PredictionsofTeslaStockPricebasedonLinear Regression, SVM, RandomForest,LSTM and ARIMA

SATHISHKUMAR R[1],THIRUESWARAN V[2]

[123]*ComputerScienceandEngineeringDepartment,*

Rajalakshmi Engineering College Chennai

210701515@rajalakshmi.edu.in

210701290@rajalakshmi.edu.in

ABSTRACT - The stock market of a country is an important financial market. A booming stock market promotes the effective use of social capital, the prudent deployment of economic resources, and the expansion of the country's macro economics. Making more informed decisions as an investor is made possible by the development of trustworthy equity market models. A trading model allows market participants to select corporations that pay the highest dividend payments while lowering the risks associated with investing. However, batch processing methodologies make stock market research more challenging as a result of the strong connection between stock prices. The advent of technological achievements like universal digitization has elevated share market forecasting into a highly advanced age. Through the research and comparison of several methodologies, this article tries to discover the most accurate approach for predicting Tesla stock closing prices.Predictions are made using statistical approaches such as ARIMA and machine learning methods such as SVM, LinearRegression, RandomForests, and LSTM. Following athorough examination of all approaches, it was discovered that the accuracy of machine learning methods in predicting stocks is higher than that of statistical methods and integrated algorithm technologies like Random Forest have excellent anti-interference and anti-overfitting characteristics, which are more suitable for evaluating high- volatility stocks like Tesla.

Keywords - Tesla; stock market;machine learning; Future price prediction.

1. Introduction

The stock market is a group of stockbrokers and dealers who buy and trade stock shares. A stock exchange is where many large firms' shares are traded. This improves the stock's liquidity, making it more enticing to investors [1]. A vast number of investors put large sums of money into the stock market. However, it is risky since stock values may quickly rise or decrease [2]. Because of this, forecasting stock prices is a difficult subject on which many scholars are working. To predict stock prices, a variety of approaches have been utilized. Because statistical methodologies are linear in nature, they perform poorly in the event of a rapid spike or decrease in stock prices. Because stock data is unpredictable, volatile, non-stationary, and dependent on a variety of technical characteristics, statistical techniques have been proven to be insufficiently accurate [3]. The open stock exchange trades stocks of publicly listed companies, whereas the private share exchange trades stocks of privately held businesses. Investments made through the mixed-ownership share trade are made in companies with ordinary shares that may only be exchanged publicly on rare occasions. Several countries, like the UK's London Stock Exchange and the US's New York Stock Exchange, have stock exchanges with mixed ownership [4-9].

Thus, all previous work has been centered on reducing some metric that leads to estimates near the true stock price. The experiment first collects data from Kaggle. Reprocess the data to remove any null values in the time series. And transform and normalize the data to make it compatible with the input of the model. Then use Matlab to implement the Linear Regression model , SVM model Random Forest model, LSTM model, and ARIMA model. Finally, compare the performance of different models and propose further work and project expansion.The purpose of the experiment is to successfully realize the linear regression model, SVM model, Random Forest model, LSTM model, and ARIMA model of Tesla stock data. Finally, assess the model's accuracy and provide any updated recommendations.The model is trained and tested using daily data from Tesla stock from June 29, 2021, to July 12, 2022. The data collection consists of 6 columns and 3031 rows. Data preparation comprises looking form is sing values and removing them from the data collection, as well as looking for category values and removing extraneous information from the data source. Training data and test data make up the two sections of the data set. In this case, 2121 data points were considered as training data, while the remaining 909 data points were maintained for testing. The training data includes 2121 days, from June 29, 2010, to September 7, 2018, while the test data extends 909 days, from July 10, 2010, to July 12, 2022. Furthermore, to reduce the variable ranges, all the data is scaled with a typical scaler. Data from various methodologies can be scaled and compared in similar situation.

2. APPROACH

A basic statistical technique called linear regression creates a linear relationship between the independent variables, or features, and the dependent variable, or the price of Tesla's shares. It aims to capture the pattern and directionality of price movements over time by fitting a line to past data points.However, Linear Regression assumes a linear relationship, which may not adequately capture the inherent non-linearity and complexity of stock price dynamics, especially in the case of Tesla, where market sentiment and innovation play significant roles.

The powerful supervised learning method Support Vector Machines (SVM) looks for the hyperplane that divides data points into the best classes. SVM looks price movements in the context of stock price prediction.By mapping historical data to a higher-dimensional space, SVM can potentially capture intricate relationships that linear models may overlook. However, SVM's performance heavily relies on parameter tuning and kernel selection, which can introduce challenges in optimizing its predictive accuracy.

Random Forest,a powerful ensemble learning technique, builds numerous decision trees and combines their projections to reduce over fitting and increase accuracy. Through the utilization of numerous trees collective wisdom,Random Forest can precisely capture the vast array of variables influencing Tesla's stock price., including market sentiment,investor behavior,and macro economictrends.

