BIRMINGHAM CITY UNIVERSITY

DATA MINING REPORT

NAME	STUDENT ID
HARDIK N JAIN	23111876
SAIRAJ SALVI	22233637
ABBAS ALI PATEL	22175253
GROUP NO.	09

MODULE CODE	CMP7206
MODULE NAME	DATA MINING
MODULE LEADER	DR. MUHAMMAD AFZAL
DATE OF SUBMISSION	21ST AUGUST, 2023

Contents

1] DOMAIN OF INTEREST:

Stock market determining is a huge and complex field that includes various methods, hypotheses, and areas of interest. The securities exchange is impacted by various variables, from macroeconomic pointers to international occasions, and that's only the tip of the iceberg.

- The field of stock advertising determination can be complex, but it is also always evolving. By analyzing an assortment of components such as the current financial climate, world events, and corporate profit reports, speculators can be different designs and patterns that will help them make educated choices.

- Careful investigation and cautions through of these components are key to weathering the ups and downs of the stock market and accomplishing long-term speculation objectives.

In finance, opinion polls focus on the general mood or expressed in news, financial reports, or virtual entertainment to predict market developments (Nagaya et al., 2009). Quantitative analysis for grasping the way of behaving utilizing numerical and factual demonstrating, estimation, and exploration (Seemann, 2011). This approach includes considering cost design and showcasing patterns utilizing verifiable information, essentially through charts. The point is to figure out future cost developments (Chong and Ng, 2008). Macroeconomics includes considering wide financial measurements like GDP development, intrigue rates, inflations, and unemployment to anticipate advertising courses (CHAN and LAKONISHOK, 1995).

2] PROBLEM DEFINITON:

Stock showcase forecast is the act of attempting to decide the long-standing time value of company stock or other money-related instruments exchanged on a monetary trade. Forecasting stock prices may be a major challenge due to the complexity of components affecting the markets (1--8. and 2011, n.d.). To foresee the long-term closing cost of a given stock over a characterized period, leveraging chronicled information and other significant indicators.

- Accurately predicting the stock market axis on defining the problem at hand. It involves pinpointing specific stocks to concentrate on determining, the timeframe for predictions, and establishing a crystal-clear understanding of the desired level of accuracy(Zhang et al., 2019).

- Without a proper definition of the problem, devising effective prediction strategies and measuring success can prove to be challenging (Obthong et al., 2020).

- Stock costs are in numerous cases affected by subjective news events, making them non-fixed. The money-related trade could be a one-kind system with numerous interrelated variables. Financial circumstances and ways of carrying on alter over the long pull, in a few cases rapidly (Rather et al., n.d.).

Foreseeing securities exchanges is a complex issue that requires a mix of quantitative methods, space skills, and a profound comprehension of the fundamental information sources and market mechanics. Indeed, even with cutting-edge procedures, expectations won't ever be 100 percent exact, and techniques ought to represent chance and unpredictability.

3] LITERATURE REVIEW:

In finance, it has long been difficult to predict stock market movements, which has drawn researchers from a variety of backgrounds. The paradigms of stock market prediction have undergone significant change as a result of the development of sophisticated computational techniques, and deep learning (Hiransha et al., n.d.). Major analysis, calculations, and specialized research all have distinct pros and cons. Before selecting the method that best meets unique needs, one must carefully consider each one. Keeping up with the most recent developments in this area can increase the chances of accurate predictions and trading securities.

- Using patterns and trends from the past to forecast the future, time series forecasting is a popular technique for predicting stock market values (Khan et al., 2022). The efficient market hypothesis in finance contends that is impossible to consistently outperform the market using anticipation.

- Reinforcement learning is a sort of machine learning that is applicable to the stock market and entails teaching algorithms to make judgments based on incentives and penalties (Nabipour et al., n.d.).

- Time-series estimating work on the 'ARIMA' model established a groundwork for grasping time series information for stock cost expectations, notwithstanding, the ARIMA model, which catches direct connections, frequently misses the mark because of the non-straight nature of stock costs. The specialized investigation gives complete procedures utilizing cost and volume to gauge market development (Ariyo et al., 2014).

- It’s a staple for merchants yet has confronted analysis for its subjective nature. The methodology prior to choosing the one that best suits your extraordinary necessities. The most recent progressions in this field can build your possibilities of making precise forecasts and making progress in the securities exchange (Zhang et al., 2019).

Financial exchange expectations strategies have evolved from simple time series models to composite profound learning calculations. An encompassing approach to this mind-boggling task is implied by the consolidation of feeling analysis and hybrids models. Even though progress is being made, the inherent unusualness and excess of influencing factors suggest that maintaining a high level of accuracy in securities exchange expectations will be challenging (Applications and 2021, n.d.). Future research directions could search into more multi-structures, deep learning for methodology exchange, and coordinates various information to create a more comprehensive expectation system.

4] DATASET DESCRIPTION:

The dataset has all the earmarks of being a verifiable record of stock costs for a particular organization, catching day-to-day exchange exercises. From the information, Scrape gave about the organization are NETFLIX(NFLX), TESLA (TSLA), and UBER (UBER).

This dataset used for the:

- Investigate stock cost patterns over the long run.
- Foresee future stock cost utilizing time series techniques or Intelligence methods.
- Compute stock and bring measurements return.
- Comprehend exchanging volume designs.

The dataset catches everyday stocks costs, and without strengthening datasets like macroeconomics markers, organization news, and profit reports, one can't get a total comprehension of the elements impacting the stock cost. Moreover, information for the end of the week and public occasions will be missing since financial exchanges are closed at the moment.

Link to dataset: https://github.com/jainhrk/DATA_MINING_DATA_SET

5] DATASET PRE-PROCESSING:

This article explains the pre-processing handling steps applied to securities exchange information of three organizations: Netflix, Tesla, and UBER. Pre-process is a vital stage in information examination to guarantee the information's quality and resulting investigations or demonstrations. Information Stacking exchange information for each organization was imported from their particular csv documents. Missing values containing invalid or NA values were eliminated and took sufficient values information.

Include scaling to ensure highlights are on a comparative scale, the numeric section open, high, low, and close were standardized and saved as a new CSV document.

The datasets for Netflix, Tesla, and Uber have been effectively prepared to result from examinations such as patterns, time series, and building investigations.

Link to the database after pre-processed:
https://github.com/jainhrk/DATA_MINING_PRE-PROCESSED_DATA

Architecture stock market prediction

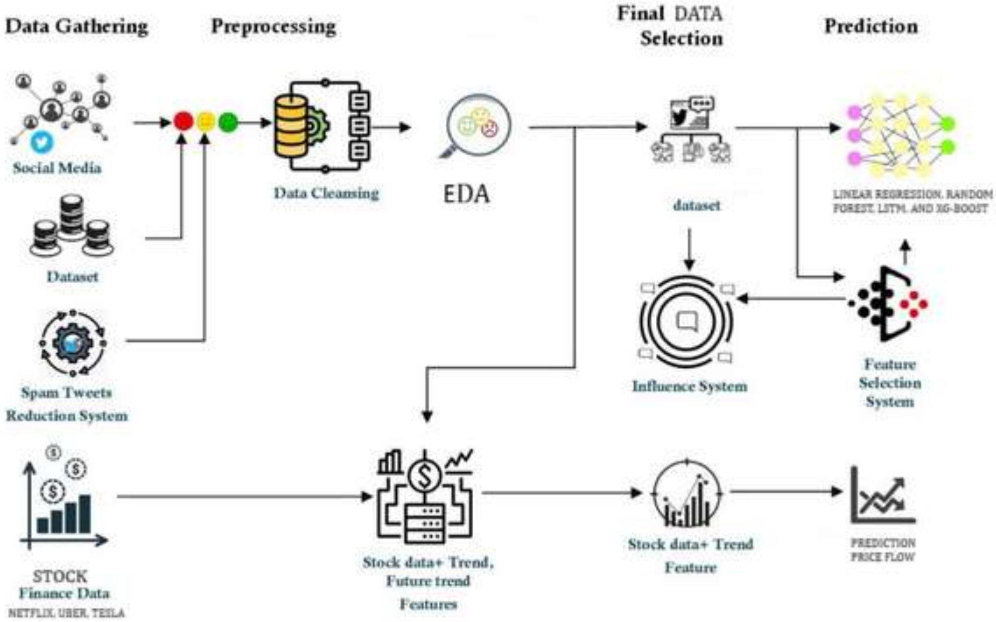

Figure 1 STOCK MARKET PREDICTION ARCHITECTURE

6] EXPLORATORY DATA ANALYSIS

In this segment exploratory information examination on three securities exchange datasets which are TESLA, UBER, and NETFLIX. This will produce four separate portraying the progressions in a particular part of the stock cost on each dataset.

