1. Project Title and Team Members

Feature Engineering and Machine learning for Stock Analysis and Prediction

Group 1: Paul Phillips, Qi Cai, Xin Gao

GitHub link at https://github.com/pcp0019/CSCE-5222-Group-1/

2. Idea description

This project is based on the Yahoo Finance Dataset. We will be performing feature engineering work using machine learning models on this dataset. We will perform data visualization, then preprocess the data, then select several financial instruments based on the importance of the features. By using several feature engineering techniques, we can find the relationships between those features and try to analyze the influence of each financial instrument. This way, we can test the performance of the features so that the best features can be selected, improving the quality of analysis and predictions. Afterwards, it will be possible to obtain stock predictions using machine learning models.

3. Goals and Objectives:

Our goal is to use the raw finance dataset to get preprocessed stock data, and after several treatments we can find the relationship between features. Based on this, we may work deeper to discover or predict the performance of the stock in the future. This will in turn enable us to create a more advanced machine learning model for stock price prediction.

4. Motivation

There are a number of factors that affect a stock's price, and these factors change every second. Finding the relationships between these factors is key to determining the value of a stock. This raises the question: how can we know which feature is more important than the other? How can we know how one feature affects another feature? What we need to do is to use feature engineering techniques to get this answer, and alter features as needed to improve predictive modeling. After we find out or get some patterns among them, then we may use these patterns and several regression machine learning models to apply to make stock predictions.

5. Significance

A stock's price changes often, and if we can use feature engineering to improve predictive models it will enable a better understanding about what goes into a stock's valuation and the factors that can cause this valuation to change. This knowledge can be useful for companies looking to improve their present or future stock value and prevent factors that will lead to valuation decline.

6. Literature Survey

Sidra Mehtab, Jaydip Sen & Abhishek Dutta have used NIFTY 50 index values of the National Stock Exchange (NSE) of India and built eight regression models that predicted the open values of NIFTY 50 for the period December 31, 2018 till July 31, 2020. [1]
Rebwar M. Nabi and his research team collected data from NASDAQ and S&P, constructing new features and used the WEKA evaluation method. They finally proposed a new feature engineering approach for stock prediction.[2]

7. Objectives

Perform feature engineering techniques to examine and alter a feature set to create a better machine learning model for predicting stock price.

8. Features

The features are: Adj Close, Close, High, Low, Open, Volume.
We will use these original features to get more complex features by using different mathematical methods such as Certain days MA (moving average price), H-L (High price minus Low price), and O-C (Open price minus Close price).

9. Expected outcome

After this project, we expect to get a cleaned, preprocessed dataset. Also, several histograms to visualize the features and certify the patterns among them. We will also show which features have more impact on the stock price. Another expected outcome is that our use of feature engineering will improve a model for predicting future stock price.

10. References

[1] Mehtab, Sidra, Jaydip Sen, and Abhishek Dutta. "Stock price prediction using machine learning and LSTM-based deep learning models." Machine Learning and Metaheuristics Algorithms, and Applications: Second Symposium, SoMMA 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 2. Springer Singapore, 2021. (Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models | SpringerLink)

[2] Nabi, Rebwar M., Soran AB Saeed, and Abdulrahman MW Abdi. "Feature Engineering for Stock Price Prediction." Int. J. Adv. Sci. Technol 29.12s (2020): 2486-2496.(https://www.researchgate.net/profile/Rebwar-Nabi/publication/342339410_Feature_Engineering_for_Stock_Price_Prediction/links/5eee9f1092851ce9e7f78075/Feature-Engineering-for-Stock-Price-Prediction.pdf)