Stock Market Sentiment Analysis

Ajwad Javed

August 2019

The goal of this project is to run sentiment analysis on a company primarily operating in Pakistan and use that to predict its future rise or dip in stock price. I looked up the Top Holdings of companies on `https://finance.yahoo.com/quote/PAK?p=PAK&.tsrc=fin-srch` and decided to start off with a list of the following companies:

- Oil Gas Development Co Ltd (8.86)

- Habib Bank Ltd (4.84)

- Engro Corp Ltd (8.80)

- Pakistan Petroleum Ltd (8.21)

- Lucky Cement Ltd (5.20)

- Fauji Fertilizer Co Ltd (5.02)

Note: The number in the bracket represents the percentage of assets owned.

Currently I have decided to a pretrained neural network trained on word embeddings from the IMDB datasest, I have not yet decided where I will get data of people expressing their opinion abuot a particular company since our social media isn't focused on that sort of stuff (I've rarely ever seen people discussing stock prices on Reddit/Twitter/Facebook but a simple search of the hashtag Engro shows us the following tweets which may later be of use:

- Asad Omar as chief of ENGRO ILLEGALLY rented out its EVTL Chemical Terminal to its sister company EPTL for handling LNG cargoes WITHOUT getting any indispensable requisite NOCs as per LNG

Policy LNG Rules which is contempt of SC decision on Constitutional WP No. 11/2013 - Shahid Nauman Afzal

- The @SAFoundationN in collaboration with TharFoundation managed supported by Engro is building "Thar Hospital" for the poor. We donated PKR 100mn last year for this cause and Insha'Allah we will support our brothers and sisters of Thar with this 120 Bed Hospital. HopeNotOut - Shahid Afridi

- Shahid Khaqqan Abbasi gave a Deal to Engro corporation for LNG Terminal. Invest $130 million. Get $270k a day. 365 days for 15 years. That is $8 million a month. $97 million a year. Almost $1.5 billion in 15 years. That is 1100% guaranteed return from tax payers money. - Mir Mohammad Alikhan

Extracting the tweets and their dates shouldn't be that hard using the Twitter API (the names aren't relevant but this poses another problem: how do we filter out anomalies or if a person has a bias towards a single company and continously keeps posting fake news about them? Verification is a different challenge to tackle and I might just filter out my list manually or add a fake news detector as an extension to this project). After that, matching the dates of the tweets or any other social media posts (including newspaper like BBC, WSJ, Dawn?) with the stock price data and checking any correlation (through numpy) would also be pretty easy, model building is going to be the easiest part of all of this. Getting and cleaning the data is going to be hard.

Another issue thats coming to my mind right now (if we restrict our task just to Twitter) is for example if we take the recent Pakistan IMF bailout, stuff like that is not going to show up on the hashtag Engro, or the recent drama in Kashmir must make investors slightly hesitant to invest in Pakistan? Or the recent news announcement of the next PSL being entirely hosted inside Pakistan, is that going to effect the stock price of HBL? The issue is how far do we search? Or just for the sake of completing this task within this month I might restrict the API to a few select hashtags (or whatever website I choose to search on) and then find the trigger words over there based on date and check any correlation between the POSITIVE/NEGATIVE rating of the comments and the UP/DOWN price fluncuations of the stock.

Also I just found this website called `https://pkfinance.info/kse/stock/hbl`, experimenting with this is going to be fun.