The visualization graphically the way that the stock's open, high, low, and close values have developed over the long run by plotting these outlines. These representations offer a clever examination of the authentic value examples, patterns, and instability of stock (Bhoite et al., 2023).

I. NETFLIX:

- Open cost: This figure represents the progressions in the stock's initial evaluation over the time period. The patterns and examples in the initial cost of the stock by contrasting the blue line, which addresses the open costs at different dates. The x-plot indicates the time regarding dates, and the y-plot shows the stock's open costs in Figure 22.

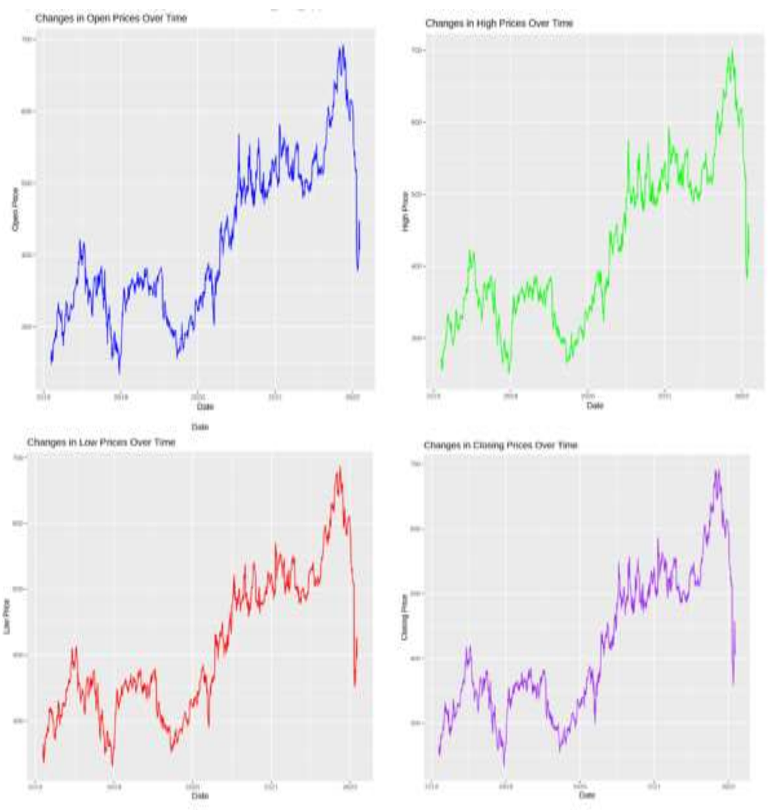

Figure 2 NETFLIX EDA PERFORMANCE

II. TESLA

- Low cost: This chart shows the advancement of the stock's most minimal costs through time. The stock's cost presentation at its bottommost extremes is shown by the red line, which records low costs on different days. The x-plot on this outline addresses dates, and the y-plot addresses the low costs Figure 33.

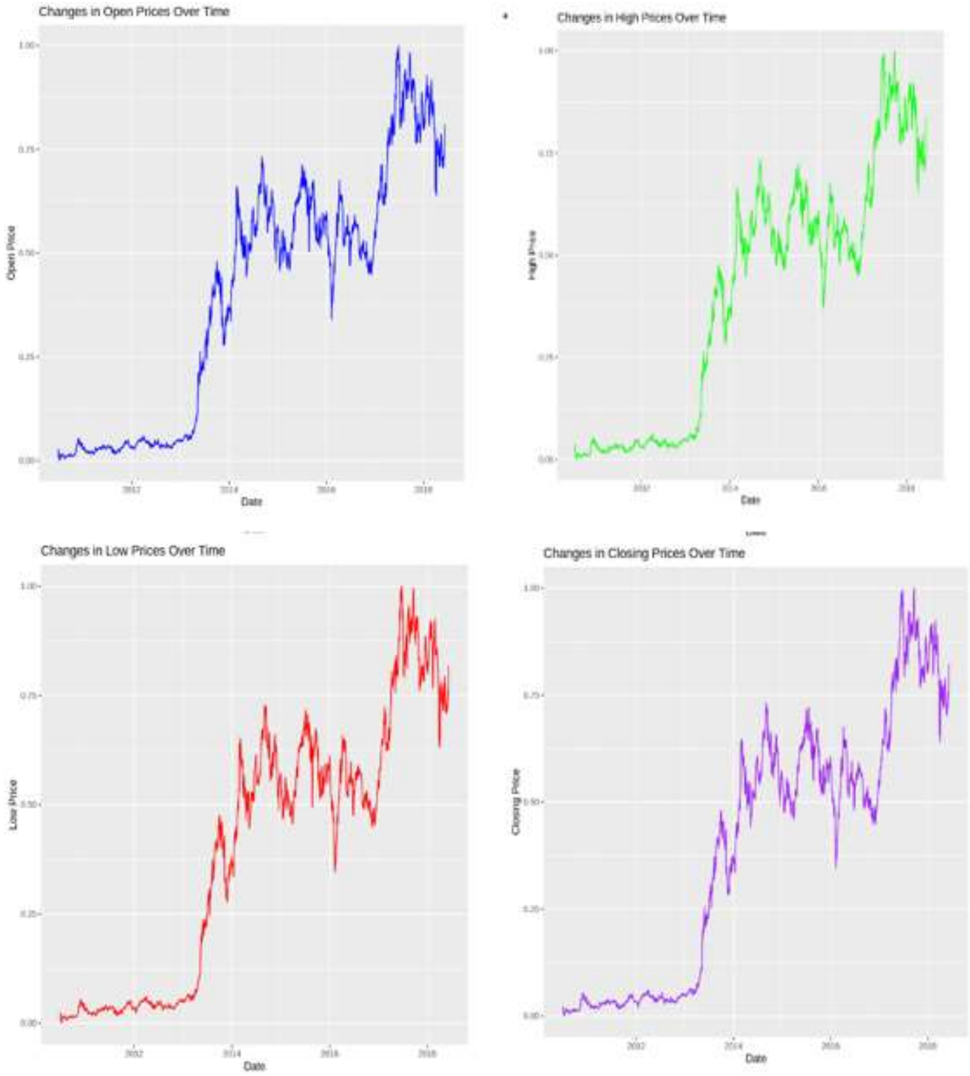

Figure 3 TESLA EDA PERFORMANCE

! Close cost: The last diagram indicates the progressions in the end costs of the stock throughout different time periods. The stock's present at the end of exchanging meetings by taking sessions at the purple line, which shows the end costs of different events. Dates are addressed on the x-pivot, and shutting costs are displayed on the y-pivot in Figure 4 4.

