Github Link:
https://github.com/anjanshrestha123/Sentiment-Analysis-for-Stock-Price-Prediction
Project Title: **Sentiment Analysis for Stock Price Prediction**
Team Members: **Anjan Shrestha, Aashish Pandey, Umesh Jaiswal, Puja Dhungana**

1. **Motivation**

Due to the availability of large data sets, Natural Language Processing (NLP) techniques and Machine Learning (ML) models have become an integral part of finance research. Many researchers have analyzed the text documents to see how investors' sentiment affects the stock price movement. Twitter data was used to predict the investors' mood and used it to analyze the stock markets movements (Mittal and Goel). They formed a naïve portfolio management strategy for their analysis and obtained 75.56% accuracy using Self Organizing Fuzzy Neural Networks on the Twitter feeds. Bollen et al. (2011) predicted the stock market with 87% accuracy using a similar technique. Similarly, Kim et al. (2018) used textual data from blogs, financial reports, and news to predict the stock price movements. Specifically, they used 8-K financial reports of firm's sector by sector and found that their approach improved the stock price prediction by 25.4%.

Although many managers and academic researchers have published their works on stock prediction, it is still identified as an important empirical problem in the finance field. In this paper, we will use data from different social media to predict public mood and stock price movements. We will also compare our study with some previous works to see if our approach improves prediction performance.

2. **Significance**

Stock market follows a random pattern and contains many calculated and uncalculated risks for the investors. This makes the stock market more fragile. The stock market sentiment is directly impacted by different factors such as politics, news and industry. Social media has become a common platform to share, discuss and give opinion on these factors. Thus, the positive and negative reviews on social media largely affects the stock market. Positive reviews increase the stock value while negative reviews decrease the stock value.

In this project we are analyzing such sentiment of people over social media to minimize some uncalculated risk. This could predict the possible future fluctuation of a stock price. It

could attract many new investors. Moreover, it could help active investors to make decisions to enter or exit from the stock.

The stock market cannot be solely predicted by sentiment analysis. However, combining it with other fundamental analysis and technical analysis could help to make a more precise decision on particular stock.

3. Objectives

The main goal of this project is to predict stock prices based on market sentiment and provide a summary of people's sentiment using text summarization techniques. As we know, stock movement is largely based on people's feelings and emotions. So, if we can identify the correct emotions and sentiments in the market, there is a huge chance that we can predict the price of the stock market.

When people talk about stocks in blogs, chats and articles, they generally include positive sentiments such as happiness, hopefulness, enthusiasm etc. or negative sentiments such as disappointment, hate, sadness, etc. in their sentences. In the present world, social media has become the perfect platform to get those positive, negative or neutral sentiments from the people around the world.

So, our main objective is to extract those sentiments from different social media by using their APIs or by performing various web scraping techniques, to train and test our model out of those data, to predict the price of various stocks and to summarize findings in the form of text. In other words, our model should be able to suggest the best stocks that have a higher chance of price increase based on sentiment analysis.

4. Features

In this project, we will create a ML model to analyze the text data and perform the prediction on upward / downward movement of overall stock price (also the stock price for a list of highly volatile companies). We will exploit the sentiments expressed during communication or discussion in social media groups to perform the training and make predictions. Our data source will be various social networks like twitter, discord and facebook groups. We will also use the data from the New York Stock Exchange, publicly available in Kaggle [5] , to map the daily conversations with daily movement in stock price. Furthermore, we will generate the summary

of conversations where the stocks are comparatively highly volatile by combining abstractive and extractive text summarization techniques.

Reference:

1. Bollen, J., & Mao, H. (2011). Twitter mood as a stock market predictor. Computer, 44(10),
0091-94.

2. M. Kim, E. L. Park and S. Cho, "Stock price prediction through sentiment analysis of corporate disclosures using distributed representation", *Intelligent Data Analysis*, vol. 22, no. 6, pp. 1395-1413, Jan. 2018.

3. Mittal,A., & Goel,A. Stock Prediction Using Twitter Sentiment Analysis.
4. https://towardsdatascience.com/sentiment-analysis-for-stock-price-prediction-in-python-bed40c65d178

5. https://www.kaggle.com/dgawlik/nyse