**ECE 579 Intelligent Systems, Winter 2024  
Project Progress Report**

**Project Title:** MNCs Stock Market Close Price Detection

**Students in the project group:** Sai Sanjith Sivapuram, Devyani Deore

|  |  |  |
| --- | --- | --- |
| **Task** | **Responsibility** | **Status** |
| Data Identification | Sanjith Sivapuram, Devyani Deore | Completed |
|  |
| Literature survey | Sanjith Sivapuram, Devyani Deore | Completed |
|  |
| Data engineering | Devyani Deore | Completed |
|  |
| Data pre-processing | Devyani Deore | Completed |
|  |
| EDA | Sanjith Sivapuram | Completed |
|  |
| Model training and testing | Sanjith Sivapuram, Devyani Deore | In process |
|  |
| Model evaluation | Sanjith Sivapuram, Devyani Deore | Not Started |
|  |
| Final project report | Sanjith Sivapuram, Devyani Deore | Not Started |
|  |

We aim to finish the project by mid-week of April 2024.

**Flow Chart:**

**A diagram of data processing

Description automatically generated**

**Project Description:**

This project focuses on predicting the closing stock prices of multinational corporations (MNCs) through the application of predictive models and data analysis techniques. We aim to enhance prediction accuracy by exploring various methods like normalization, and data smoothing methods.

**Data Description:**

* The project utilizes stock market data sourced from [Kaggle](https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm/input?select=all_stocks_5yr.csv) i.e. the S&P 500 dataset. The data contain around 61k records.
* The dataset is characterized by the following attributes: Date, Open, High, Low Close, Volume, and Name of the company.

**Proposed Method:**

We will employ three distinct predictive models for Stock Price Forecasting:

**Random Forests:** Random forest models leverage multiple decision trees to provide more accurate forecasts from the data. By considering various combinations, these models facilitate increased information gain from the dataset, allowing us to comprehend non-linear relationships within the data.

**ARIMA (Autoregressive Integrated Moving Average):** ARIMA is renowned for its effectiveness in handling time series data, particularly stationary data. Given that our dataset encompasses historical stock prices for diverse companies, the stationary nature of the data makes it well-suited for ARIMA models to deliver robust performance.

**LSTM (Long Short-Term Memory) Neural Network:** LSTM models, built upon recurrent neural networks (RNN), excel in processing time series data. These models adeptly handle non-linear relationships inherent in the data.Notably, LSTM's ability to retain long-term dependencies within the data proves invaluable, especially considering our dataset spans approximately five years.

**Model Evaluation:  
RMSE (Root Mean Square Error):** RMSE is a widely used metric for evaluating the performance of regression models. It measures the average deviation of predicted values from the actual observed values. The lower the RMSE, the better the model's predictive accuracy. RMSE is calculated by taking the square root of the average of the squared differences between predicted and actual values.