III. UBER (EDA) :

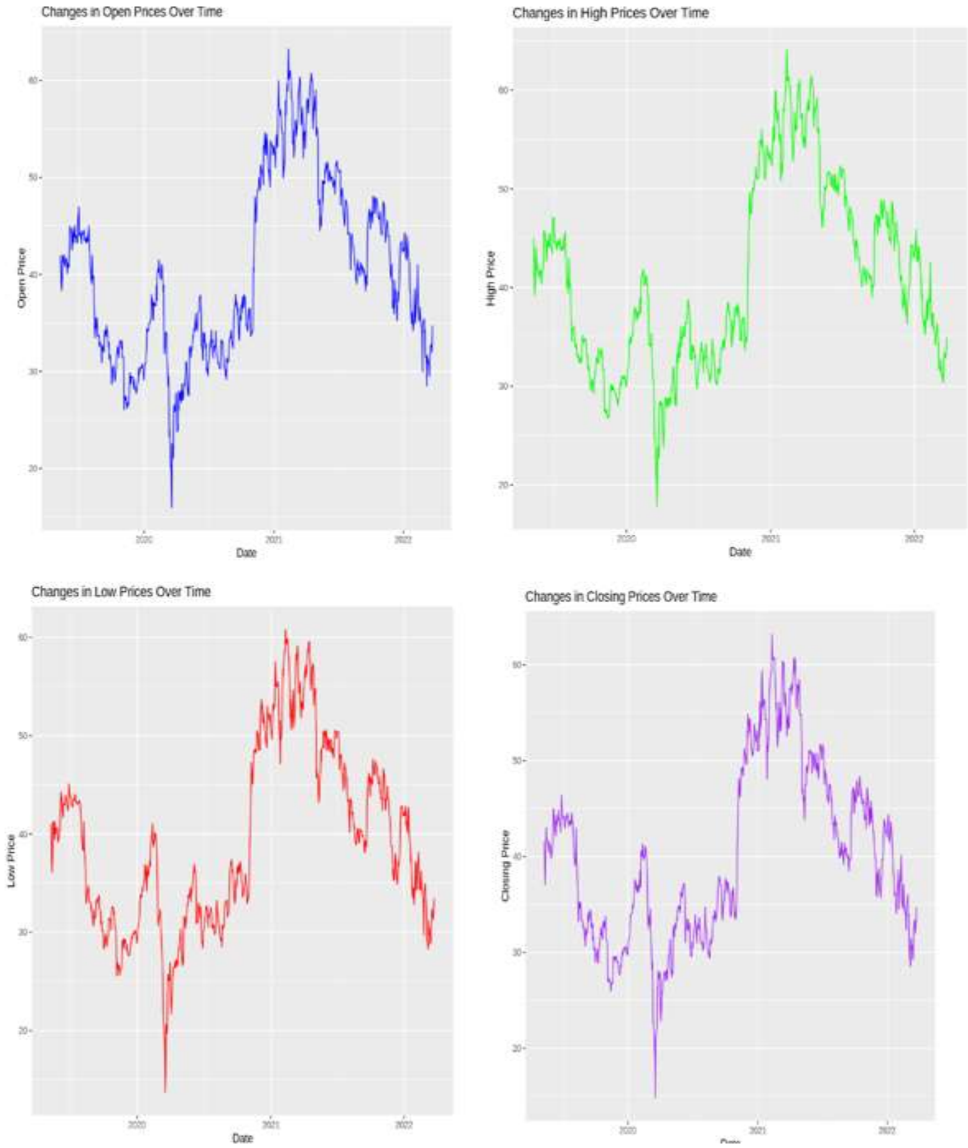

Figure 4 UBER EDA PERFORMANCE

! High Price: The subsequent chart shows the fluctuations in the stock's greatest qualities over the different time spans. The stock's cost exhibition as far as it tops taking at the green line, which records the high costs at different periods. The x-plot on this graph shows the dates, and the y-plot shows the top costs.

7] EXPERIMENTS:

A. RANDOM FOREST

Numerous domains have used Random Forest, an ensemble learning technique renowned for its adaptability and high accuracy. Due to its capacity to handle non-linear relationships, financial analytics has been used to predict stock market movements with encouraging results, frequently outperforming conventional techniques (Luckyson Khaidem et al., 2016).

To deliver a more exact and solid expectation, the random forest calculation makes various choices of trees and joins them. Its strength originates from the way that it develops a 'timberland' of different trees, making it less inclined to overfitting and more qualified to catch perplexing examples.

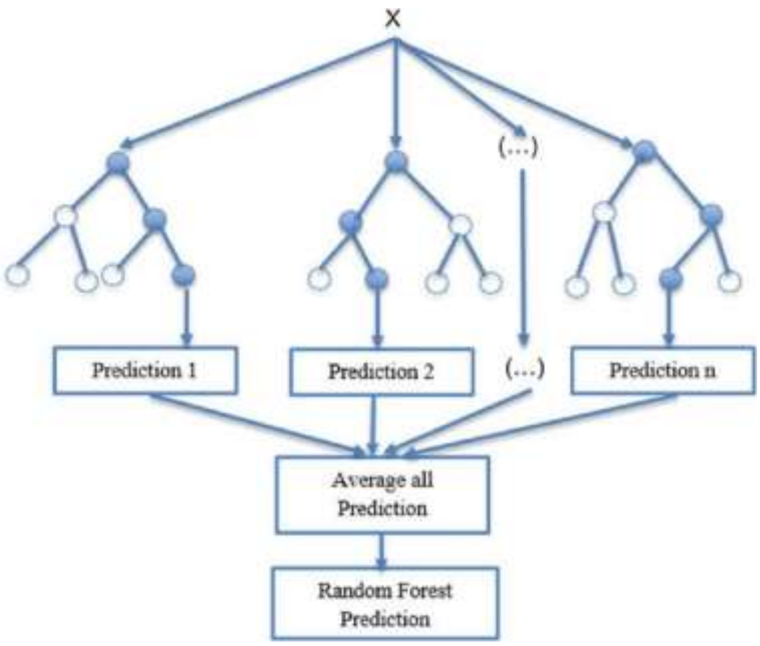

Figure 5 Random Forest for Stock Market Prediction.

REFERENCE – The given Figure 55 is used from (Majumder et al., 2022).

Convectional cross-approval strategies might bring about information spillage because of the successive idea of stock costs. The significance of time-series cross approval for preparing random forest models on securities exchange information has been accentuated in the writing. Forest in stock market prediction, its challenges, and future scope in the rapidly evolving landscape of financial analytics (L Khaidem et al., 2016).

B. LINEAR REGRESSION

The financial exchange forecast utilizing straight relapse rotated around involving past costs or returns as indicators. On the chance markets are really productive, stock costs ought to follow an irregular path and, in this manner, be innately unusual. This represents a fundamental test for utilizing past costs to foresee future ones (Bhuriya et al., 2017). Straight relapse, perhaps the most rudimentary and generally utilized factual procedure, is established in foreseeing a result in light of the connection among dependent and free factors Figure 6.

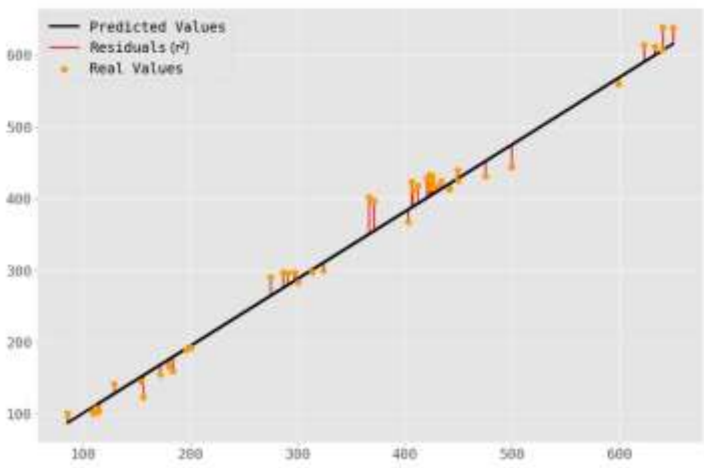

Figure 6 Linear Regression for stock market prediction

Consolidating these experiences from the writing will give a nuanced comprehension of the potential and entanglements of a direct relapse in financial exchange expectations. The exploration scene recommends a nonstop development of procedure, driven by the consistently changing elements of monetary business sectors (Gharehchopogh et al., 2013).

