Sentiment analysis of tweets and their effects on the Stock Market

Hrishikesh Mahajan*
Department of Computer Science
Binghamton University - State University of New York
hmahaja1@binghamton.edu

Jainil Parikh†
Department of Computer Science
Binghamton University - State University of New York
jparikh1@binghamton.edu

ACM Reference Format:
Hrishikesh Mahajan and Jainil Parikh. 2020. Sentiment analysis of tweets and their effects on the Stock Market. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Twitter has a large audience potential. Currently it attracts an estimated average of 271 million users every month. It is therefore not surprising that the rich and continuous mass of data made available by these platforms is being harnessed with the purpose of studying individual and group behavior as well as global patterns especially in regards to sentiment towards brands, products, events, recent news and social and political issues.

Stock markets is a highly volatile and dynamic platform where events happening in real time have a significant capability to alter the course of the market. Since we have discussed how Twitter has the potential to sway public opinion, we were interested to find out if tweets, its popularity among the community and the sentiment of the tweets could affect the stocks of trending companies. We thus plan to perform a sentiment analysis on twitter data and try to find out if any such co-relation exist.

In this project, we plan to implement a dashboard to showcase different plots resulting from our data analysis. We implement 4 graphs as a part of this project on the analysis we did in project 2.

2 PREVIOUS IMPLEMENTATION AND ANALYSIS

Our project is based on sentiment analysis based on tweets related to stocks. Our data collection system has been gathering data from both, the Twitter and the stock market website and storing them into our database. As a part of the Project 2, we performed analysis on the Tweets filtered from the 1 percent stream and the stock market data collected for the Top 30 trending stocks. Our analysis was focused on 2 important events that were captured. First, Tesla joining the S & P 500 index, which led to a drastic spike in the stock market price of Tesla and second, The Federal Aviation board lifting the ban on Boeing 737 Max which have had massive accidents killing more than 300 people on board.

Our analysis revealed that there was a causation of relationship between the tweets posted during both the events and the tweets posted has had a significant role in influencing the stock market prices of both the symbols.

3 INSTRUCTIONS TO CONNECT TO THE DASHBOARD

3.1 Connecting to Sentiment Dashboard

Login to our VM, the details are following

- Username-jparikh1
- Password-PXqqeDf42

After logging in, if chrome is already running, just hit refresh. If not, open the PyCharm window and in the Terminal tab, run the Sentiment Dashboard by the command:

- steamlit run sentiment-dashboard.py

Our dashboard is actively running via VM on google chrome window, where you can navigate, visualize the graph according to the filters provided. Before you run anything, it is advised to refresh the page. Since the final plot produced is a combination of 4 and more graphs, please make sure you scroll to the bottom in order to fully visualize the charts.

3.2 Connecting to Time Series Dashboard

Since, both the dashboards run on the same localost and port, we cannot run both simultaneously. So you need to force stop the execution of the Sentiment Dashboard in order to see this one. Press ctrl+c in Terminal tab of pycharm to stop the running dashboard. Run the Time Series Dashboard by the command:

- steamlit run Final-dashboard.py

Once the dashboard is open, you will have to upload the CSV files. Click on browse files, go to downloads select the file "Tesla-stockstopredict.csv". The next step is to scroll down select forecast

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

horizon. Forecasts become less accurate with larger forecast horizons.

4 DETAILS OF DASHBOARDS

In this section the we will be discussing about the analysis plots, chart tools we have included in the live dashboard. We also have a time series dashboard which needs to be run separately.

4.1 Plots included in Sentiment Dashboard

Our Dashboard will have 4 different plots in the sentiment dashboard which can be interacted via a set of drop downs and buttons. The time series dashboard can run only when we exit the sentiment dashboard. For the scope of this project, we have limited our data to the dates from November 8th 2020 to November 20th 2020. Since the fourth graph is a combination of 4 and more charts, please make sure to scroll down to the end of the page to see all the charts. We can describe each plot as follows:

- The first plot has 2 parts. First is to display any random tweet based on the sentiment you assign. Figure1 represents any random tweet based on the sentiment selected. Second is the visualization of the number of tweets based on sentiments. Figure2 shows the total number of Tweets per sentiment. There is a drop down on this part which converts the graph into a scatter plot. Select a visualization type and check the show button.

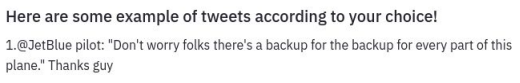

Figure 1: Random sentiment (Positive) Tweet

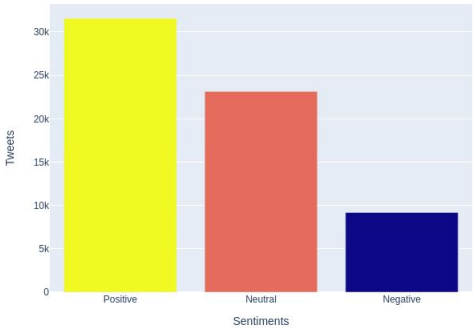

Figure 2: Total Tweet count based on sentiment

- The second plot is the Breakdown of tweets for Tesla and Boeing based on the sentiment. This is also a histogram showing either of the two or both histograms at the same time. In the bar, select either of the two or both and then click show to see the plot. Figure3 represents the visualization for sentiment based count for a stock

