**ECE 579 Intelligent Systems, Winter 2024**

**Technology Survey Report**

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

**Names of the students in the group:** Devyani Deore, Sai Sanjith Sivapuram

**Responsibilities of each student:**

1. **Devyani Deore:** Devyani would be involved in Data Engineering, Data Pre-Processing, and partly in the Exploratory Data Analysis. It would be a combination of understanding the data by performing analysis of the structure and the properties transformation, and cleaning data wherever necessary**.**
2. **Sai Sanjith Sivapuram**: Sanjith will perform the other necessary Exploratory Data Analysis to understand the features and behavior of the data and perform appropriate feature extraction and modeling on the data with the retrieval of evaluation metrics for the data.

The members will be equally involved in the documentation and presentation of the project.

**Introduction:**

* Our focus will be on applying various predictive models to stock market data, primarily from multinational corporations (MNCs), and then comparing the performances of the various models. Our approach involves a thorough examination and analysis of diverse stock market datasets from MNCs.
* We aim to understand the different models used, and the feature extraction methods implemented, and explore additional techniques that could be applied to the data. This includes normalization methods and data smoothing techniques, all to enhance the accuracy of predictions on the stock market dataset.
* Pandas, Matplotlib, Long Short-Term Memory (LSTM), and Autoregressive Integrated Moving Average (ARIMA) are certain technologies and libraries involved.
* We are going to use the K-fold technique for normalization and to get the average accuracies of the model from the data.
* We check the RMSE for all the models and conclude the one with the lowest score.

**Description of Technologies Related to the Project:**

1. **Random Forests:**

* Random forest models build on multiple decision trees and enable us to retrieve more accurate forecasts from the data.
* This helps us to get more combinations enabling more information gain from the data.
* These models help us to understand the non-linear relations in the data.

1. **ARIMA (Autoregressive Integrated Moving Average):**

* ARIMA is well known for working with time series data (preferably stationary data).
* As our dataset contains the history of the stock prices for different companies hence the data is stationary and suitable for models like ARIMA to perform well.

1. **LSTM (Long Short-Term Memory) Neural Network:**

* LSTM models are based on RNN (recurrent neural networks) and work well with time series data.
* The main advantage of LSTM is its ability to hold long-term dependencies on the data which is crucial for our data as we have data of around 5 years and models like LSTM take the whole data under consideration.
* These models also handle non-linear relationships well.

1. **Decision Trees:**

* Decision Trees are known to work well with both regression and classification types of data.
* It helps to get information gain and relations between the features in the dataset enabling more accurate predictions.

**Pros And Cons of the Investigated Technologies:**

1. **Pros:**

* Compared to Decision Trees, and Random Forest the other models LSTM and ARIMA have shown better results as they work on finding the relationships in the data and work through time.
* All these models can work on large-dimension datasets and do not showcase the problem of overfitting the dataset.

1. **Cons:**

* Training and performing prediction using LSTM or ARIMA can be expensive concerning computation.
* The dataset needs to be checked for its stationarity to be able to use ARIMA in its best form.
* Retrieving new information from the current data and checking its correlation with the other features.
* Trying to find out the best features from the final form of data (old and new information) for performing prediction for better results.

**Conclusion:**

We conclude by using LSTM and ARIMA as our models for the stock market dataset. We deviate from the regular approach of predicting the closing price by retrieving additional features from the data and using them as part of modeling and predictions. We aim to experiment with different models, compare their results, and conclude an intelligent system to predict the closing price for the given stock market data to the best nearest accurate value.

**References:**

* <https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm/input?select=all_stocks_5yr.csv>
* <https://www.sciencedirect.com/science/article/pii/S1877050920304865>
* <https://www.academia.edu/download/51057076/Z05120207212.pdf>
* <https://www.researchgate.net/profile/Dr-Polamuri/publication/347994783_Stock_Market_Prices_Prediction_using_Random_Forest_and_Extra_Tree_Regression/links/5fec36eb299bf1408859fdf3/Stock-Market-Prices-Prediction-using-Random-Forest-and-Extra-Tree-Regression.pdf>