STOCK PRICE ANALYSIS AND PREDICTION USING MACHINE LEARNING TECHNIQUES AND LSTM

INTRODUCTION:

In this proposed project, I have used entire data life cycle (DLC): Data Acquisition, Exploratory Data Analysis, Data Modeling and Management, Data Analysis, Data visualization and Reporting for the stock price analysis and prediction. As, a data science student it is vital to implement the DLC on any dataset that is used for the analysis. So, I think any proposed project should necessarily follow this life cycle for the prediction.

We aimed to undertake new work in this study by using a Machine Learning technique to forecast or sense the stock behavior tracking of Amazon. Linear regression, Support Vector regression, Decision Tree, Random Forest Regressor, are Machine Learning models. Recurrent Neural Networks (RNN) and the Long-Short Term Memory model (LSTM) that have been successfully used to forecast stock prices and characterize the activity between buyers and sellers of securities. Based on the closing value and stock price, we forecasted the stock price of Amazon. An algorithm with high accuracy is regarded a superior method for predicting stock price when we compare the accuracy of each model.

STATEMENT PROBLEM:

In this paper, I’m forecasting the stock market price using Various ML techniques and the Long-Short Term Memory model (LSTM). I was going to collect the dataset from the Finance.yaaho.com and clean the dataset If it has missing values or numbers. Then, I’m going to do the data analysis, and pre-process the data for feeding the data to machine learning algorithms like Linear regression, Random Forest, Decision Tree, and the Long-Short Term Memory model (LSTM) to predict the stocks using the historical data. Then, I’m going to visualize the stocks by using Tableau. Then I’m going to represent my findings from the dataset, and I will conclude with the results.

LITERATURE REVIEW:

From the recent invasion the stocks have been a quick expansion and expanding possibilities for the market has gained more attention from overseas investors. As the stock market has increased in prominence, researchers have paid increasing attention to studies on various categorization strategies for determining stock performance. According to the research paper, they are two traditional approaches for the stock analysis they are: fundamental and technical analysis whereas technical approach is preferrable due its prediction of the future stocks based on the historical data and to analyze the financial time series (ANN). Due to the technological developments, SMP is using Machine Learning methods for the prediction of patterns in the data. Out of any, models SVM is preferable It's a pattern classification algorithm that minimizes error and improves geometrical tolerances.

According to the research paper, Decision tree and Random Forest are the best algorithm for the prediction of the stocks compared to remaining models. If the amount of data is less than we can use linear regression for the prediction if the data is huge, we should use python and continued with higher level algorithms like LTM and STM. Data-mining approaches (ANN and SVM) are more equipped to identify share price manipulating than multidimensional statistical techniques such as regression models or LR, since data-mining techniques surpass methodologies in classification performance. As per Manoj S Hegde et al. [2], Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) capable of solving involute linear problems, and there is also a discussion on the usage of RNN (Recurrent Neural Networks) to anticipate share prices.

Objectives of the study:

To forecast the stock price of a company using various Machine learning techniques and Long-Short Term Memory model (LSTM) and compare the performance of each model with best performing model.

Research design and methodology:

Data is scraped from the internet for the chosen stock (Amazon) . After the data is collected, the dataset is pre-processed to be in line for the model to consume this data. The cleaned data is now treated for outliers if required. After the pre-processing, data is now fed to the Machine learning models and LSTM based model for training using 80 per cent of the data and the rest of the 20 per cent is used for testing the model. The model is tuned for best performance and now the model is set for testing.

Datasets:

Data is scraped from <https://www.kaggle.com/datasets/vaishubarama/data-set-for-project-info5082>.The data contains features like Date Open, High, Low, Close, Volume, Open Int, Stock. Data is collected for the past 6 months and 80 per cent of this goes for training and the rest for the testing.

Conclusion:

Stock prices are very volatile and depend on various factors. Here we are trying to figure and find out any pattern that exists in the stock price history. So, we will be utilizing the history of the stock prices of the chosen company, to predict or forecast the future prices using an LSTM based model which is known for predicting the use cases with sequential data.

References:

[1] Raut Sushrut Deepak, Shinde Isha Uday, Dr D. Malathi, “Machine Learning Approach In Stock Market Prediction”, IJPAM 2017

[2] M. S. Hegde, G. Krishna and R. Srinath, "An Ensemble Stock Predictor and Recommender System," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1981-1985.

[3] M. Roondiwala, H. Patel and S. Varma, "Predicting stock prices using LSTM," International Journal of Science and Research (IJSR), vol. 6, no. 4, pp. 1754-1756, 2017.

[4] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," PloS one, vol. 14, no. 2, p. e0212320, April 2019.