Tesla stock prediction: a comparative study between four models

Hao Li*

Department of Mathematical Sciences, University of Southampton, Southampton, the UK

* Corresponding Author Email: hl23g22@soton.ac.uk

Abstract. One of the most significant components of the economy is the stock market. Due to the impact of many industries and market conditions, Tesla stock prices are continually fluctuating. Stock market forecasts are becoming more precise as artificial intelligence develops. The performance of the four methods is compared in this article, which predicts Tesla's stock using linear regression, super vector regression, RFR, and LSTM. The study's findings show that all four methodologies are capable of accurately predicting Tesla's stock prices according to four parameters (R-squared, MSE, RMSE, and MAE). Linear regression stands out among them due to its highest R-squared value (0.85) for Tesla from 2020 to 2022, compared to the other three models. The findings of this study provide empirical evidence for investors.

Keywords: Tesla stock price prediction, linear regression, super vector regression, random forest regressor, long-short term memory

1. Introduction

Shares of publicly traded firms are a company's measure of value. The state of a nation's economy is seen to be reflected in its stock market. On the stock market, these shares are transferred and traded. The price of stocks may also rise or fall in the short term. For both buyers and sellers, accurate stock predictions can result in enormous profits. It's a common misconception that stock price movements may be predicted through analysis rather than being random. The stock market, market trends, and price predictions may all be successfully made using machine learning, typically with high levels of accuracy. In studies, the application of machine learning to stock forecasting has gotten a lot of attention [1].

Machine learning techniques are widely used in stock prediction. Methods like linear regression(LR), Super vector regression(SVR), Random Forest regressor(RFR), and Long-short term memory(LSTM) are used in this paper. This paper aims to make a comparison in terms of performances between the four chosen models by measuring parameters.

2. Literature review

Stock forecasting is extremely risky until a trustworthy and accurate computation model is developed. In recent years, a considerable deal of study has been done on predicting the value of a company's shares. By utilizing the stock data of various corporations or entities from the global financial market, it is possible to anticipate changes in stock indices in the future, either through the application of machine learning or not.

This technique is simple but extremely useful in fields such as finance and economics. In the case of an investment, linear regression can be used to quantitatively relate the investment and the risk and to forecast the risk in the future [2].

A new approach is presented to stock price forecasting based on the theory of super vector regression to enhance the study of the stock market price trend. First, a floating search approach for factor selection based on support vector regression and the leaving-one method is proposed. This method uses non-linear expansion of data using SVR to identify the time series model order. The test outcome is satisfactory. There is a research state that the SVM algorithm works on large-scale data values derived from the global financial markets. SVM is not overfitting [3].

A machine learning algorithm known as a random forest (RF) classifier categorizes data by using a group of decision trees. A forest of trees is created by combining the decision trees. And then trained using a different subset of data and features. Each decision tree is used to predict by the algorithm during projection.

Deep learning LSTM method exhibits great accuracy in predicting stock values when accuracy, precision, and standard deviation are measured and compared across all machine learning techniques.

The remaining of the paper is structured as follows. Session II introduces dataset and the four methods. Session III summaries the experimental results, including the performance of each model under the measuring parameters. Session IV concludes.

3. Method

3.1. Dataset

It is needed to choose a specific dataset when utilizing machine learning techniques to forecast stock prices because slight changes in the data could affect a difference in the outcomes of the predictions [4]. The Tesla stock market data have been used in this research that is collected from Kaggle. The dataset has 562 rows and 7 columns in total. Each row represented the prices for a stock trading day. Table.1 indicates variables including the data set and definition for each variable.