C. Long Short-Term Memory (LSTM)

The unpredictable securities exchange, with its successive and many-sided designs, fills in an alluring space for the use of LSTMs. Long Short-Term Memory (LSTM) networks, a subtype of Recurrent Neural Networks (RNN), have arisen as a main instrument for succession expectation undertakings because of their ability to catch long-haul conditions (Pawar et al., 2019). The unpredictable securities market, with its successive and many-sided designs, fills in as an alluring space for LSTMs.

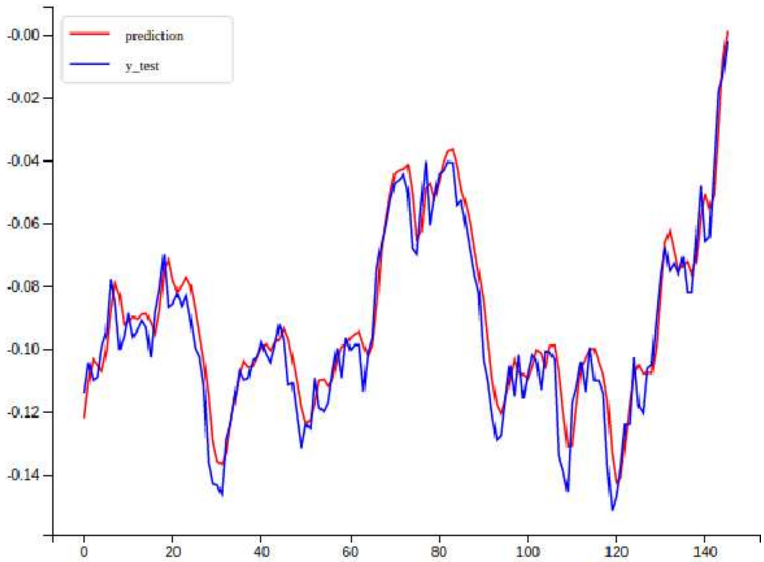

Figure 7 LONG SHORT-TERM MEMORY ON STOCK MARKET PREDICTION

LSTM's capacity to hold data from past information while being delicate to late changes makes them well-suited for securities exchange expectations in Figure 77. One of the challenges of deep learning models like 'LSTM' is overfitting, especially when learning from limited data, which is a common scenario in stock market forecasting. For instance, incorporating LSTM with other brain structures like Convolutional Neural Networks (CNNs) for include extraction or even conventional models (Singh et al., n.d.).

D. XGBOOST

A high-level slope helping implementation called Extreme Gradient Boosting (XGBOOST) is an advanced implementation of gradient boosting, designed for speed and efficiency. Due to the complexity of financial information, XG Boost has gained some traction in financial exchange forecast errands to its ability to handle missing information and catch non-straight connections (Wang and Guo, 2020).

The ability of XG Boost to rank highlights I relation to their importance is one of its strengths. It can provide insights into which indicators have the greatest influence on stock developments in financial exchange forecasts in Figure 88.

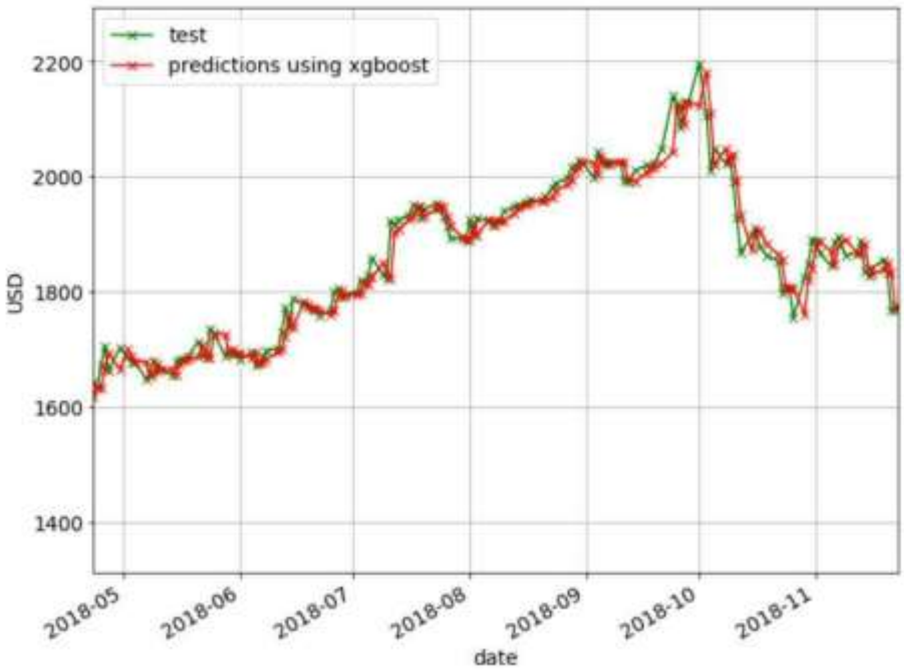

Figure 8 XG-BOOST ON STOCK MARKET PREDICTION

XG Boost is powerful, but the securities exchange forecast poses. Its prescient accuracy can occasionally be constrained by exogenous shocks, black swan occasions, and the inherently chaotic nature of the financial business sector (Dezhkam and Manzuri, 2023). XG Boost, with its slope-helped trees, offers a robust and productive procedure for securities exchange.

8] ANALYSIS & RESULTS:

A. *RANDOM FOREST*

i. TESLA

- The mean squared error is 18.13611 which displays the average difference between the actual result and the predicted result. The representation showcases accurate is the prediction of the MSE. The lower value of dataset which is 18.13611 is considered the prediction is same as the actual data.

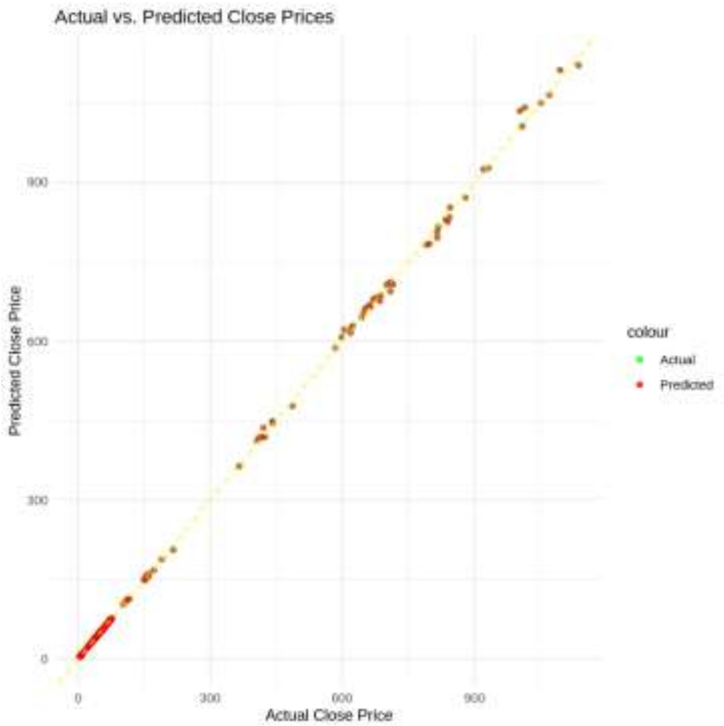

Figure 9 TESLA SCATTER PLOT ON PREDICTION PRICE ON LINEAR REGRESSION MODEL

- The R-Squared value is 0.9997435 in which a number is closer to 1 indicates the model effectively for almost all the variation in the response of the data.
- The predicted future close in which the stock price is expected to be projected at 153.0303 based on the data and technique of the random forest model in the given table. The previous data indicates 143.739627 so the next expected predicted value is going to be 153.0303 as shown in Figure 99.

ii. NETFLIX:

- The average square difference between the actual outcome and the expected result is shown by the mean squared error which is 4.724658 is taken as the forecast and matches the exact data.

