**Natural Language Processing for Determining the Sentiment of Stock Tickers**

Jayanth Rao

Venkat Ramaraju

Arizona State University

Ajay Bansal, Ph.D. (Thesis Director)

James Smith (Second Committee Member)

**Abstract**

1. **Introduction**

It is no secret that the public takes a tremendous interest in the stock market. Many have attempted to study market trends and use this knowledge into making fruitful investments. This paper attempts to determine and justify a curated sentiment of a stock ticker and leverage this analysis to advise a user on whether or not to buy, sell, or hold a given stock. It aims to detail the process behind generating a sentient polarity for a given string of text – and how integrating a stock-based ontology into developing a sentiment analysis model that is able to accurately identify the perception of a given stock. Our analysis was done on data scraped from condensed online discussion platforms, daily journalist headlines, and qualified financial stock analysts. We have incorporated dynamic, real-time changing data into this model in an effort to produce an accurate daily sentiment. Each demographic’s sentiment was weighed relative to their involvement in the market and sentiment scores were generated after aggregating many such data points.

We have verified our predictions against industry standard analyst predictions in order to gain insight into the accuracy of such predictions. However, in addition to this paper, the code for such sentiment and prediction generation is open-sourced, as opposed to industry analyst reviews, which work on a pay-per-view/monthly subscription model. This paper aims at showing its users and readers a holistic perspective of their favorite stocks – with a conservative, verified estimate of what to do with them as well. We believe that everyone should have access to such data and models, as they are generated by the users themselves.

The stock market is changing by the day, and it is people’s perceptions that are affecting them. Common traders are increasingly able to swing prices – solely based on their outnumbering of financial institutions. This paper aims to bring the information together and present it to users to give them a good idea of how to handle and manage their stocks.

1. **Background**

Unsupervised social media platforms have allowed users to express themselves free of accountability and repercussions. A tremendous amount of false information has begun to flood various platforms like Facebook, Instagram, Twitter, and other mainstream social media sites. The advent of bot accounts has only exacerbated the situation into skewing the mind of an average user regarding a particular topic. Misinformation can spread quickly through bot accounts resharing and retweeting information without verifying the facts (Ferrara et al., 2016). In this context, misinformation is the factually incorrect information that is spread via social media (ASU article). While bot accounts play a role in the spread of misinformation, the basic aspect of social media allowing anyone to publish what they please are pitfalls users must be aware of. The controversy surrounding the 2016 Presidential Election is a prime example of how misinformation distributed via social media can affect the way people view the current political landscape. Spam, rumors, and “fake news” are all forms of misinformation that can alter decision-making.

With the advent of social media, the stock market has become ever more accessible to the public, but with this accessibility comes a risk of baseless opinions with an amplified voice. In 2017, the Security and Exchange Commission (SEC) announced an investigation into entities that published fraudulent articles promoting certain stocks over others. Public companies had hired writers to publish articles without disclosing that the articles had been sponsored (SEC Press Release, 2017). Fraudulent articles like these have the power to drastically alter how the general public view stocks and what stocks to invest in. However, there are other factors to consider as well. Online forums, as mentioned previously, have the power to create echo chambers. The same opinions can be parroted to double-down on one way of thinking. Investment forums like Yahoo! Finance conversations and Reddit’s r/WallStreetBets can perpetuate positivity bias, where members may be unnecessarily bearish or bullish on a given stock. Studies have shown that this can lead to excessive trading during times when that may be fiscally inadvisable (Tang et al., 2017).

1. **Literature Survey**

In this section, we present related literature to utilizing sentiment analysis to gauge a given stock’s sentiment. Our work is focused mainly on how conversations from Yahoo! Finance’s “conversations” feature and journalistic headlines contribute to a stock’s sentiment, and whether that derived sentiment can be viable guidance in terms of whether an individual should buy, sell, or hold that stock. Das and Chen (2004) present a derivation of sentiment from Yahoo’s old stock market-focused message board. They note five separate classification algorithms for determining sentiment, which provided a better tracking across indexes instead of individual stocks. <*we should consider this as well*>. Similarly, Mittal and Goel (2011) analyzed Twitter conversations to compare the sentiment provided by public discourse to the sentiment seen by market fluctuation. Paired with the sentiment were the Dow Jones Industrial Average’s (DJIA) previous day values to provide a prediction for stock movements in the future. The authors accounted for Twitter feeds from June 2009 to December 2009 and implemented a sentiment analysis classification to divide the data in four “fuzzy” categories. This classification was performed by Self Organizing Fuzzy Neural Networks (SOFNN). The prior DJIA values are preprocessed and are paired with the classifications, which are then fed into the SOFNN models to determine future DJIA values. They then corresponded these predicted values with their stock portfolios to perform buys or sells for given stocks. Zhang, Fuehres, and Gloor (2010) performed a similar study, where they analyzed six months’ worth of Twitter feeds. Instead of a neural network to perform classification like Mittal and Goel, the authors utilized a randomized subsample of the full volume of the tweets, and then measured the “hope”, “fear”, “happy”, “worry”, and other such mood words to tag each tweet. They also factored in the follower count per day, with the reasoning that both positive and negative sentiment can carry farther the larger the audience a Twitter user has. Finally, the authors analyzed the amount a tweet was “retweeted”, supported with the hypothesis that the more an opinion is reshared, the more relevant it is.

Compared to these above approaches, the main difference is the type of data that is collected. Similar to Das and Chen’s data, we are using Yahoo! Finance’s conversations feature to collect tuned data. As mentioned in the *Background* section of this paper, the members who post in Yahoo! Finance conversations can be assumed to be knowledgeable about the stock market. Journal headlines regarding the stock market are also helpful in terms of tuning our prediction model with data that may not necessarily be affected by the echo chamber phenomenon, also noted in the *Background* section.

1. **References**

The Rise of Social Bots: <https://arxiv.org/pdf/1407.5225.pdf>

$FAKE: Evidence of …: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/viewFile/17871/17055>

A large scale study to understand…:https://www.computer.org/csdl/pds/api/csdl/proceedings/download-article/12OmNzzP5HP/pdf

ASU Article: <https://www.public.asu.edu/~huanliu/papers/Misinformation_LiangWu2019.pdf>

“SEC.gov” SEC: Payments for Bullish Articles on Stocks Must , 10 Apr. 2017, [www.sec.gov/news/press-release/2017-79. Accessed 29 Jan. 2021](http://www.sec.gov/news/press-release/2017-79.%20Accessed%2029%20Jan.%202021).

Das and Chen: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.6418&rep=rep1&type=pdf>

Mittal and Goel: <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>

Zhang, Fuehres, Gloor: <http://www.ickn.org/documents/COINs2010_Twitter4.pdf>