Table 1. variables and definitions

Variables	definitions
Open price	The opening price of the day
Highest price	The highest price for each sample
Lowest price	The lowest price of the sample
Close price	The close price of the day
Adj. close price	The closing price after adjustments
Volume	The volume of shares traded of each sample day

3.2. Normalization of data

3.2.1 Data splitting

Before splitting, it is required to look for any missing values and remove such data from the dataset. It is also necessary to look for categorical values and remove any extraneous data from the data collection. The training set and the testing set, which account for 80% and 20% respectively of the total data, will then be separated. There are 449 rows of data considered as training sets and the left 113 rows of data is the testing set. And the training data(449 rows) is from 2020-01-02 to 2021-10-12; the testing data(113 rows): from 2021-10-13 to 2022-03-24.

3.2.2 Data scaling

Normalization efficiently scales the data to fit within a specific range when working with a lot of data. Gradient descent is accelerated and becomes more precise after normalization [5].

3.3. Prediction models

The LR, SVR, RFR, and LSTM are chosen out of all the machine learning techniques. Four models are first trained using training data, followed by prediction using testing data. Next, evaluate the model's performance using the parameters to determine whether it is a good fit.

3.3.1 Linear regression

A valuable tool for forecasting quantitative responses is linear regression. Due to its simplified model, linear regression also produced findings that were plausible after normalization without the need for parameter tuning, however, the accuracy was not as good as it could have been if using the data to form a portfolio. Although it might seem a bit dull compared to the other three models

mentioned in this study, linear regression is still an effective and popular statistical learning technique. Linear regression uses a given linear function for predicting:

$$y = \beta_0 + \beta_1 X + \varepsilon. \quad (1)$$

In this formula, Y is a continuous value, X represents independent values and β_0, β_1 are coefficients.

3.3.2 SVR

SVR is a subset of the Support Vector Machine (SVM), which is a common technique. This model is popular due to the wide range of applications it can be used for, from pattern recognition to regression estimation with a generalized performance [6]. Super vector regression can handle both classification and regression on linear and non-linear data. And this model is convenient when finding more complex relationships. The following equation shows how SVR works:

$$y = wx + b \quad (2)$$

3.3.3 RFR

Random Forest regression can handle datasets containing continuous variables. The following describes the algorithm:

Step 1: Select N randomly chosen records from the dataset.

Step 2: Create a decision tree based on N records.

step 3a: Repeat steps 1 and 2 after selecting the number of trees from your algorithm.

Step 3b: In the case of a regression issue, each tree in the forest forecasts a value for Y (output) for a new record [7].

3.3.4 LSTM

The forecasts' heavy reliance on data—often on the market's extensive historical data—is the main reason for using this model in stock market prediction [8]. In the LSTM model, input was taken, and the LSTM algorithm was used to recall data and results over time [9]. To forecast the further stock prices of Tesla. The framework of the LSTM approach is illustrated in fig. 1.

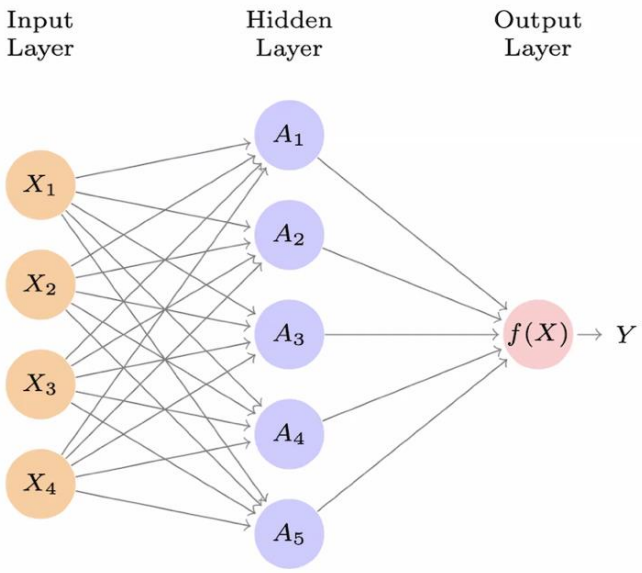

Fig. 1 LSTM single-layer neural network

3.4. Evaluation parameters

RMSE was employed as the main metric for the optimization to conduct the evaluation [10]. Four parameters are utilized in this research for evaluating the performance of four models. They are R-squared value, MSE, RMSE, and MAE. The following part will introduce the specific formulae.