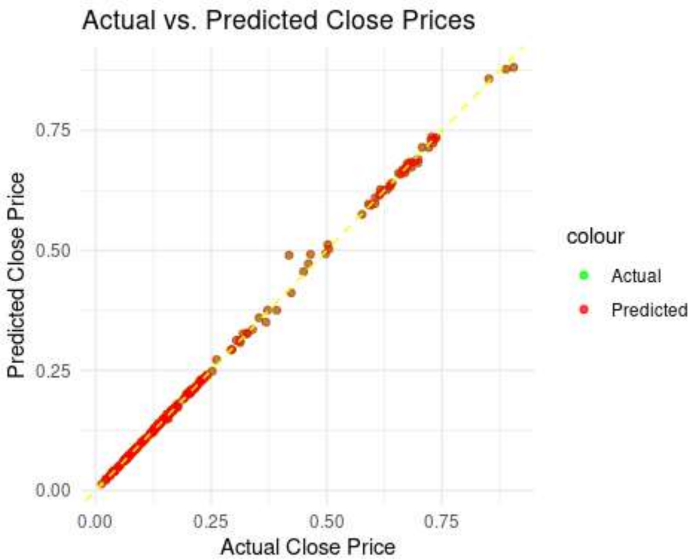

Figure 10 NETFLIX PRICE PREDICTION ON LINEAR REGRESSION MODEL

- The R-squared score is 0.9990947, a value that is near to 1 suggesting that the model successfully explained virtually all of the variation in the data's response. Based on the data from the random forest model approach, the stock price is forecasted to be proposed at 0.8964885 at the predicted future close. The previous data value was 0.62596387, and the projected value is 0.8964885 in Figure 1010.

iii. UBER

- The mean squared error, which shows the typical squared difference between the actual result and the projected outcome is 2.084096 this figure shows how precise the MSE forecast the dataset's lower value is taken as the point at which the forecast and the observed data are identical.

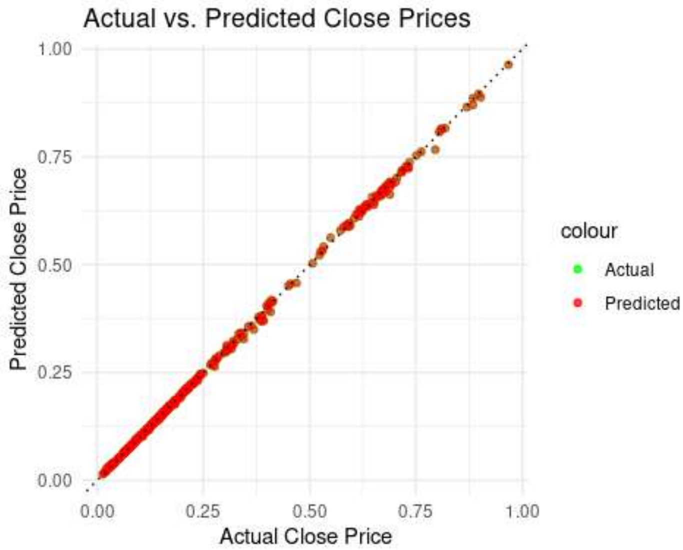

Figure 11 LINEAR REGRESSION ON UBER STOCK PRICE

- The model successfully accounts for virtually all of the variation in the response of the data as indicated by the R-Squared value of 0.9996201 which is near to 1. The values predicted demonstrates the model's excellent performance.

- Using the information and methodology of the random forest model, it is anticipated that at the next close, the stock price will be forecasted to be at 0.9047609 where the previous value was 0.5910511 which anticipated value in Figure 111.

B. LINEAR REGRESSION

i. TESLA:

- The average squared difference between projected and actual closing prices, or mean squared error (MSE), was found to be 1.555227. This demonstrates how closely the model's predictions match the actual values.
- R-Squared (R2): The R2 value, a gauge of how effectively the model accounts for closing price variation, was found to be 1. An R2 value of 1 denotes a perfect fit, with the model appearing to account for all variability.

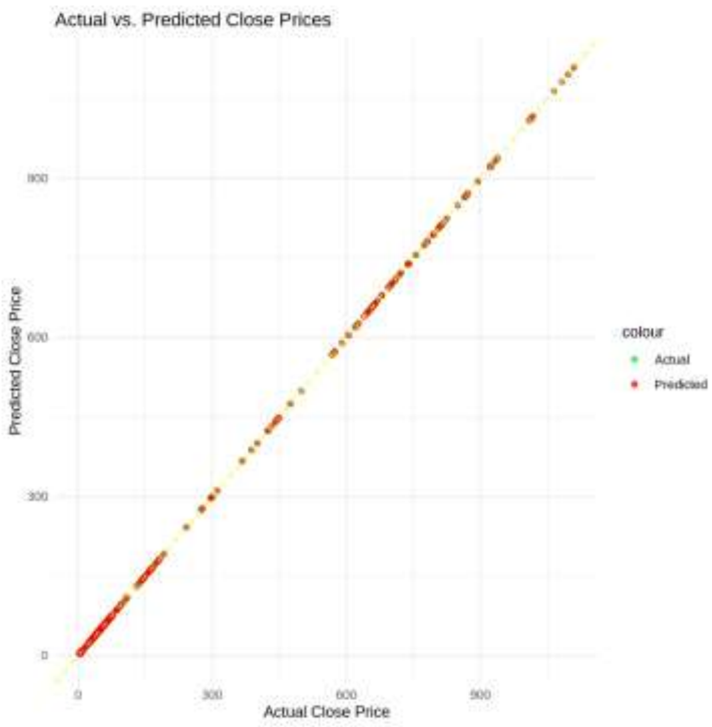

Figure 12 RANDOM FOREST MODEL ON TESLA STOCK ON LINEAR REGRESSION

- From the above Figure 122 scatter plot, we can see that the actual value and the predicted values are overlapping each other which states that the predicted values are accurately same as actual values.

ii. NETFLIX:

- The mean squared error (MSE), or the average squared difference between the forecast and actual closing prices, was discovered to be 1.736988e-06. This shows how closely the model's predictions correspond to the measured values.
- R2: R-Squared: The R2 value, which measures how well the model takes into account closing price volatility, was discovered to be 0.9999673. The model appears to account for all variability when the R2 value is R-squared: 0.9999673, which indicates an almost accurate match with 99% of accuracy.

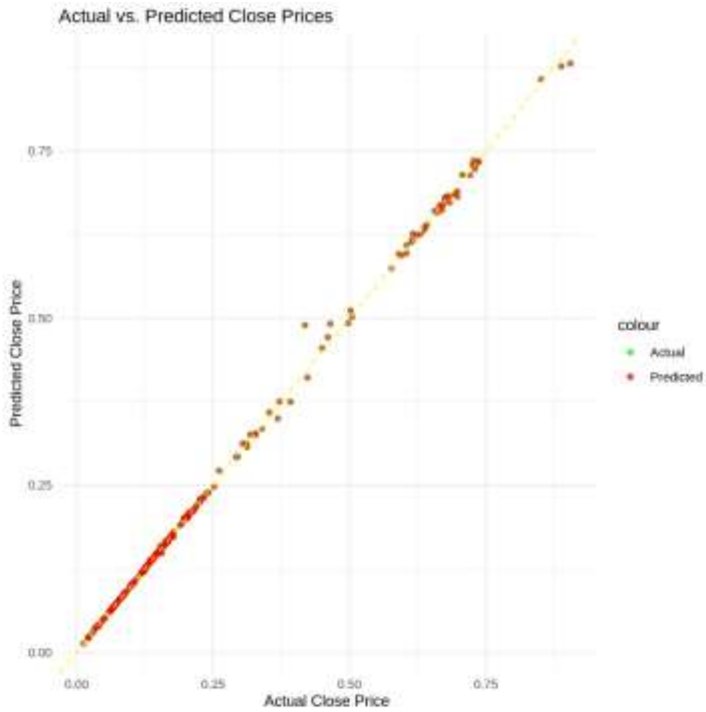

Figure 13 NETFLIX STOCK PRICE GENERATES ON RANDOM FOREST

- The real value and the anticipated value are almost overlapping in the scatter plot below, indicating that the predicted value is a nearly accurate representation of the actual value in Figure 133.

iii. UBER:

- Mean Squared Error (MSE): 1.736988e-06 was determined to be the MSE, which is a measurement of the average squared difference between anticipated and actual closing prices. This incredibly low score shows that the model's predictions and the actual values agree very well.