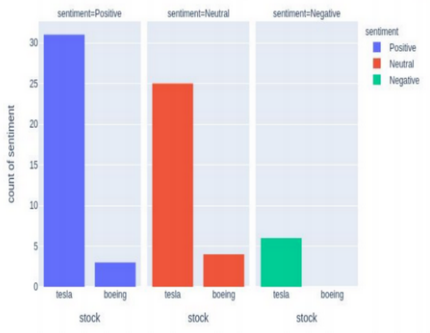

Figure 3: Sentiment based Tweet count for Tesla and Boeing stocks

- The third plot is a word cloud. This word cloud can be interacted using 3 radio buttons each button showing different word clouds being plotted for the 3 sentiments namely, positive, negative and neutral. Select any one sentiment and check on show to see a word cloud generated on the sentiment you input.Figure4 shows the Word-Cloud

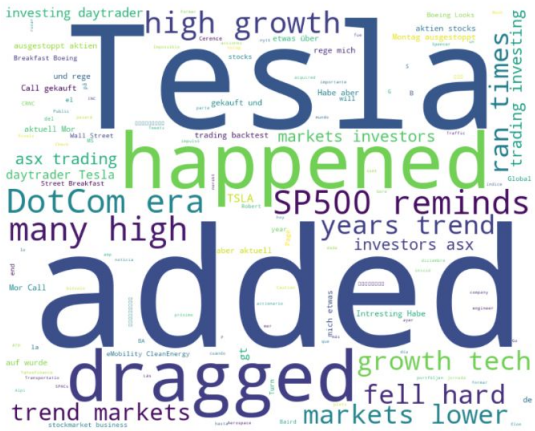

Figure 4: Sentiment based Word-Cloud

- The fourth and the final plot is a combination of plots showcasing closing price, High price, low price, and volume for any one of the 30 stock symbols we analysed. From the drop down, select any Stock you want to see, select the start and end dates, and click on generate charts. Figure 5 shows one of the graph of the trend of the closing price of Tesla's stock.

4.2 Plots included in Time Series Dashboard

The time series dashboard shows us the prediction of the Stock price within the given time period. It uses the Prophet library to produce predictions. For the scope of this project, we have limited this dashboard to Tesla's stock and between the time period of October 22nd to December 3rd. It shows a trend of the stock, the

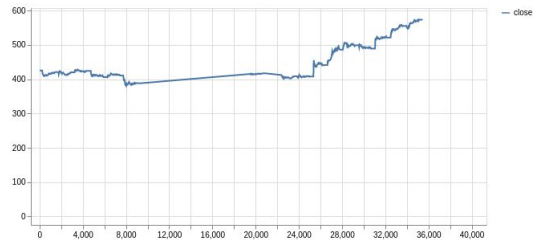

Figure 5: Trend of Closing Price (Tesla)

weekly and the daily performance forecast. Figure 6 shows the black dots as the actual price against the line which represents the prediction for a 5 day period. Figure7 represents the trend of the stock in given time period and Figure 8 represents the trend over the week. One main thing to note here however is that if you increase the horizon period, it will lead to less accuracy.

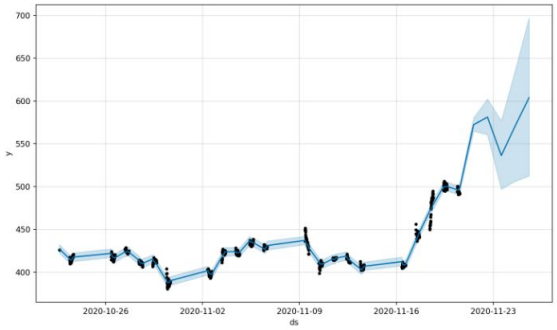

Figure 6: Prediction of Stock Market Price

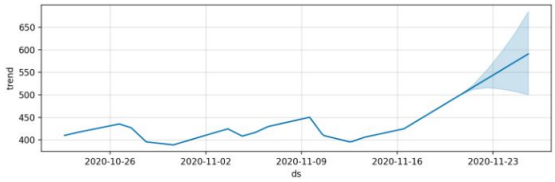

Figure 7: Trend of the Stock

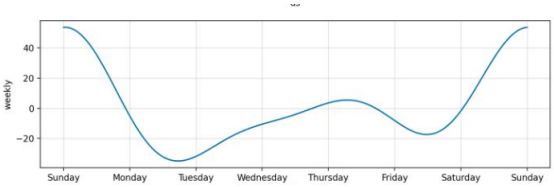

Figure 8: Weekly Trend of the Stock

5 TOOLS USED IN DASHBOARD

We plan to use following plotting libraries:

- Streamlit: Turn data scripts into sharable web app also provides inbuilt plot options.
- Plotly: This makes it easy to create, deploy, and share interactive web apps, graphs, and visualizations.
- Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.