3.4.1 R-squared value

The formula for the R-squared value is shown in (3). The higher R^2 value represents a smaller difference btw the observed data and the fitted values.

$$R^2 = \frac{\sum_{t=1}^{N}(\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2} \tag{3}$$

3.4.2 MSE

The formula for MSE is shown in (4). The lower MSE value means the closer is forecast to the actual.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4}$$

3.4.3 RMSE

The formula for RMSE is shown in (5). The lower RMSE value indicates better accuracy.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{Y}_i - Y_i)^2}{n}} \tag{5}$$

3.4.4 MAE

The formula for MAE is shown in (6). The MAE measures the average error. The smaller, the better.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{6}$$

4. Experimental results

Historical data on Tesla stock is taken from Kaggle. Machine learning methods, linear regression, Super vector regression (SVR), and random forest regressor models with one of the deep learning methods LSTM are used to train the data and predict stock prices.

Fig. 2, Fig. 3, Fig. 4, Fig. 5 shows predicted and original stock prices of Tesla Inc. which has been predicted using SVR, Random Forest regressor and long-short term memory (LSTM) respectively. Considering the performance indicator of four models as summarized in table 2, linear regression is the most effective strategy for the Tesla stock market.

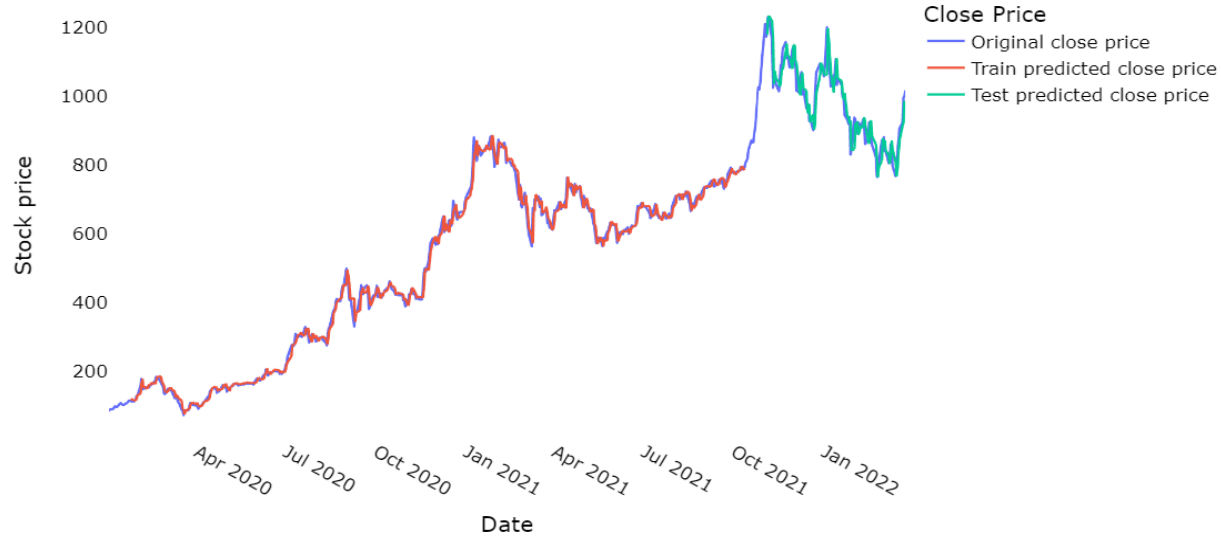

Fig. 2 Original stock value and the predicted values for Tesla based on the linear regression model