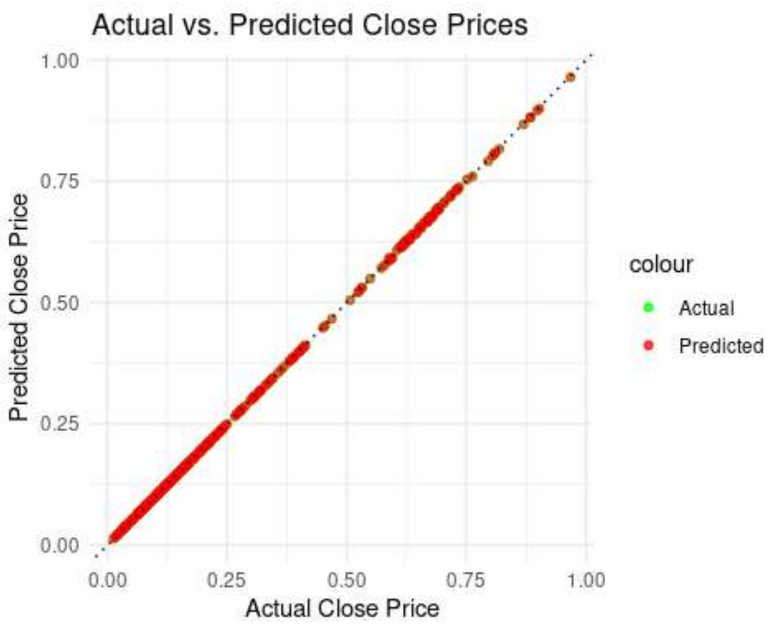

Figure 14 UBER PERFORMANCE ON THE RANDOM FOREST MODEL

- R-Squared (R2): A measure of how effectively a model accounts for volatility in closing prices, the R2 value was found to be 0.99. A high R2 value means that the model effectively accounts for most of the data variation Figure 144.

C. LONG SHORT-TERM MEMORY (LSTM)

i. TESLA:

- Mean Squared Error (MSE): 2.801021 was calculated as the MSE, which is a measurement of the average squared difference between anticipated and actual closing prices. This number represents the difference between the model's predictions and the actual values.

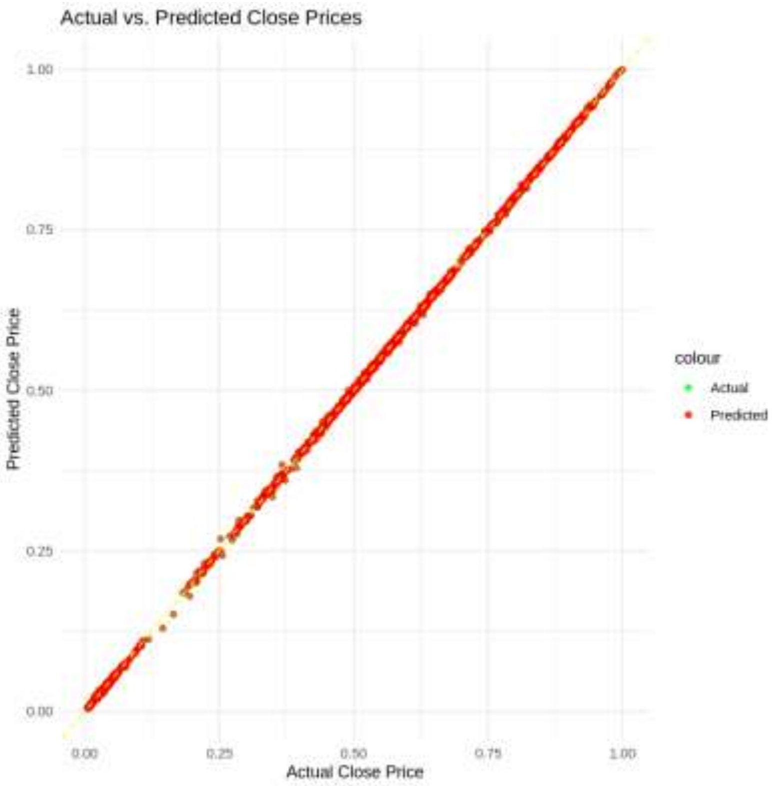

Figure 15 LSTM ON TESLA STOCK PRICE

- R-Squared (R2): The R2 value, which assesses how well the model accounts for the closing price variance, was discovered to be 0.99. This high R2 value illustrates the model's capacity to explain a sizeable percentage of the data variation Figure 155.

ii. NETFLIX:

- Mean Squared Error (MSE): 6.890584e-06 was calculated as the MSE, which is a measurement of the average squared difference between anticipated and actual closing prices. This shows a very slight difference between the predicted values and the actual ones.

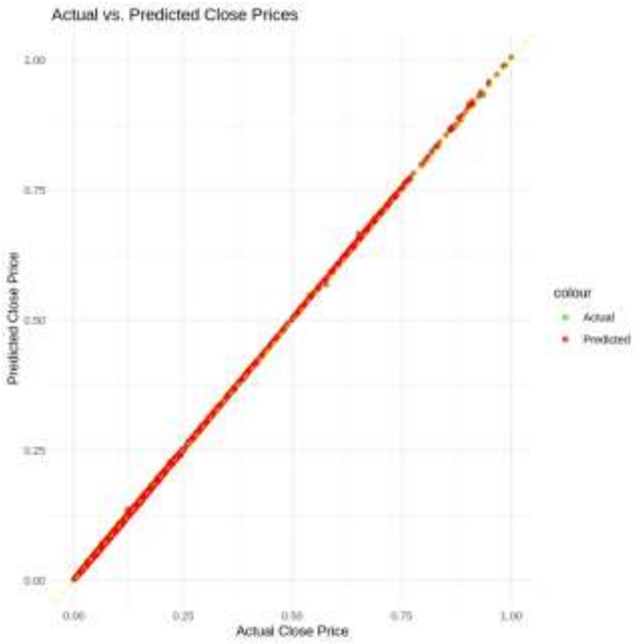

Figure 16 NETFLIX STOCK PRICE ON LSTM MODEL

- R-Squared (R2): The R2 value, which measures how well the model can account for the variation in closing prices, was discovered to be roughly 0.9999634. This indicates that the model successfully captured a very high level of variance Figure 166.

iii. UBER:

- Mean Squared Error (MSE): A measurement of the average squared difference between the anticipated and actual closing prices, the estimated MSE was 1.334957e-06. This suggests that there was very little forecast inaccuracy.

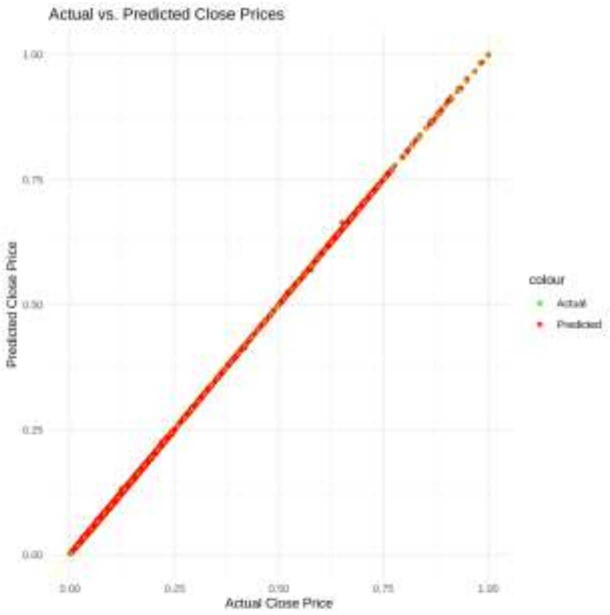

Figure 17 LSTM MODEL PERFORM ON THE UBER STOCK DATA

- R-Squared (R2): An estimate of 0.999978 was made for the R2 value, which measures how well the model can account for the variation in the closing price. This shows that the model has a very high level of variance covered Figure 177.

