Analysis of Stock Price Prediction in Context of Machine Learning Models for Tesla

Jiayuan Xu [1, *, †] and Yi Yang[2, †]

[1] Eberly College of Science, Pennsylvania State University, Pennsylvania, USA

[2]School of Science, Xi'an Jiaotong-Liverpoor University, Suzhou, China

*Corresponding author: jfx5074@psu.edu

†These authors contributed equally

Abstract. Contemporarily, Investors spend plenty of time to speculate and predict the growing trend of the stock price in order to gain extra return from the stock market. Nowadays, the problem of natural resources and global warming has put oil-fueled automotive into controversial dispute. Therefore, the importance of environment-friendly automotive is remarkable in global scale. Tesla (TSLA), as one of the leading electric automotive builders, widely attracted the attention of investors around the world. In this article, we will adopt several state-of-art models in machine learning to predict the stock price of Tesla including ARIMA, LSTM, Linear Regression to analyze the stock price of TSLA. 80% of data is used to be the training set and 20% as the contrast group to verify the accuracy of the prediction. According to the analysis, the outcome of ARIMA model is quite accurate, and LSTM model is better than linear regression model. These results shed light on guiding further exploration of electric vehicle, the new blood of automobile industry.

Keywords: Tesla; machine learning; ARIMA; LSTM; Linear Regression.

1. Introduction

Stock has both high returns and high risks. The stock market is affected by many factors, and the stock price is elusive. The research on stock price prediction is of great value. However, in modern world, artificial intelligence and machine learning is taking over many of the tedious work. Consequently, there are many scholars trying to predict the stock price with the help of models in machine learning, which will be helpful and crucial for investors to gain extra returns (i.e., the alpha) from the market [1]. However, neither human brain nor machine could give a 100% accurate prediction in the stock market. Nikou et al. also concludes that machine forecasts are not accurate, but the algorithm utilized determines how inaccurate they are. An appropriate way for researchers to do is decreasing the error using the best-fitting model.

In modern field of machine learning, there are plenty of choices of models to choose e.g., ARIMA, GARCH, LSTM [2]. Another thing besides the use of machine is to perceive the trend in global extent. For instance, due to the increasing scarcity of the natural resources, as well as the increasing severity of global warming, public concern regarding the impact of oil-fueled automotive on global environment have reached an unprecedented level. In burning fuel, automobiles contribute to the depletion of the world's limited reserves of oil [3]. In addition, the price of oil has been increasing due to political reasons. The fluctuations of the gas are positive relevant to the enhancement of the demand for storing [4], not mentioning the impact of trade war and gas dispute between competing countries. Hence, the demand for environment-friendly automotives is consistently increasing worldwide. Human beings have not invented an 100% environmentally friendly automotive, but the most common emission-free automobile that mankind have is electric cars.

Tesla, as one of the earliest electric automotive companies, has a vast group of loyal customers. In addition, the interrelation with cryptocurrency (e.g., Bitcoin and Dogecoin) has put Tesla into the sight of a wider range of stockholders [5]. Plus, Elon Musk, the CEO of Tesla, established the SpaceX program. This program connects Tesla with the development of the latest technology that human has, and transfer Elon Musk from an entrepreneur to a leader in the entire human development. For example, the reusable rocket will probably overthrow our current transportation methods. All these bold thoughts that Elon Musk made is letting Tesla becoming a popular choice of investment. As a result, we chose Tesla as our study target and focused on developing the best-fit model for predicting Tesla's stock price.

In the rest part of the paper, we will be using several models in machine learning to forecast the price, including ARIMA, LSTM, and Linear Regression. Methodology of doing the research will be provided, and actual results and contrast between different models will be further discussed. Limitations on our research will also be shown. Elements related to analyzing the research will also be taken into consideration, e.g., the data of cryptocurrency and more competing electric car brands.

2. Methodology

2.1 Data

In order to construct the models, we collected the stock price of TSLA of the last five years from Yahoo Finance and Stooq website. Tesla's stock is denoted by the symbol TSLA, which also includes the date, open price, closing price, high price, low price, and volume. In this paper, the expected stock price is the dependent variable.

2.2 ARIMA Model

To create a model, we initially utilized ARIMA, which is a basic, mature and stable statistical models for prediction. ARIMA is a linear modeling technique [6], which means that it is very simple, just using historical data to forecast future time series values [7]. To be specific, the dependent variable is solely regressed on its lagged value, the present value, and the lagged value of the random error component once the non-stationary time series are converted into stationary time series [8]. There are three terms, which are frequently represented by the letters p, d, and q. These letters stand for the order of the AR term, the number of differentiating transformations that the time series underwent in order to become stationary, and the order of the MA term, respectively. The formula is defined as follows:

$$y_i = \mu + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} \tag{1}$$

In the equation above, α_i stands for the weight that historical data y_{t-i} to the predicting value. β_i represents the weight of ε_{t-i} to the predicting value at the moment of $t-i$.

