Cixin Gao, Lize Chen, Ziyao Tang
Professor Kevin Gold
DS 110
Nov 5, 2022

Research Proposal

Topic to Research:
Stock Prediction Through Using Machine Learning

Background:
Social media, as an integral part of people's daily lives, is widely welcomed by all the phone users. Instagram, one of the most popular social media platforms, is used to get the latest news and posts, or photos and videos, from people cared for by users, including celebrities, family members and friends. Instagram's stock quotes and market prices are influenced by various factors from time to time. Many predictions are made to forecast the future direction of Instagram's stock. Our group's goal is making more reliable predictions of Instagram's future stock movements by using machine learning.

Explanation of why we would like to research this topic:
The financial field is more inclined to combine with data to make stock investment decisions nowadays. More people have begun to explore the direction of using machine learning to predict future stock trends. There are many machine learning methods that can help us to some extent in financial forecasting and predicting stock trends. From this, we are very interested in the algorithms of K-NN and LSTM, and want to compare which of these two algorithms will learn and predict better for stocks. The Federal Reserve Bank decided to further raise interest rates to fight inflation. The higher interest rate has severely hurt the performance of the stock market, especially the IT industry. So we are not optimistic about META's stock trend forecast for a period of time in the future, and want to use the better fit of the two algorithms above to verify this point of view.

Data Sets and Approaches:
To begin with, we break our project down into several steps. Firstly, we decided to use the historical prices and data of Instagram from Yahoo Finance (https://finance.yahoo.com/quote/META/history?p=META). Downloading the primary data from Yahoo Finance, we will clean the data and adjust those data to the format that can be used for different machine learning methods. For future operation of data, k-nearest neighbors and LSTM methods are going to be used during data training and testing. To predict the future data more accurately, we first train the stock's data from 01/01/2019 to 12/31/2021 in both methods, then we are going to test the model using data from 01/01/2022 to 10/31/2022. After plotting the prediction and the real history data, we will compare two plots that are plotted by two methods we use and decide which method is better for META stock. Finally, we will use the better method to predict future stock in the end of 2022 and the first half year of 2023, after training the history data from 01/01/2019 to 10/31/2022, and plot the prediction. Based on the plot we will simply

analyze the trend of the stock to prove whether our hypothesis based on macro perspective the stock will fall is correct or not.

For the k-nearest neighbors method, all columns of the dataset are supposed to be used while predicting, containing the date, open price, close price, high price, low price, and the volume, as these are all crucial elements to stocks. The k, cv, and the parameter grid will be provided also during predictions to achieve better performances. Rest of the parameters will use the default. While for the LSTM method, only the date column and the close price column are necessary to be used. The return_sequence， input_shape, units, loss, optimizer, epoches, batch_size, and verbose will be provided also during predictions to achieve better performances. Rest of the parameters will use the default.