Long Short-Term Memory (LSTM) networks, a subtype of recurrent neural networks (RNNs), are appropriate for time-series forecasting applications, which are intended for sequential data processing, such as stock price prediction. By maintaining memory for lengthy periods of time, In the past stock values of Tesla, LSTMs are able to identify subtle patterns and temporal connections, enabling more accurate predictions even in the presence of volatility and irregularities.

Autoregressive Integrated Moving Average (ARIMA) models, on the other hand, concentrate on modeling the intrinsic temporal dependencies in the time series data. ARIMA can efficiently capture data by taking into account the auto-regressive (AR), differencing (I), and moving average (MA) components, the underlying trends, seasonality, and unpredictable variations impacting Tesla's stock price, offering insightful information for short- to medium-term projections.

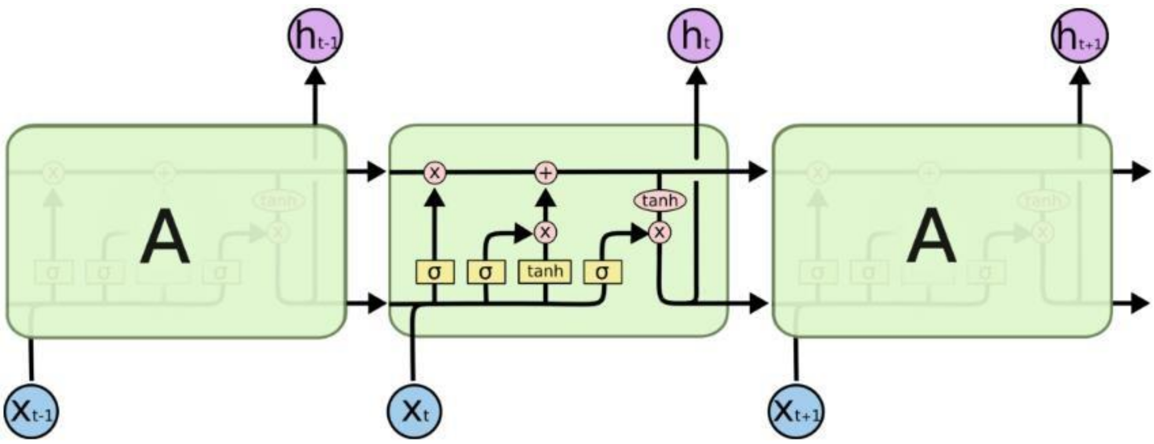
3. PROPOSED WORK

The prediction of Tesla's stock price by including a variety of machine learning and statistical models offer an alluring avenue for in-depth investigation. To predict when Tesla's stock will rise price with accuracy and dependability, the suggested work intends to take advantage of the advantages of Linear Regression, Support Vector Machines (SVM), Random Forest, Long Short-Term Memory (LSTM) networks, and Autoregressive Integrated Moving Average (ARIMA) models.

To guarantee data quality and consistency, the dataset containing past Tesla stock prices and pertinent predictors will first be gathered and preprocessed. Features engineering techniqes will be utilised to extract helpful featurs, which capture the basic mechanics of the changes in Tesla's stock price. For the purpose of preparing the data inputs for further modeling, this preprocessing phase is essential,with each of the selected algorithms.

Following data preprocessing, the proposed work will involve the implementation and evaluation of each model individually. Linear Regression will establish a baseline predictive model by demonstrating the linear association between the chosen predictors and the price of Tesla's stock. After that,SVM will be used to simulate complicated patterns and non-linear correlations in the data,providing reliable forecasts in challenging market circumstances. Random Forest will use group dynamics. learning to aggregate predictions from multiple decision trees,thereby enhancing predictive accuracy and generalizationcapabilities.

Subsequently, Long Short-Term Memory (LSTM) networks will be utilized to capture temporal dependencies and sequential patterns in Tesla's stock price data. LSTM's ability to retain memory It is ideally suited for time-series forecasting applications because it can capture both short- and long-term changes and trends over extensive periods of time.Combining the forecasts from each model, the suggested job aims to offer a thorough and strong structure for predicting. Tesla's stock price, facilitating informed decision-making in financial markets.



4. SystemProcessFlow:

The system process flow for predicting Tesla's stock price using Linear Regression, SVM, Preprocessing and data gathering are the first steps in the Random Forest, LSTM, and ARIMA models. Historical stock prices of Tesla, along with other relevant financial indicators and market data, are gathered from reliable financial databases. This raw data undergoes cleaning to handle missing values, outliers, and inconsistencies. The basic mechanics of stock price fluctuations are then captured by informative features that are created through the application of feature engineering approaches. Moving averages, trade volume, volatility indices, and other technical indicators are a few examples of these qualities.