2.3 LSTM Model

As a matter of fact, LSTM, as a improve model of Recurrent neural networks (RNN), provides a unique path to address the issue of gradient missing for RNN [9]. Recurrent neural networks can effectively handle time series, support multi-step time series input, and have linked hidden layers [10]. Compared with normal neural networks, it can process sequential changes in data. Since it uses the hidden nodes from t-1 time slices as the input for current time slices at t time slice, as seen in Fig. 1, RNN performs exceptionally well on time series data. Here,

$$h_t = \sigma(x_t w_{xt} + h_{t-1} w_{ht} + b) \tag{2}$$

In RNN, the neural network unit additionally accepts h_{t-1} as an input in addition to the conventional DNN hidden nod expression. The initial purpose of the hidden node h_{t-1} is to determine the expected value at the instant $\hat{y}_t$: $\hat{y}_t = \sigma(h_t w + b)$. The hidden node state of the following time slice, h_t, is determined by the second function.

The neural network nodes' long-term reliance is caused by the fact that they have covered the characteristics of earlier rather lengthy time slices after many iterations of calculation. Each time slice in an RNN utilizes the same parameters because it is a chain structure. To address the long-term reliance of RNN, LSTM incorporates a gate mechanism to manage the flow and loss of features. A number of LSTM units make up LSTM. Its chain structure is shown in Fig. 2.

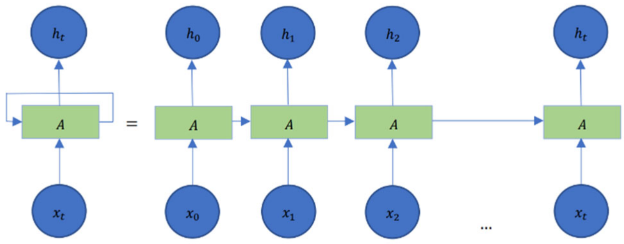

Fig. 1. A sketch of RNN

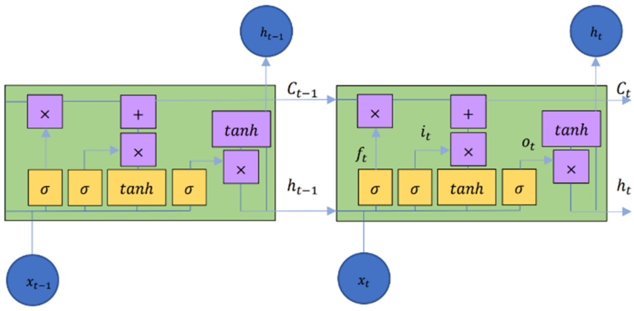

Fig. 2. A sketch of LSTM.

A neural network layer is represented by each orange box and consists of weights, bias, and activation functions. An element level action is represented by each purple box. The vector direction is shown by the direction of arrow, intersecting arrows show vector concatenation, and forked arrows show vector duplication. The core part of the LSTM is the top conveyor belt, which is commonly called the cell state. It exists throughout the chain system of the LSTM.

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{3}$$

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4}$$

where f_t is the forgetting gate. f_t is a vector with each element in $[0,1]$. It uses the features of C_{t-1} to calculate C_t. For the gate mechanism of LSTM, Hadamard product is the most significant one. It represents the unit multiplication relationship between f_t and C_{t-1}.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{6}$$

Here, the function of i_t is to control the features of $\tilde{C}_t$ and to update C_t. It is used in the same way as f_t. Lastly, to compute the predicted value $\hat{y}_t$ and create the whole input for the next time slice, compute the output h_t of the hidden node.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \tanh(C_t) \tag{8}$$

Here, h_t is obtained from the output gate o_t and the unit state C_t, where o_t is calculated in the same way as f_t and i_t. For the importance of the state in LSTM, the forgetting gate is more importance than the input gate, and the output gate is the last.

