PREDICTION OF TESLA STOCK PRICE BASED ON LINEAR REGRESSION, SVM, RANDOM FOREST, LSTM AND ARIMA

A PROJECT REPORT
submitted by

THIRUESWARAN V (210701290)
SATHISH KUMAR R (210701515)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE,

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this project report **titled "Prediction of Tesla Stock Price prediction on Linear Regression, SVM, Random Forest, LSTM, and ARIMA**
" is the bonafide work of **"Thirueswaran v (210701290), Sathish Kumar v (210701515)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Karthick. V Associate Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai – 602015
Submitted to Project Viva-Voce Examination held on ____________________

Internal Examiner **External Examiner**

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 macroeconomics. 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 , Linear Regression, Random Forests, and LSTM. Following a thorough 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 highvolatility stocks like Tesla.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work.

We convey our sincere and deepest gratitude to our internal guide, **KARTHICK V** Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

THIRUESWARAN V – 210701290

SATHISH KUMAR R– 210701515

TABLE OF CONTENTS

CHAPTER 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.

1.1 GENERAL

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.

1.2 Objectives

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.

CHAPTER 2

LITERATURE REVIEW

The paper which we refered[1] 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].

2.1 Existing System

The existing system for predicting Tesla's stock price typically involves a combination of quantitative and qualitative methods. Quantitative approaches include statistical models and machine learning techniques that analyze historical price data, trading volumes, financial indicators, and macroeconomic variables. Common models used are time series analysis (like ARIMA), regression models, and more advanced neural networks such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs). These models can identify patterns and trends that might not be immediately obvious to human analysts.Integrating these approaches, sophisticated trading algorithms and prediction systems continuously update their forecasts based on new data. These systems may also factor in real-time events such as earnings reports, product launches, regulatory changes, and broader market movements. The goal is to create a dynamic and adaptive model that can provide accurate and timely predictions to inform investment strategies.

2.1 Proposed System

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.

CHAPTER 3

MODULE DESCRIPTION

Data Preprocessing Module:

Data preprocessing for Tesla stock price prediction involves several key steps to ensure the dataset is clean, consistent, and suitable for modeling. First, collect historical stock prices, including open, high, low, close, and volume data, from a reliable source. Handle missing values by using techniques like forward or backward filling. Next, convert date columns to datetime objects and set them as the index for time series analysis. Normalize or standardize the data to scale features appropriately.

Model Training Module:

To train a model for predicting Tesla stock prices, start by collecting historical stock price data along with relevant financial indicators and macroeconomic factors. Preprocess the data by normalizing the values and handling missing data. Split the dataset into training and testing sets. Use a time series forecasting model such as Long Short-Term Memory (LSTM) neural networks, which are well-suited for sequential data.

Model Evaluation Module:

Model evaluation for predicting Tesla's stock price involves several key metrics to ensure its accuracy and reliability. The primary evaluation criteria include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) values. MAE provides the average magnitude of errors in predictions, while MSE and RMSE penalize larger errors more heavily, with RMSE offering a more interpretable scale.

Prediction Module:

A prediction model for Tesla's stock price could leverage a combination of historical stock prices, financial indicators, and macroeconomic factors. Using a machine learning approach like a Long Short-Term Memory (LSTM) neural network, which is adept at handling time-series data, the model would be trained on past stock prices, trading volumes, and other financial metrics like moving averages and RSI (Relative Strength Index). Additionally, external data such as market sentiment derived from news articles and social media, as well as broader economic indicators like interest rates and GDP growth, could be incorporated. By capturing both the historical trends and external influences, the LSTM model would aim to provide accurate short-term predictions of Tesla's stock price, continually refining its forecasts as new data becomes available.

