Machine Learning in Tesla's Stock Price Prediction

Yalong Kong*

Faculty of Science, The Chinese University of Hong Kong, Hong Kong, China

* Corresponding Author Email: 1155157170@link.cuhk.edu.hk

Abstract. With the development of new energy battery technology and people's attention to environmental issues, new energy vehicles have gradually become one of the important choices for people to travel, and stock investors have also begun to keep an eye on the field of new energy vehicles. However, since the new energy industry is an emerging industry, the current stock market analysis in this regard is not complete. Machine learning is one of the important means of stock forecasting at present. This article will use 6 machine learning models such as linear regression, polynomial regression, XGBoost, ARIMA, Prophet and LSTM to analyze and predict the stock of Tesla, a leading company in the field of new energy vehicles, so as to judge the investment prospects of new energy vehicles. The outcomes demonstrate that LSTM has the lowest error in predicting the price of Tesla's stock.

Keywords: Machine Learning, Tesla, LSTM.

1. Introduction

Changes in the stock market also reflect the prospects of the industry to a certain extent [1]. For many years, investors have always hoped to be able to accurately predict stock prices to profit from it. However, given that a number of factors have an impact on the stock market such as policies, market conditions and stockholders' confidence, the stock price has the characteristics of dynamic, non-linear and noise. This also makes predicting the price and movement of stocks quite difficult [2].

Deep learning technology has been widely used in market analysis in recent decades as a result of the rise of big data and the continuous mature of AI technology. Deep learning and big data analytics are presently used in SMP (Stock Market Prediction), which can lead to more optimal decision-making than previous frameworks. Vijh et al. used ANN and Random Forest to predict the stocks of five manufacturers including Nike [3]. Jorgenson et al. used LSTM to analyze stock change of India [4]. YUN et al. used hybrid GA-XGBoost algorithm for prediction [5]. With the oil crisis, environmental pollution and the breakthrough of new energy battery technology, new energy vehicles are regarded as an important choice to replace fuel vehicles. However, the current research on the new energy vehicle stock market is not as sufficient as other fields, and cannot provide data support for investors who want to invest in the new energy field.

This paper will use 6 machine learning methods such as linear regression, polynomial regression, XGBoost, ARIMA, Prophet and LSTM to analyze Tesla's stock price changes. According to the analysis of the four criteria such as MSE, RMSE, MAE and stock price trend, LSTM has the smallest error and the most accurate prediction value. The result shows that the LSTM model can be used more in future new energy stock price forecasts.

2. Introduction to Tesla and Data

2.1. Introduction to Tesla

With its headquarters in Austin, Tesla is a transnational American automotive and clean energy corporation. Additionally, as of 2023, Tesla will not only be the most valuable corporation in the world but also the most valuable automaker. The company dominated the battery electric car industry in 2022 with an 18% share.

2.2. Data Information

This paper will use Tesla stock data from 2016 to 2021.The source of the data set is Kaggle, a useful Machine Learning and Data Science Community. Some basic information of the price of Tesla is shown below in Table 1.

Table 1. Data information

	Open	High	Low	Close	Adjusted Close	Volume
mean	183.274022	187.265251	178.941224	183.351102	183.351102	4.441050e+07
std	227.979867	232.845126	222.443374	227.987027	227.987027	3.035737e+07
min	36.220001	36.945999	35.397999	35.793999	35.793999	8.297500e+06
25%	54.985001	55.930001	53.929001	54.994501	54.994501	2.478255e+07
50%	66.602002	67.950001	65.445999	66.756000	66.756000	3.498150e+07
75%	165.516998	168.243504	161.504005	163.856503	163.856503	5.211238e+07
max	891.380005	900.400024	871.599976	883.090027	883.090027	3.046940e+08

In this set of data, there are 7 variables in totalThe main variable in this study is Adjusted Close. In order to visualize the data, this paper also graphs each group of data separately (See Fig. 1).