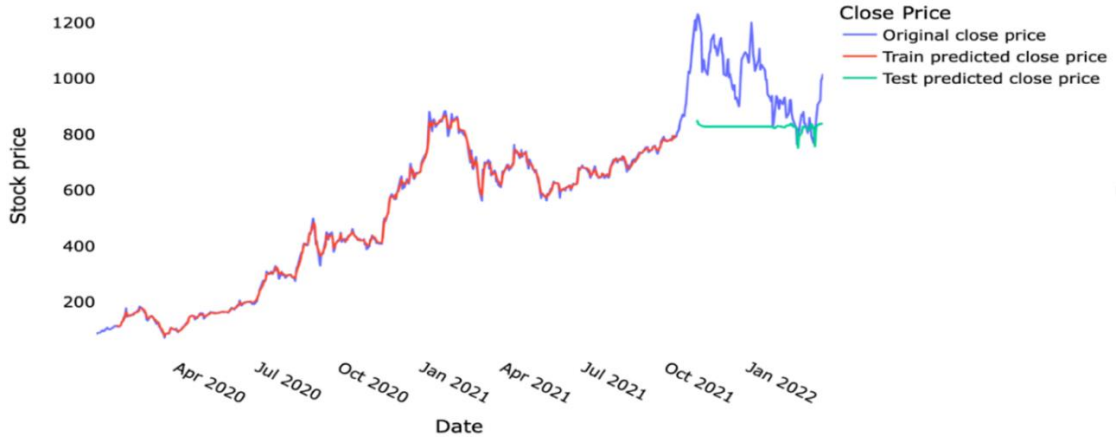

Fig. 3 Original stock value and the predicted values for Tesla based on the SVR model

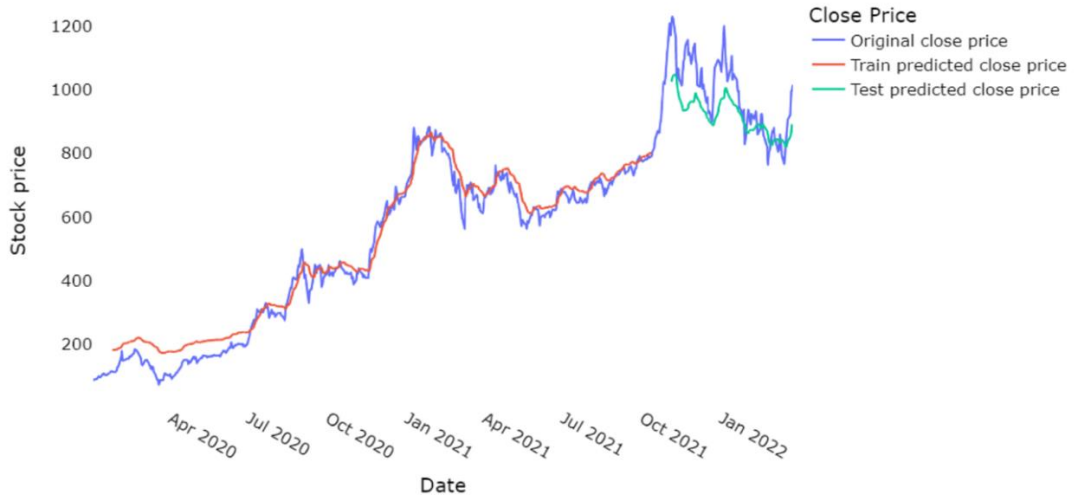

Fig. 4 Original stock value and the predicted values for Tesla based on random forest model