Software Requirements

- NumPy - Used for large computations
- Pandas - Load data into 2D format
- Seaborn - Used for data visualization
- Matplotlib - Used for data visualizations
- SkLearn -> contains libraries to perform tasks

Tools used:

- Jupyter
- Pycharm
- Streamlit

Hardware Requirements

- Processor: Intel Core i5 minimum, i7 recommended
- Memory (RAM): 8 GB minimum, 16 GB recommended
- Storage: 256 GB SSD minimum, 512 GB SSD recommended
- Graphics Card: Integrated Graphics minimum, NVIDIA GTX 1060 recommended

CHAPTER 4

PROJECT DESCRIPTION

4.1 SYSTEM ARCHITECTURE

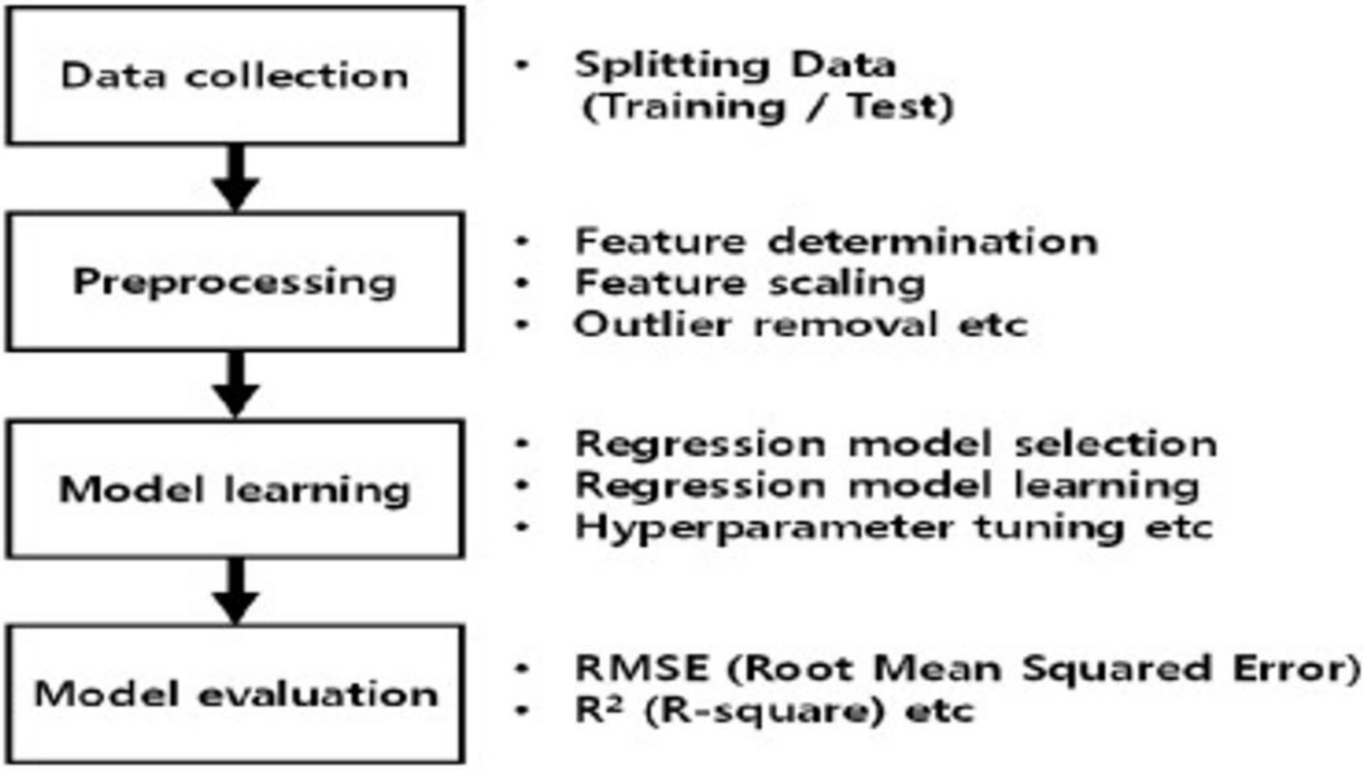

Fig 4.1 System Architecture

4.2 METHODOLOGY

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.

CHAPTER 5

RESULTS AND DISCUSSION

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. Since becoming the company's CEO in 2008, Musk has consistently surpassed expectations and overcome obstacles. The LSTM, SVM, Random Forest, and Linear Regression models' training outcomes are shown in Tables 1, 4, 7, and 10. The results show how reliable and strong the deep learning model is at predicting the current value based on the training data. Low forecast error rates are observed in Random Forest, as measured by the RMSE (0.56248).

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 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 shoulduse 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.

6.2 Future Work

Future 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.

REFERENCES

[1] Dhankar R S. Stock Market Operations and LongRun 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/unde rstaing-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),

[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,