2.4 Linear Model

Supervised learning in machine learning algorithm falls into two categories, which are classification and regression. Classification is to judge the category of the target variable. Regression analysis is the analysis of quantitative dependence of objective things, which is a mathematical method of dealing with relationships among multiple variables. In regression analysis, there are unary linear regression and multiple linear regression. For the first part, a line can approximate the relationship between an independent variable and a dependent variable. For the second part, the regression analysis contains two or more independent variables and there is a linear relationship between them. Given data set $D = (x^{(1)}; x^{(2)}; \dots; x^{(m)})$, where $x^{(i)}$ means that the i^{th} sample point $x^{(i)} \in \mathbb{R}$. The linear regression model can be expressed as follows:

$$h_\theta\left(x^{(i)}\right) = \theta_0 + \theta_1 x^{(i)} \tag{9}$$

where $h_\theta(x^{(i)})$ represents predictive value, θ_j represents model parameters.

2.3.2 Loss Function and Least Square Method

The loss function calculates the degree of dissimilarity between the predicted and the real. The larger the loss function is, the worse the model is. It is also hard for model to fit the data.

$$L(\hat{h}_\theta, y) = \frac{1}{m}\sum_{i=1}^{m}(\hat{h}_\theta^{(i)} - y^{(i)})^2 = \frac{1}{m}\sum_{i=1}^{m}[(\theta_0 + \theta_1 x^{(i)}) - y^{(i)}]^2 \tag{10}$$

$$\theta_1^*, \theta_0^* = \underset{\theta_0,\theta_1}{\operatorname{argmin}} L(\theta_0, \theta_1) = \underset{\theta_0,\theta_1}{\operatorname{argmin}} \frac{1}{m}\sum_{i=1}^{m}[(\theta_0 + \theta_1 x^{(i)}) - y^{(i)}]^2 \tag{11}$$

With the loss function, the process of machine learning is dissolved into the process of optimizing the pairwise loss function. For least squares, it is an optimization problem.

$$MSE(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{12}$$

The value of θ_j is determined by minimizing MSE, which minimizes the Euclidean distance from all samples to the line.

$$(\theta_1^*, \theta_0^*) = \underset{\theta_0,\theta_1}{\operatorname{argmin}} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{13}$$

First step, taking the derivative of θ_0 and θ_1:

$$\begin{cases} \frac{\partial MSE(\theta_0,\theta_1)}{\partial \theta_1} = \frac{2}{m}\left(\theta_1 \sum_{i=1}^{m} x^{(i)^2} - \sum_{i=1}^{m}\left(y^{(i)} - \theta_0\right)x^{(i)}\right) \\ \frac{\partial MSE(\theta_0,\theta_1)}{\partial \theta_0} = \frac{2}{m}\left(m\theta_0 - \sum_{i=1}^{m}\left(y^{(i)} - \theta_0\right)x^{(i)}\right) \end{cases} \tag{14}$$

Secondly, leting the equations of (14) equal to 0:

$$\begin{cases} \theta_1 = \dfrac{\sum_{i=1}^{m} y^{(i)}\left(x^{(i)} - \bar{x}\right)}{\sum_{i=1}^{m} x^{(i)^2} - \frac{1}{m}(\sum_{i=1}^{m} x^{(i)})^2} \\ \theta_0 = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\left(y^{(i)} - \theta_1 x^{(i)}\right) \end{cases} \tag{15}$$

Therefore, the form of θ_0 and θ_1 could obtain.

Typically, the data set is $D = \left(x^{(1)}, y^{(1)}\right); (x^{(2)}, y^{(2)}); \dots; (x^{(m)}, y^{(m)})$, where $x^{(i)}$ means that the i^{th} sample point $x^{(i)} \in \mathbb{R}^n$.

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_n x_n^{(i)} \tag{16}$$

where $h_\theta(x^{(i)})$ represents predictive value, θ_j represents model parameters, θ_0 is the bias of model. It can be generalized in the form of a matrix $h_\theta(x^{(i)}) = \theta^T x^{(i)}$, where

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}, x^{(i)} = \begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix} \tag{17}$$

and $x_0^{(i)} = 1$. θ^T is also a vector form and it indicates the corresponding weight of each feature. The loss function is as follows,

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{18}$$

The matrix form is

$$MSE(\theta) = \frac{1}{m}(X\theta - Y)^T(X\theta - Y) \tag{19}$$

$$J(\theta) = \frac{1}{2m}(X\theta - Y)^T(X\theta - Y) \tag{20}$$

Here,

$$X = \begin{bmatrix} x^{(1)^T} \\ x^{(2)^T} \\ \vdots \\ x^{(m)^T} \end{bmatrix} = \begin{bmatrix} x_0^{(1)} & x_1^{(1)} & \cdots & x_n^{(1)} \\ x_0^{(2)} & x_1^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ x_0^{(m)} & x_1^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \tag{21}$$

represents the training set,

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \tag{22}$$

is a vector that is made up of all the training sample outputs. The gradient is essentially a vector that says that the directional derivative of a function at that point is maximized along that direction. The function changes most rapidly at that point along that direction (the direction of the gradient) and has the highest rate of change (the magnitude of the gradient). The gradient descent method is to find the direction in which the function decreases fastest in order to find the optimal solution.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \tag{23}$$

$$\theta_j := \theta_j - \frac{\propto}{m} \sum_{i=1}^{m} \left(\left(y_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \right) \tag{24}$$

where α is the step size and $\frac{\partial}{\partial \theta_j} J(\theta)$ is the rate of decline.

