Shareholders' Paper

In this paper, a short recap of the assignment will be made.

This assignment has focused on analysing Trump tweets. More specifically, the relationship between Trump tweets and the market index of certain stock markets.

Data

Four datasets are used for this assignment. First off, the Trump Twitter Archive has been used to extract tweets from Donald Trump from the 11th of October 2016. This date has been chosen, as it's about this date it became clear, that Donald Trump was very likely to be elected President of the United States. The remaining three datasets are of the major stock indexes S&P500, DAX, and Shanghai Stock Exchange. Data from the indexes has also been extracted from the 11th of October 2016. Index data is from Yahoo! Finance, and we have downloaded the data to Excel, cleaned it up (deleting unnecessary rows etc.) before saving as a csv-file, and uploaded to our GitHub repository.

The datasets etc. can be downloaded through the GitHub page:
https://github.com/JBOzol/m2-group
Similarly the notebook are set up to download the datasets, so no action needs to be taken for this.

The following is a description of the variables in the four datasets. Note that only two datasets are described, as all three indexes has the exact same structure.

Trump tweets:
source: this shows if he's using Twitter for iPhone, Android and so on
text: this is the actual tweet
created_at: date and time the tweet was created
retweet_count: number of retweets
favorite_count: number of favorites
is_retweet: boolean that is True if the tweet is a retweet
id_str: ID of the tweet

Stock indexes:
Date: date
[index name]_close: adjusted closing price of the index, adjusted by Yahoo! Finance
[index name]_return: the calculated return between the previous trading day and the current trading day

[index name]_op: a dummy variable with the value 1 if the market goes up on a given trading day, and zero if it goes down

Problem

With Donald Trumps very active and criticized Twitter presence, we would like to examine the interaction between Donald Trumps trade war related tweets and the three indexes, which we use to represent the overall stock market in the three regions: North America, Europe, and Asia.

Thus we define the hypothesis:
Does the markets react to Donald Trumps tweets, and if so: is it pure hysteria?

To answer this question, we focus on Natural Language Processing, tokenizing Donald Trumps tweets, and using machine learning techniques to see if we can predict if the market goes up or down, based on the tweets themselves. Before doing this, however, we first try to use some explorative data analysis to see if the markets react to Trumps trade war related tweets. The following three graphs show the index prices, with Trumps trade-war related tweets marked on the plot line.

In the HTML document found in the github repository, there's a short presentation of the 3 index evolutions. Here are shown 2 main groups, consisting of 4 graphs each. These groups each show an interactive presentation of the index evolution over time, with a hover ability, showing Trump tweets during the period. First group is consisting of only trade related words, while the second group contains no filtration of words, so all Trump's tweets are found in said presentation. First graph in each group is a combined plot of the 3 index, with China and USA close values being on the left axis and DAX on the right. The X axis consists of dates, and a hover option shows you the Tweet of that day. The next 3, are individual evolution of Index for USA, China and DAX.

As can be seen from the graphs, it seems like some of the big spikes happens at the same time or shortly after Trump has made a trade war related tweet.

This might imply that the markets do in fact react to Donald Trump's tweets. This can be further examined by vectorizing Trumps tweets, and use those to predict market changes using machine learning techniques.

Machine learning techniques

By using supervised learning, the goal is to use Trumps' tweets to predict market movements. This has been done by vectorizing his tweets, and using those word vectors to predict market movements.

Conclusion

Keeping in mind, that the analysis is heavily biased, and doesn't take any other factors in mind, other than Donald Trump's tweets, it can be concluded that we haven't been able to predict the markets reactions to Donald Trump's tweets.