, for each day, over the time period. Scaled oil price is also included to present a general idea. Below you can see a table of the correlation and p-values in the dataset for different lexicons.

Method	Correlation	P-Value
NRC	-0.0015351	0.8475
Bing	-0.004049246	0.6119
Afinn	0.02177592	0.006357

Evaluation results

The Evaluation was done on all 3 lexicons, below you can see the accuracy, recall, precision and F1 score for each of them.

Method	Accuracy	precision	f1
Nrc	0.693373	0.748027	0.730368
Afinn	0.713721	0.757445	0.746052
Bing	0.731949	0.716467	0.747974

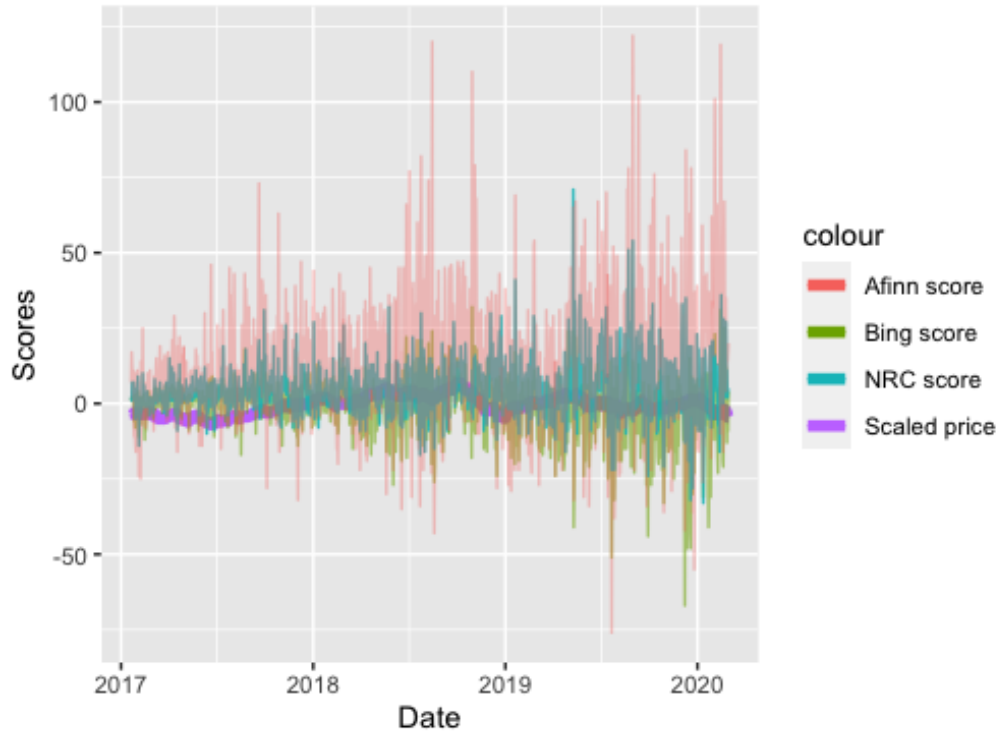


Figure 4: Sentiment Scores VS Date

Section V: Discussion

Figure 4 reveals some interesting facts! As we can see in this plot scores are going too high or too low, considering the fundamentals of each method's score, the reason for this could be a lot of same sentiment tweets in one day. For instance, over 20 positive tweets in one day and maybe 30 negative tweets the day after. There might be very interesting psychological data to be discovered here.

One other noticeable thing on Figure 4 is the different sum scores per day in early 2017 and early 2020, apparently his tweets have more and more positive and negative per day outcomes. either he is tweeting a lot more or using stronger words in each tweet.

On the evaluation table, a higher accuracy for Bing and AFINN, is achieved, compared to NRC even though they have smaller word database, which means using the bigger lexicon will not always give us the best result.

Correlations between sentiment scores and oil price was very low. The most correlation achieved was with Afinn lexicon which still was quite low, and also the only positive one! nrc and Bing showed negative correlation which would initiate that Trumps negative tweets will raise oil price, if the correlation scores were higher.