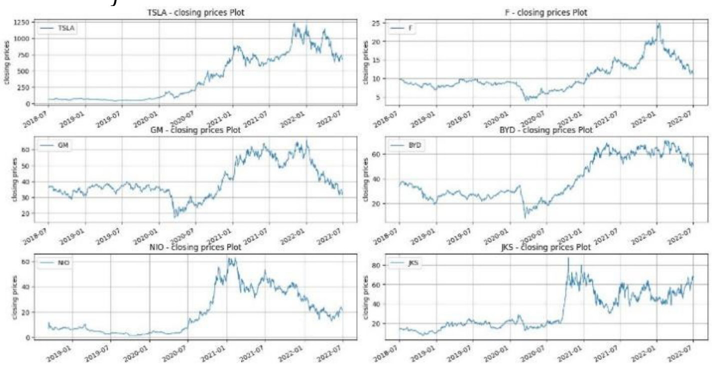

Fig. 3. Stock price of each company.

Fig. 4. Stock price of six companies.

3. Empirical analysis

3.1 Data processing

In order to explore the relevance of each company to Tesla, get data of these companies from the Internet by pandas_datareader. TSLA represents Tesla, F represents Ford, GM represents General Motors, BYD and NIO represent two Chinese electric vehicle companies, JKS represents Faraday future. The stocks of different car companies are shown in Fig. 3 and Fig. 4. For different electric car companies and fuel care companies, they could have a different impact on the stock price of Tesla.

The correlation matrix of these companies is given in Table. 1. For BYD, NIO and JKS, they are all electric car companies. They have a high correlation with Tesla. Although electric vehicles have begun to take over some of the market for fuel cars, the stock prices of electric and fuel cars are currently positively correlated. The reason for this might be that they are all in the auto industry. The ups and downs of their stock prices have a close bond to the whole industry. If we assume that the fuel is scarce and expensive many years from now, the correlation might change.

Table 1. Correlation matrix

	TSLA	F	GM	BYD	NIO	JKS
TSLA	1.000000	0.817316	0737912	0.890344	0.790815	0.841560
F	0.817316	1.000000	0.818275	0.869219	0.529219	0.583638
GM	0737912	0.818275	1.000000	0.858739	0.779537	0.620614
BYD	0.890344	0.869219	0.858739	1.000000	0.741779	0.731870
NIO	0.790815	0.529219	0.779537	0.741779	1.000000	0.833866
JKS	0.841560	0.583638	0.620614	0.731870	0.833866	1.000000

3.2 Prediction

For the ARIMA model, the cumulative return plot is shown in Fig. 5. In this model, we use 80 percent of the historical data to serve as our training data, the rest to be as the contrast group to see if the prediction result is accurate. In our diagram, we have the training data colored blue. In the prediction part, the prediction results are presented as green dots, and the actual stock price is colored red. Interesting, the result is really accurate, it predicts the stock very successfully.

Fig. 5. Prediction results of ARIMA.

Even thought that we have got an accurate result, there are limitations and disadvantages regarding this model. The ARIMA model is not as a good choice when it refers to long-term predictions. Furthermore, the prediction is not as sensitive to the outliers. For the LSTM model, the whole data set is divided into training set and test set and the proportion of training set is 70%. Set the time step as 100. Reshape the input data because LSTM model requires 3D input data. The LSTM model is constructed where sets epochs=25, batch_size=32. The results are presented in Fig. 6. In order to predict the stock price of Tesla in 30 days with linear regression model. Dividing the whole data into train set and test set, the results of linear regression is illustrated in Fig. 7.