D. XGBOOST:

i. TESLA:

- In this following analysis of TESLA, the present of XG-Boost model prepared on the securities exchange information with two principles metric Mean Squared Error (MSE) and R-Squared. The closer to the MSE is to nothing, the more precise and better model's expectations are considered.

- A worth moving towards 1 demonstrates the model has almost consummated its illustrative power on the information in the given Figure it shows the predicted closed price on x-plot and actual closed priced on y-plot in which the blue dots show the actual and red dot shows the predicted values Figure 18.

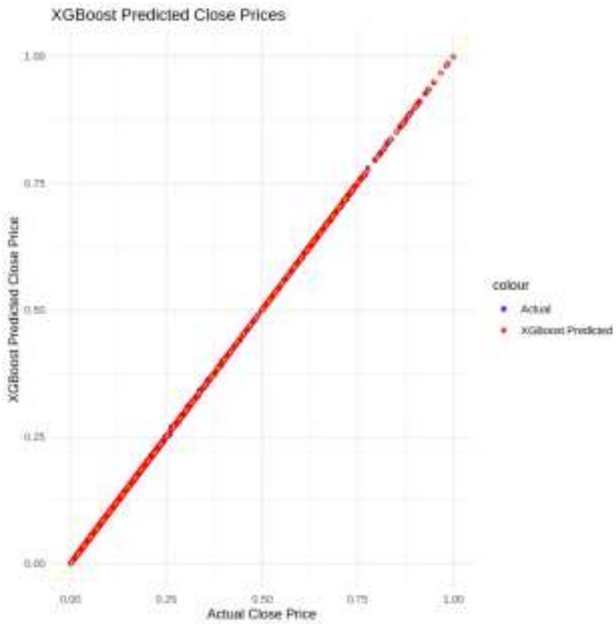

Figure 18 XG-BOOST MODEL USED ON TESLA STOCK DATA

- This microscopic MSE lingerie the deviations between the model's expectations and connoting a model of remarkable exactness. A R-Squared worth of 0.9999912 declares the model explains around 99.99% of the difference in the objective variable in the given Table 1.

Table 1 MSE AND R-SQUARED VALUES FOR TESLA STOCK

Mean Squared Error	7.99464
R-Squared	0.9999912

ii. NETFLIX:

- This analysis represents a top bottom examination of XG-Boost model custom-made at foreseeing the stock costs of Netflix, utilizing two key metrics Mean and R-Squared. This metrics works out the normal of the squared inconsistencies between the perception and the expectations given by the model.

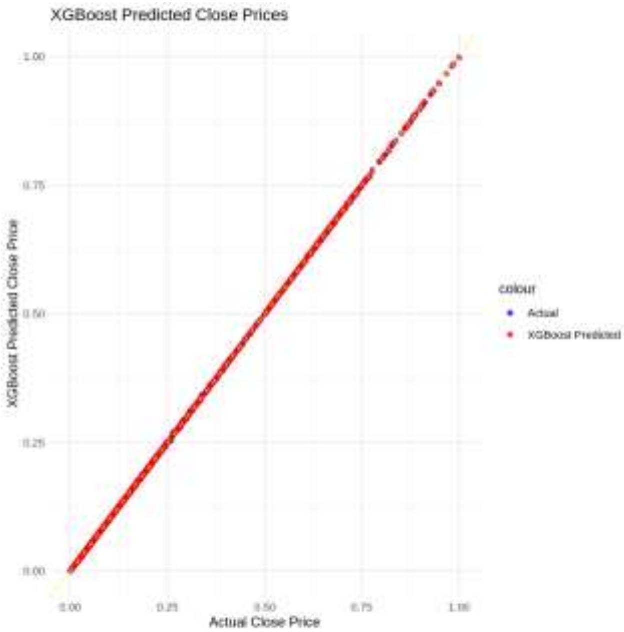

Figure 19 NETFLIX LSTM MODEL STOCK PRICE

- A R-squared esteem poking towards 1 implies a model that makes sense of a significant piece of the information's fluctuation in the given figure where the XG-Boost shows the actual and predicted vales for the organization Figure 19.

- Such a low MSE recommends that model's expectations at Netflix stocks costs are strikingly near the genuine qualities, featuring its accuracy, with of 0.9999852, the outlines come to 99.99% of the fluctuation in Netflix's stock costs and the table shows the MSE and R-Squared values Table 2.

Table 2 NETFLIX STOCK VALUES ON XG-BOOST MODEL

Mean Squared Error	8.107431
R-Squared	0.9999852

iii. UBER

- In this analysis, the present of XG-Boost model prepared on securities exchange information utilizing two essential measurement Mean Squared Error and R-Squared. The typical squared distinction between the noticed real qualities and the anticipated by the model. A R-Squared esteem near 1 proposes that a huge of changeability in the result has been made sense by the model of organization 'UBER'.

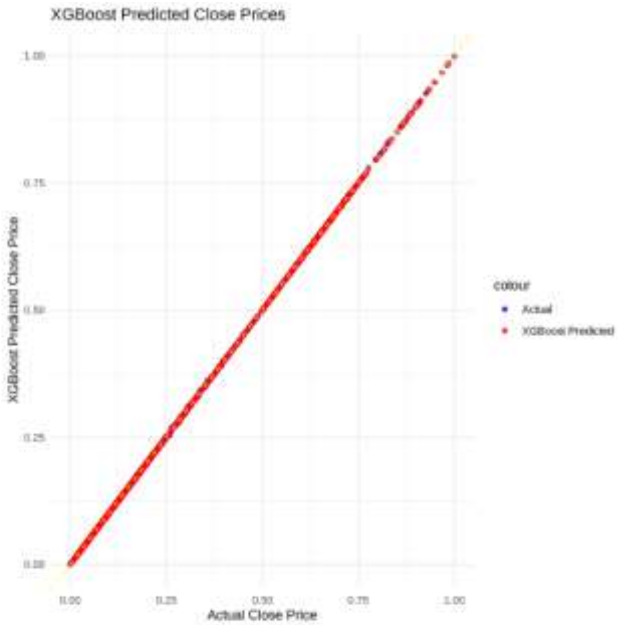

Figure 20 EG-BOOST ON UBER STOCK DATA

- The very low MSE esteem proposes that EG-Boost model’s expectations are exceptionally precise, with minute mistakes when contrasted Figure 20.

- The R-Squared worth of around 1, for this situation 0.9999852, shows that model can make sense of around 99.9% of the fluctuation in our objective variable in the given Table 3.

Table 3 UBER MSE AND R-SQUARED VALUES ON XG-BOOST

Mean Squared Error	9.231431
R-Squared	0.9999952

9]COMPARSION OF TECHNIQUES:

Companies	TESLA (MSE)	NETFLIX (MSE)	UBER (MSE)	TESLA (R-SQUARED)	NETFLIX (R-SQUARED)	UBER (R-SQUARED)
RANDOM FOREST	18.13611	4.274658	2.084096	0.9997435	0.9990947	0.9996201
LINEAR REGRESSION	1.555227	1.736988	1.736988	1	0.9999673	0.9999673
LONG SHORT-TERM MEMORY(LSTM)	6.115383	6.890584	1.3349	0.9999346	0.9999634	0.999978
XG-BOOST	7.99464	8.107431	9.231432	0.9999912	0.9999852	0.9999952

Table 4 COMPHERSION TABLE OF TECHNIQUES

In the above Table 4 it shows the Mean Squared Error (MSE) and R-Squared for three organization where the deep learning and techniques are used.