To verify the robustness of the models, the dataset is divided into training and testing sets after the data has undergone preprocessing. The first is linear regression, implemented to establish a baseline model by fitting a linear relationship between the features and Tesla's stock price. This involves calculating the regression coefficients that minimize the difference between the predicted and actual stock prices in the training set. The model's performance is then evaluated using the testing set to assess its predictive accuracy.

Random Forest models and Support Vector Machines (SVM) are implemented as part of the ongoing process.By converting the data into a higher-dimentional space and finding the perfect hyperplane that splits different price movements, SVM is use to model non-linear interactions. Random Forest, an ensemble learning method, builds multiple decision trees using Various subsets of the data and integrating their forecasts to enhance precision and minimize overfitting. Both models are evaluated on the testing set after being trained on the training set in order to compare their results to the baseline Linear Regression model.

Long Short-Term Memory (LSTM) networks and Autoregressive Integrated Moving Average (ARIMA) models are put into practice in the next stage. LSTM networks, that is a type of recurrent neural netwok (RNN), is particulary fitting for time-series forecsting due to its capacity to retain long-term dependancies. The LSTM models are trained on sequential datas to capture temporal paterns and trends in Tesla's share price. ARIMA models are used on the time-serie data to include auto-regressive, differencing, and moving average elements, efectively modeling the price dinamics. modeling the stock price dinamics. In order to give a thorough and reliable framework for predicting Tesla's stock price, the forecasts from all models are finally combined, maybe using an ensemble technique. This ensures that each model's advantages are maximized while minimizing its disadvantages.

5. Experimental Result

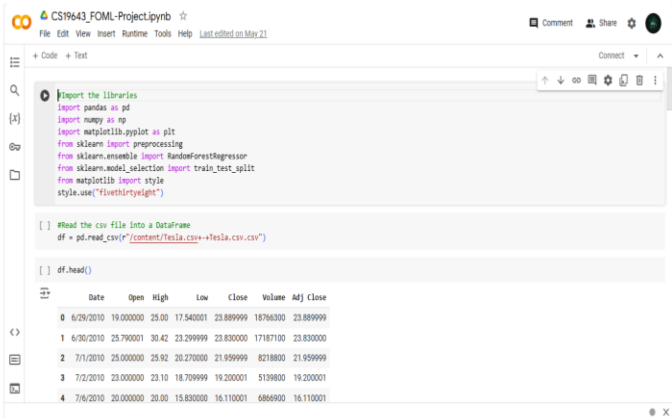
Extensive insights were obtained from the predicting Tesla's stock price with different models, such as Linear Regression, SVM, Random Forest, LSTM, and ARIMA. At first, the Linear Regression baseline model had a moderate level of predictive the capacity to precisely recognize linear relationships between input factors and Tesla's stock price. More intricate models, however, outperformed it in capturing the non-linear dynamics and patterns found in financial markets.

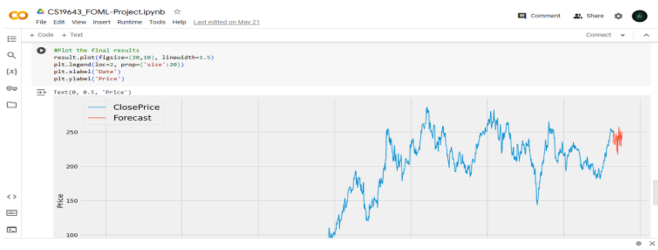
When it was time for simulating non-linear interactions, Support Vector Machines (SVM) performed better than other models because they could convert the data into higher-dimensional places and find the greatest hyperplanes to separate price changes.This made it possible for SVM to identify complex patterns in the data on Tesla's stock price, resulting in more accurate estimates than those produced with the Linear Regression baseline. In a similar vein, Random Forest outperformed in capturing the intricacy of variations in stock prices, improving forecast precision, and reducing overfitting by utilizing the combined intelligence of numerous decision trees.

The implementation of Long Short-Term Memory (LSTM) networks revealed promising outcomes, especially in identifying sequential patterns and temporal connections in Tesla's stock price data. LSTM's ability to retain memory over extended periods proved beneficial In forecast both long-term patterns and short-term changes correctly. Meanwhile, stuff like Autoregressive Integrated Moving Averages (ARIMA) provided useful information on the fundamental characteristics of Tesla's share cost by efficiently examining auto-regressive, differencing, and moving average elements in the time-series-data.

Overall, the ensemble of predictive models, including SVM, Random Forest, LSTM, and ARIMA, did better than the starting Linear Regression model in predicting Tesla's share price. By combining the strong points of each model and reducing its specific weaknesses, Even so, the together strategy gave a full and reliable base for guessing shifts in share costs, letting knowledgeable financial market decision-making. However, more looking into and making the models better might be needed toimprove how accurately they predict and how they can change with the markets as they change.