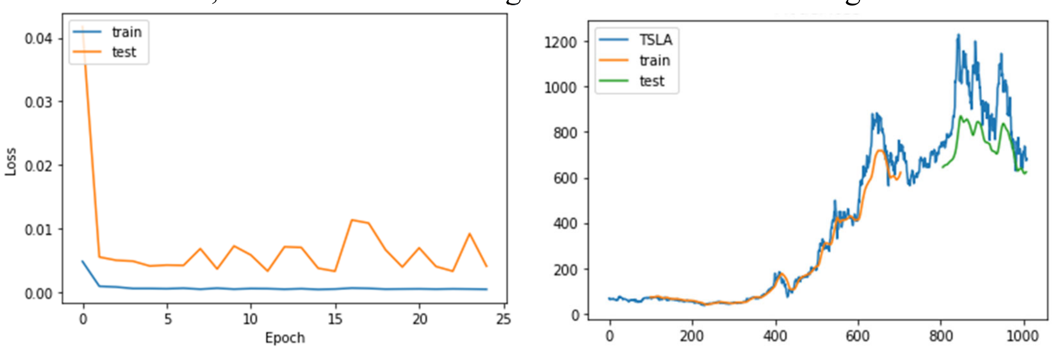

Fig. 6. The loss as a function of Epoch and prediction results for Tesla of LSTM.

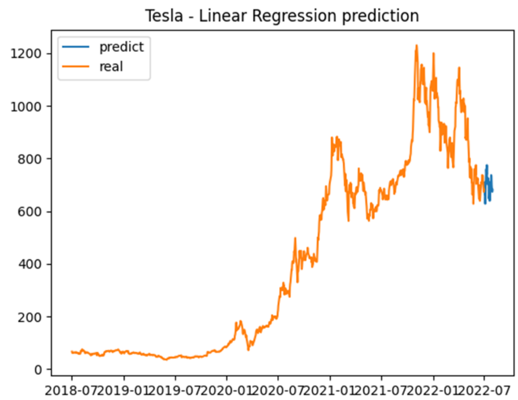

Fig. 7. Prediction results of linear regression.

3.3 . Comparison

Although the Tesla stock price was correctly predicted, there is still a discrepancy between the outcome and the actual value. Both the LSTM approach and the linear regression model have undesirable root mean square errors. It shows that there are still some problems in data processing, fitting, and prediction. However, it can be explained that the prediction result of LSTM model is better than that of linear regression model.

4. Summary

In conclusion, this paper investigates the better model on predicting TSLA stock price based on ARIMA and LSTM model. According to the analysis, the outcome of the entire research is mind-blowing. We successfully predicted the stock price of Tesla using multiple models and took other factors like other electric cars into consideration. The data used in this study is not based on the entire database, but five years in total due to unity purposes. The results could be better with more database trained. The stock price of Tesla has been predicted successfully, there is still a gap between the predicted result and the real value. There might be several reasons. Firstly, models are simple and the number of epochs is small. The training and forecasting process may not have a very high degree of accuracy. Secondly, Economic, political and other objective factors affect stock prices. The impact of uncontrollable factors such as the epidemic on stock prices is also large. Due to the limitation of time and knowledge, this project has achieved these results. In the future, the prediction results may be better if the model is further optimized or multiple models are combined. Overall, these outcomes provide a general framework for predicting stock price. The electric car is one field that is booming, the stock price and other aspects worth further exploration and research.

Reference

[1] Nikou M, Mansourfar G, Bagherzadeh J. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. Intelligent Systems in Accounting, Finance and Management, 2019, 26(4): 164-174.

[2] Ho M K, Darman H, Musa S. Stock Price Prediction Using ARIMA, Neural Network and LSTM Models. Journal of Physics: Conference Series. IOP Publishing, 2021, 1988(1): 012041.

[3] King Ciaran. Cars and global warming. The Irish Times, May 21, 1991, pp. 13. ProQuest, Retrieved from: https://ezaccess.libraries.psu.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fhistorical-newspapers%2Fcars-global-warming%2Fdocview%2F530057143%2Fse-2%3Faccountid%3D13158.

[4] Nick S, Thoenes S. What drives natural gas prices?—A structural VAR approach. Energy Economics, 2014, 45: 517-527.

[5] Business This Week. The Economist, The Economist Newspaper, Retrieve from: https://www.economist.com/the-world-this-week/2021/02/11/business-this-week.
[6] Babu C N, Reddy B E. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Applied Soft Computing, 2014, 23: 27-38.
[7] Singh S, Parmar K S, Kumar J, et al. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. Chaos, Solitons & Fractals, 2020, 135: 109866.
[8] Banerjee D. Forecasting of Indian stock market using time-series ARIMA model. 2014 2nd international conference on business and information management (ICBIM). IEEE, 2014: 131-135.
[9] Yan B, Aasma M. A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. Expert systems with applications, 2020, 159: 113609.
[10] Zhao J, Zeng D, Liang S, et al. Prediction model for stock price trend based on recurrent neural network. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(1): 745-753.