Section VI: Conclusion

There was not any noticeable correlation between Trump's tweets and oil price fluctuations, so that means my assumption was not correct under these circumstances. which means, Oil price is not affected by what only 1 man has to tweet, even if that man is the president of the united states of America.

Another good approach for future analysis could be using a dataset based on a specific time (e.g. when there are a lot of rumors about a war in the middle east) rather than focusing on one person.

One thing I'm curious about is the accuracy of used and available methods for doing sentiment analysis over Trump's tweets and speeches as he is usually using very unusual structures of language and has a unique choice of words. A manually labeled set of data of his tweets is one way that comes to mind, to obtain a more reliable accuracy of our method functioning, other than unsupervised algorithms.

Several possibilities of improving the performance exist. Maybe using some groups of features would find a better understanding of what Trump says and ends up on better results, as we can see in (Saif M. Mohammad, Kiritchenko, and Zhu 2013), doing so, resulted in great accuracy on sentiment analysis.

Also maybe scaling lexicon scores and merging the results is something to think about, I believe that is one thing missing from my approach, The reason I did not implement it was that I could not think of an efficient and reasonable way to merge these numbers, to be more clear AFINN has a range of +5 and -5 however Bing does not have such limitation, Hence I am not sure how I should compare a 5 resulted with AFINN and a -6 given by Bing. Such concerns resulted in presenting each method separately instead of merging the results.

References

- Ahmad, Gazi, and Nikita. 2019. *Review on Sentiment Analysis of Indian Languages with a Special Focus on Code Mixed Indian Languages*. IEEE.
- archive, Trump twitter. n.d. “Trump Twitter Archive.” <http://www.trumptwitterarchive.com/>.
- Clement, J. 2019. “Twitter Users.” August 14. <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- Kohavi, Ron, and Association for Computing Machinery, eds. 2004. *KDD-2004: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: August 22-25, 2004, Seattle, Washington, USA*. New York: ACM Press.
- Mai, Feng, Qing Bai, Zhe Shan, Xin Wang, and Roger Chiang. 2015. “The Impacts of Social Media on Bitcoin Performance.” <https://pdfs.semanticscholar.org/44ab/c5e1deeb1ef1716dad13f889a5fb45b8d69.pdf>.
- Mohammad, Saif M, and Peter D Turney. 2013. “Nrc Emotion Lexicon.” *National Research Council, Canada*.
- Mohammad, Saif M., Svetlana Kiritchenko, and Xiaodan Zhu. 2013. “NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets.” *arXiv:1308.6242 [Cs]*, August. <http://arxiv.org/abs/1308.6242>.
- Nielsen, Finn Årup. 2011. “A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs.” *arXiv:1103.2903 [Cs]*, March. <http://arxiv.org/abs/1103.2903>.
- Orkphol, Korawit. 2019. “Sentiment Analysis on Microblogging with K-Means Clustering and Artificial Bee Colony.” <https://doi.org/10.1142/S1469026819500172>.
- Paracchini, Pier Lorenzo. 2016. “Package ‘Lexicon’” December 17. https://rstudio-pubs-static.s3.amazonaws.com/236096_2ef4566f995e48c1964013310bf197f1.html.
- Saini, Sonia, and Ritu Punhani. 2019. *Sentiment Analysis on Twitter Data Using R*. <https://ieeexplore.ieee.org/servlet/opac?punumber=8766340>.
- “Sentiment Analysis.” 2020. January. <https://monkeylearn.com/sentiment-analysis/>.
- Tatman, Rachael. n.d. “Sentiment Analysis in R.” <https://www.kaggle.com/rtatman/tutorial-sentiment-analysis-in-r>.
- Thomson, Reuters. n.d. “Europe Brent Spot Price FOB Dollars Per Barrel.” <https://www.eia.gov/dnav/pet/hist/RBRTED.htm>.
- Wojcik, Rafal. 2019. “Unsupervised Sentiment Analysis.” November 27. <https://towardsdatascience.com/unsupervised-sentiment-analysis-a38bf1906483>.