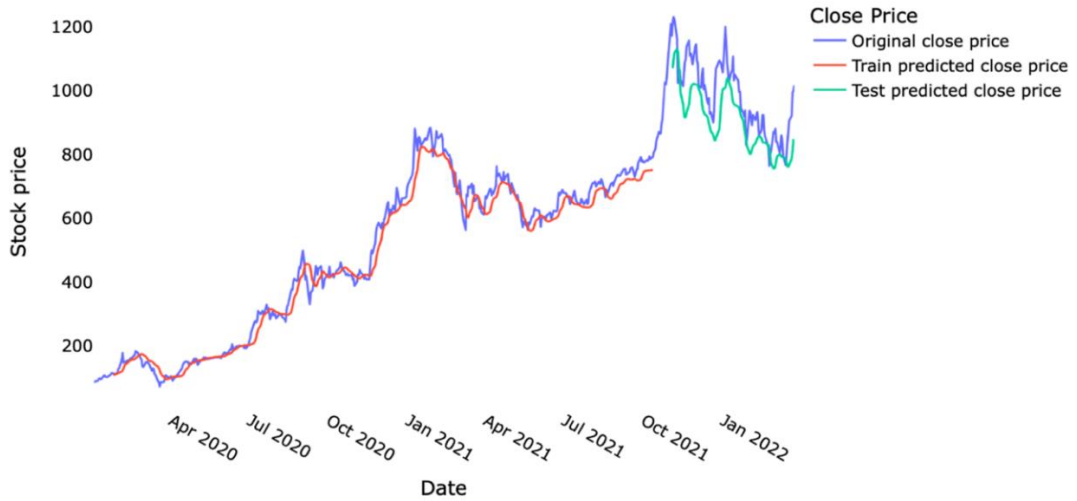

Fig. 5 predicted and original stock prices of Tesla Inc. which has been predicted using LSTM

Table 2. performance for four models

Methods	R^2 value	RMSE	MSE	MAE
Linear Regresion	0.85	44.6	1989.1	34.7
Super Vector Regression	0.32	94.0	8843.0	76.1
RandomForest Regressor	-1.76	189.5	32924.5	158.1
LSTM	0.26	97.9	9592.0	81.8

5. Conclusion

This paper predicts Tesla's stocks by using four machine-learning techniques, linear regression, SVR, random forest regression, and LSTM. The key findings include two folds. Firstly, each machine learning algorithm performs differently on various data points. Secondly, linear regression is the most effective strategy for the Tesla stock market. Tesla's stock price prediction for investors can use four indicators to forecast future stock prices: R squared value, MSE, RMSE, and MAE. The stock closing prices that were close to the initial values on the same days are predicted by the linear regression model.

However, this study only used four models, and many other methods may be more effective. More parameters can improve the accuracy of the evaluation. In the future, taking people's perspectives into account and running them through models when training the system can improve the model's accuracy for traders, investors, or stock market enthusiasts.

References

[1] Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. .. Stock market prediction using machine learning techniques. International Conference on Computer & Information Sciences. IEEE(2016).

[2] Xu, W. , Wang, B. , Liu, J. , Chen, Y. , Duan, P. , & Hong, Z. .Toward practical privacy-preserving linear regression. Information Sciences, (2022)596, 119-136.Fangfang.

[3] M, KHANDELWAL, and, PK, KANKAR, & and, et al.. Evaluation and prediction of blast-induced ground vibration using a support vector machine. Mining Science & Technology(2010).

[4] Girsang, A. S. , Lioexander, F. , & Tanjung, D. .Stock price prediction using lstm and search economics optimization. IAENG Internaitonal journal of computer science(2020)(4 Pt.2), 47.

[5] Aldhyani, T. H., & Alzahrani, A. Framework for predicting and modeling stock market prices based on deep learning algorithms. Electronics, (2022), 11(19), 3149.

[6] Nunno, L. . Stock Market Price Prediction Using Linear and Polynomial Regression Models, 2023.

[7] Casella, G. , Fienberg, S. , & Olkin, I. .. An introduction to Statistical learning with R(2006).

[8] Hore, S. , Vipani, R. , Das, P. , & Dutta, S. . Prediction of Stock Value Using NIFTY 50 Market Index and RBF-Kernel Based Support Vector Regressor (2018).

[9] Polamuri, S. R., Srinivas, K., & Mohan, A. K. Stock market prices prediction using random forest and extra tree regression. Int. J. Recent Technol. Eng, (2019) 8(1), 1224-1228.

[10] Raza, K. Prediction of Stock Market performance by using machine learning techniques. 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT). IEEE (2017).