Performing sentiment analysis on stock tickers using mined data from Twitter

Rakesh Boyina, Devendar Sai Thallam, Dake Shen, Nivedita Ravi, Rasika Sunil Acharya, Hongdi Li

Syracuse University

Introduction

- One such area which is impacted due to the social media is the people's investment strategy in the stock market.
- In this project we are trying to conduct an experiment which allows us to understand the user sentiment about specific stock or set of stocks using the social media data.
- Our approach is to extract the tweets containing a specific stock based on the stock ticker/symbol from all twitter users in a particular time frame.
- On an average, there are a total of 500 million tweets exchanged on Twitter [1].

Role of Twitter Sentiment on Stock Market:

- The root of the idea for this project is derived from the increased trading activities based on the user tweets in twitter. In the recent past, we have noticed that a tweet from Elon Musk has led to a sudden surge in the stock price of “Twitter” company stock. Similarly, there seem to be a correlation of the twitter activity with the meme stock rally during the year of 2021. In real time, a layman tends to rely on the news, articles, tweets or reddit posts for their investment strategies. This would mean, they are going with people’s pulse, and it shows there is a significant impact of people’s sentiment on the stock trading activities.
- Performing an analysis of the data would empower the user to further extend this project to define their trading strategies based on the volume of tweets and correlating this data with other parameters related they intend to evaluate before performing the trade. This project can be used as an input to third-party applications to enhance the accuracy of their predictions.