- Tesla

The Mean Squared Error (MSE) was used as a performance parameter in the evaluation of forecasting models for Tesla's stock market. The Linear Regression model had the lowest MSE, indicating the greatest performance among the investigating models when the MSE values were used as a gauge of prediction accuracy. This implies that Linear Regression model beat the other models, such as Random Forest, XG-Boost, and LSTM in predicting Tesla's stock values on MSE criterion.

- NETFLIX

The MSE was used as a performance among the analysed models, the Linear Regression model showed the lowest MSE, indicating the best performance when using the MSE values as gauge of prediction accuracy. Thus, Linear Regression beats the other models in the NETFLIX dataset also, predicting the price of stock.

- UBER

The LSTM model showed the lowest MSE among the investigated models, indicating the best performance when using MSE values as a gauge of prediction accuracy. In order to predict the stock prices of UBER, the LSTM model performed better than the other models, according to criterion.

From the above three analysis we can state that Linear Regression outperformed in the two of the analysis which are TESLA and NETFLIX and third analysis the LSTM outclasses all other three techniques with lowest of MSE.

10]CONCLUSION:

By initiating multiple approaches, such as series analysis and machine learning, stock market prediction using data mining techniques has the potential for uncovering hidden patterns and trends. Consistent long-term projections are put to the test stock market volatility and the efficient market hypothesis.

For predicting overall four models were used which were mainly Linear Regression, Random Forest, Long Short-Term Memory (LSTM), and XG-Boost. These techniques helped out the better prediction of the future value and gain the maximum profit in the portfolio. Implementing these techniques is the best techniques as provided maximum accuracy and the closet future predicted value.

Overfitting is crucial when models were accomplished well with old data and poorly with fresh data. Although data mining in stock market prediction can help with decision-making, it shouldn't be the only contemplation when buying stocks. Understanding market dynamics and also combining the insight of the data with professional judgement are very important and disciplined investment never cause any risk factor throughout the prediction of the stock market.

11]REFERENCES:

1--8., Z.X.-J.-J. of C.Science., 2011, undefined, n.d. Twitter mood predicts the stock market. cir.nii.ac.jp.

Applications, W.J.-E.S. with, 2021, undefined, n.d. Applications of deep learning in stock market prediction: recent progress. Elsevier.

Ariyo, A., Adewumi, A., 16th, C.A.-2014 Uks.-A., 2014, undefined, 2014. Stock price prediction using the ARIMA model. ieeexplore.ieee.org. https://doi.org/10.1109/UKSim.2014.67

Bhoite, S., Ansari, G., Patil, C.H., Thatte, S., Magar, V., Gandhi, K., 2023. Stock Market Prediction Using Recurrent Neural Network and Long Short-Term Memory. Lecture Notes in Networks and Systems 520, 635–643. https://doi.org/10.1007/978-981-19-5331-6_65

Bhuriya, D., Kaushal, G., Sharma, A., Singh, U., 2017. Stock market predication using a linear regression. Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017 2017-January, 510–513. https://doi.org/10.1109/ICECA.2017.8212716

CHAN, L.K.C., LAKONISHOK, J., 1995. The Behavior of Stock Prices Around Institutional Trades. J Finance 50. https://doi.org/10.1111/j.1540-6261.1995.tb04053.x

Chong, T.T.L., Ng, W.K., 2008. Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. Appl Econ Lett 15. https://doi.org/10.1080/13504850600993598

Dezhkam, A., Manzuri, M.T., 2023. Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. Eng Appl Artif Intell 118, 105626. https://doi.org/10.1016/J.ENGAPPAI.2022.105626

Gharehchopogh, F.S., Bonab, T.H., Khaze, S.R., 2013. A LINEAR REGRESSION APPROACH TO PREDICTION OF STOCK MARKET TRADING VOLUME: A CASE STUDY. International Journal of Managing Value and Supply Chains (IJMVSC) 4. https://doi.org/10.5121/ijmvsc.2013.4303

Hiransha, M., Gopalakrishnan, E., … V.M.-P. computer, 2018, undefined, n.d. NSE stock market prediction using deep-learning models. Elsevier.

Khaidem, L, Saha, S., arXiv:1605.00003, S.D. preprint, 2016, undefined, 2016. Predicting the direction of stock market prices using random forest. arxiv.org 00, 1–20.

Khaidem, Luckyson, Saha, S., Dey, S.R., 2016. Predicting the direction of stock market prices using random forest.

Khan, W., Ghazanfar, M.A., Azam, M.A., Karami, A., Alyoubi, K.H., Alfakeeh, A.S., 2022. Stock market prediction using machine learning classifiers and social media, news. J Ambient Intell Humaniz Comput 13, 3433–3456. https://doi.org/10.1007/S12652-020-01839-W

Majumder, A., Rahman, M.M., Biswas, A.A., Zulfiker, M.S., Basak, S., 2022. Stock Market Prediction: A Time Series Analysis. Smart Innovation, Systems and Technologies 235, 389–401. https://doi.org/10.1007/978-981-16-2877-1_35

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Entropy, E.S.-, 2020, undefined, n.d. Deep learning for stock market prediction. mdpi.com. https://doi.org/10.3390/e22080840

Nagaya, S., Chenli, Z., Hasegawa, O., 2009. An associated-memory-based stock price predictor, in: Lecture Notes in Computer

Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-04277-5_35

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., Wills, G., 2020. A survey on machine learning for stock price prediction: Algorithms and techniques.

Pawar, K., Jalem, R.S., Tiwari, V., 2019. Stock Market Price Prediction Using LSTM RNN. Advances in Intelligent Systems and Computing 841, 493–503. https://doi.org/10.1007/978-981-13-2285-3_58

Rather, A., Sastry, V., Opsearch, A.A.-, 2017, undefined, n.d. Stock market prediction and Portfolio selection models: a survey. Springer.

Seemann, T., 2011. The Effect of Stock Endowments on the Liquidity of Prediciton Markets. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1188877

Singh, V., Marcel, R.·, Durgesh, W., Mishra, K., Amit, ·, Shikha, J., Editors, M., n.d. Stock market price prediction using LSTM RNN. Springer.

Wang, Y., Guo, Y., 2020. Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. China Communications 17, 205–221. https://doi.org/10.23919/JCC.2020.03.017

Zhang, X., Li, Y., Wang, S., Fang, B., Yu, P.S., 2019. Enhancing stock market prediction with extended coupled hidden Markov model over multi-sourced data. Knowl Inf Syst 61, 1071–1090. https://doi.org/10.1007/S10115-018-1315-6

12]APPENDIES:

SECTION AND TASK FULFILLED BY HARDIK JAIN 23111876

- DOMAIN OF INTEREST
- DATA PRE-PROCESSING
- ANALYSIS AND RESULT – RANDOM FOREST
- R MARKDOWN FILE UBER

SECTION AND TASK FULFILLED BY SAIRAJ SALVI-22233637

- PROBLEM DEFINITION
- DATA DESCRIPTION
- ANALYSIS AND RESULT – LONG SHORT-TERM MEMORY (LSTM)
- R MARKDOWN FILE NETFLIX

SECTION AND TASK FULFILLED BY ABBAS ALI PATEL-22175253

- LITERATURE REVIEW
- EXPLORATORY DATA ANALYSIS
- ANALYSIS AND RESULT – LINEAR REGRESSION
- R MARKDOWN FILE TESLA

THE REMAINING SECTION AND TASKS LIKE ANALYSIS AND RESULT ON XG-BOOST, COMPARISON OF TECHNIQUES, CONCLUSION, AND REFERENCES ARE FULFILLED BY EVERY TEAM MEMBER IN THE GROUP.