6. Output





7. Conclusion

To predict the Tesla share's closing price, this paper employed five different algorithms, including statistical and deep learning methods. It can be seen that the Random Forest prediction algorithm has the greatest forecast accuracy for the Tesla stock among machine learning algorithms. Because it's possible that changes in the share market don't often adhere to a recognizable pattern or a constant cycle. The existence and longevity of trends will differ based on the organizations and industries. Investment results will increase with an awareness of these cycles and trends. We should use an integrated algorithm technique like Random Forest, which has excellent anti-interference and anti-over-fitting characteristics, to assess highly volatile equities like Tesla. Deep learning models that include news stories about the economy and monetary factors like income statements, trade volume, etc. can be constructed for future work to produce potentially better outcomes.

8. Future Work

The stock values of the firm Tesla Inc. are listed in the research's data. On June 29, 2010, Tesla began trading publicly. The NASDAQ index lists Tesla's shares. Elon Musk, the richest man in the world at the time of its founding, has served as its CEO and has overseen operations in several nations. This is demonstrated by the fact that there isn't a big discrepancy between the actual and anticipated figures. Because the Rsquare values of Linear Regression, SVM, and Random Forest are all quite high, and the RMSE values are all very low, they were ready to be tested throughout the training period. These figures demonstrate how the system is capable of achieving the set objectives. Investors may trade their personal equities on the stock market with other traders using stock brokerages and computerized trading platforms. Traders try to acquire stocks with rising prices and sell stocks with dropping prices on the share market. Because of its potential advantages, it could make it possible to foresee the future with credibility, which is something that most economies and people have long wished for. Understanding how to predict price fluctuations may be helpful for forecasting stock market enthusiasts. The use of artificial intelligence enables scientists to foresee with more accuracy than ever before. Its accuracy will also increase with time as both technological advancements and computational accuracy do.

9. References

[1] Dhankar R S. Stock Market Operations and Long-Run Reversal Effect in Capital Markets and Investment Decision Making India. Springer, 2019.

[2] Usmani M, et al. Ali Stock market prediction using machine learning techniques. 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 2016, 322 - 327.

[3] Grigoryan H. A stock market prediction method based on support vector machines (svm) and independent component analysis (ica). Database Systems Journal, 2016, 7 (1): 12 - 21.

[4] Nti I K, Adekoya A F, Weyori B A. A systematic review of fundamental and technical analysis of stock market predictions. Artif. Intell. Rev, 2019, 53: 3007 – 3057.

[5] Sengupta A, Sena V. Impact of open innovation on industries and firms—A dynamic complex systems view. Technol. Forecast. Soc, 2020, 159: 120199.

[6] Terwiesch C, Xu Y. Innovation Contests, Open Innovation, and Multiagent Problem Solving. Manag. Sci, 2008, 54: 1529 – 1543.

[7] Blohm I, et al. Idea evaluation mechanisms for collective intelligence in open innovation communities: Do traders outperform raters? In Proceedings of the 32nd International Conference on Information Systems, Cavtat, Croatia, 2010, 21 – 24.

[8] Del Giudice, et al. The human dimension of open innovation. Manag. Decis, 2018, 56: 1159 – 1166.

[9] Daradkeh M. The Influence of Sentiment Orientation in Open Innovation Communities: Empirical Evidence from a Business Analytics Community. J. Inf. Knowl. Manag, 2021, 20: 2150031.

[10] Seber G A F, Lee A J. Linear regression analysis. John Wiley & Sons, 2012, 329.

[11] Sunil Ray. Svm | support vector machine algorithm in machine learning [internet]. Available from:https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-examplecode/. Google Scholar. 2017.

[12] Henrique B M, Sobreiro V A, Kimura H. Stock price prediction using support vector regression on daily and up to the minute prices. J. Financ. Data Sci., 2018, 4 (3): 183 - 201.

[13] Liaw Andy, Matthew Wiener. Classification and regression by Random Forest. R news, 2002, 2 (3): 18 - 22.

[14] Kumar, Manish, Thenmozhi. Forecasting stock index movement: A comparison of support vector machines and random forest. In Indian institute of capital markets 9th capital markets conference paper, 2006.

[15] Roondiwala M, Patel H, Varma S. Predicting stock prices using LSTM. International Journal of Science and Research (IJSR), 2017, 6 (4), 1754 - 1756.

[16] Zhang Peter. Time series forecasting using a hybrid ARIMA and neural network mode. Neurocomputing, 2003, 50: 159 - 175.

[17] Siami-Namini S, Tavakoli N, Namin A S. A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE international conference on machine learning and applications (ICMLA), 2018, 1394 - 1401