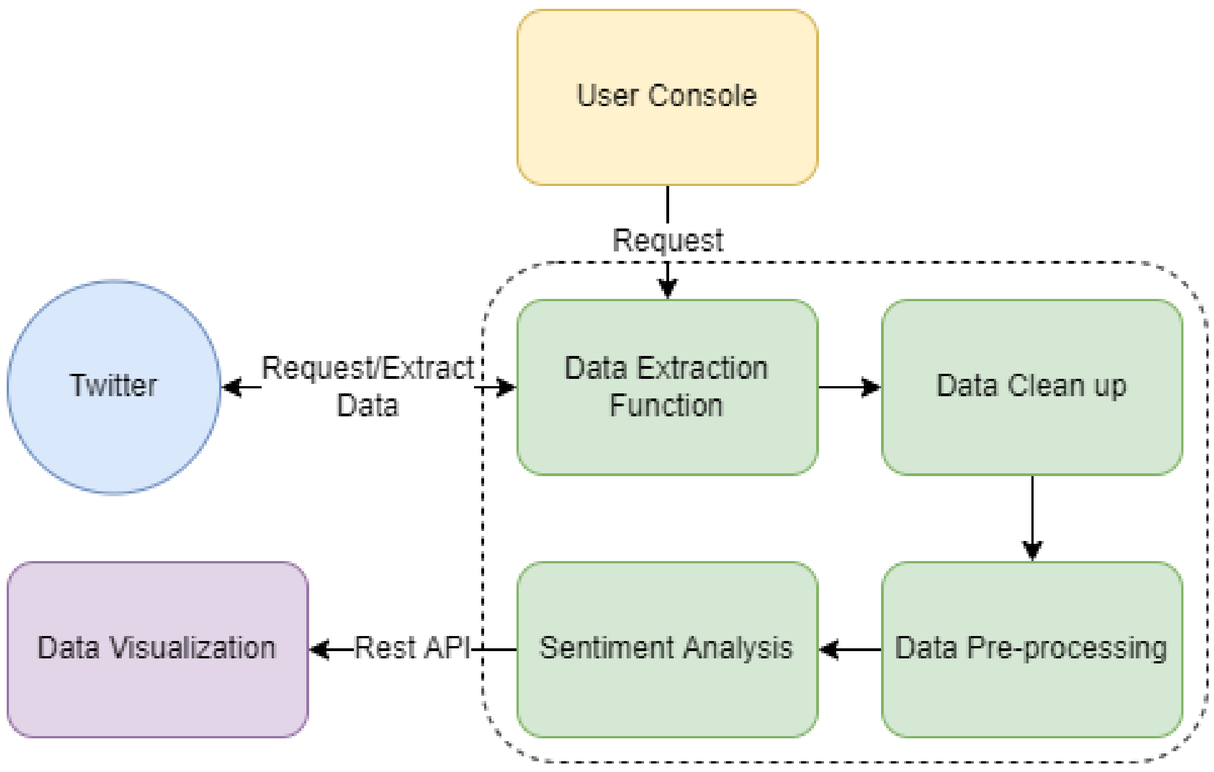
Data Set

- 7 days recent Tweets from Twitter API : Timestamp, Tweets.
- For Sentiment Analysis : Analyzed Lemma and POS tags for tweets using NLTK library.
- 7 days recent Real Time stock data from yfinance API.

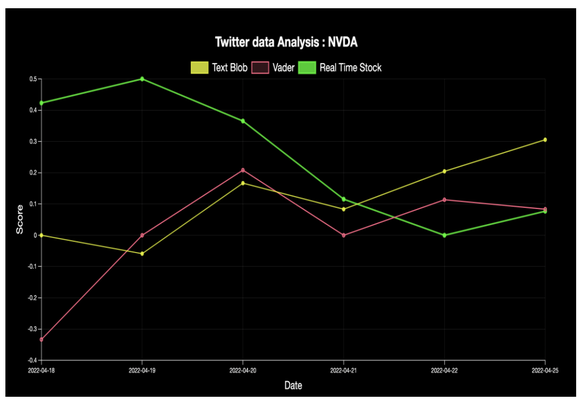
Study Methodology

- Twitter Data Extraction using Tweepy API
- Data Clean-up
- Data Pre-Processing
- NLP Sentiment Analysis
  - Text-Blob Sentiment
  - Vader Sentiment

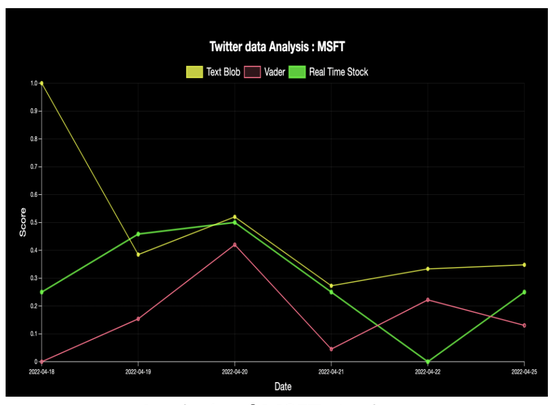
Project Design



Analysis and Results



Analysis of NVDA stock



Analysis of MSFT stock

Challenges

- Cleaning Raw Data into usable form
- Converting different timestamps from different datasets from APIs into same time format.

Sentiment Analysis

- TextBlob is a Natural Language Processing (NLP) Python library. Natural Language ToolKit (NLTK) was used extensively by TextBlob to complete its objectives.
- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is tuned in to social media sentiments.
- We used these libraries because of their simplicity and its accuracy compared to other sentiment analysis methods available. We can further improve sentiment by using most complex analysis using Machine Learning.

Conclusion

- The extraction of data from social media platforms such as Twitter and processing the data using NLP techniques taught in the CIS 668/IST 664 class have given us a fair number of insights to infer a relationship between user perception and stock market performance.
- Our initial hypothesis of user sentiment on social media platforms about a particular stock is translated to the actual performance in the real-world is substantially proven with the promising results from the results generated in this project.
- However, it is not always ideal to assume that the relationship is linear to the user sentiment and our idea is only to present the comparison of the data extracted against the stock market.

Future Work

- This project can be further extended to extract data from multiple social media networks, trading platforms etc. and develop a stock price prediction system using Artificial Intelligence/Machine Learning techniques.

Get this poster

References

1. https://www.omnicoreagency.com/twitter-statistics/