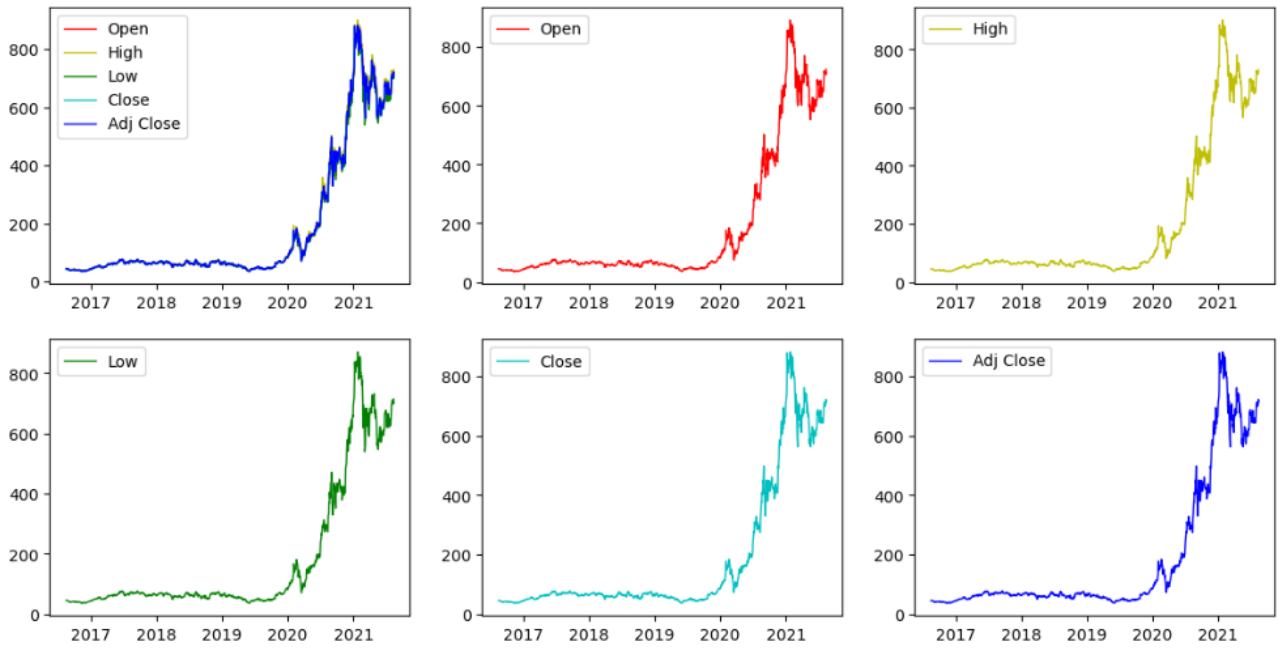


Fig. 1 Figures of variables

3. Machine Learning Methods

Since the development of artificial intelligence, a number of algorithms have been used to forecast changes in the stock market. To forecast a stock's opening price the following day or comprehend the long-term market, a combination of statistics and machine learning techniques are used [6]. Six machine learning techniques were used in this paper, namely Linear regression, Polynomial regression, XGBoost, ARIMA, Prophet and LSTM. The following are the introductions of the six models.

3.1. Linear Regression

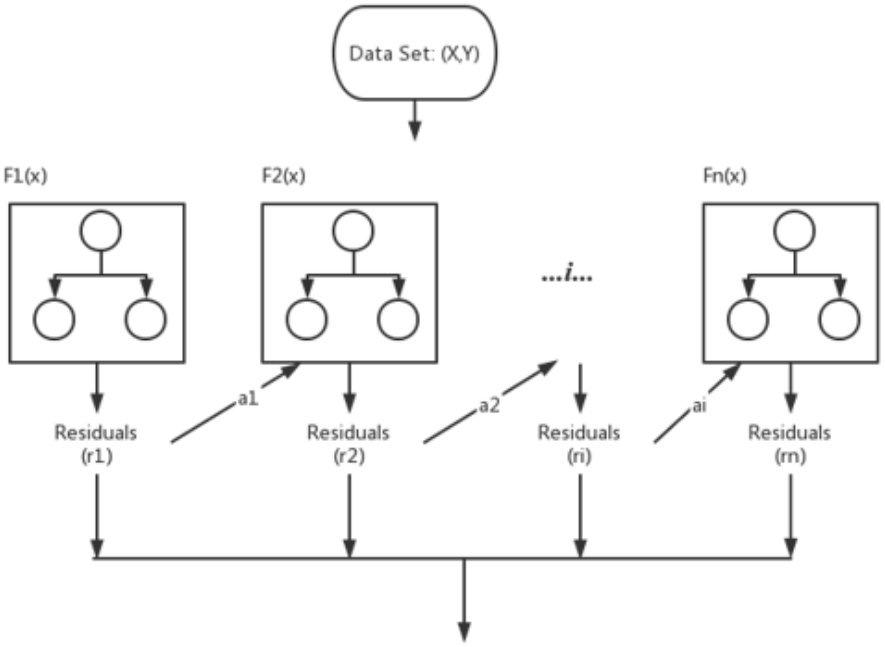
Linear regression uses the linear regression equation, a least squares function, to represent the relationship between one or more independent and dependent variables. Linear regression is the most basic model used in forecasting.

3.2. Polynomial Regression

Polynomial regression, a form of regression analysis, uses an nth degree polynomial in x to describe how the independent variable x and the dependent variable y are related. Polynomial regression can perform well in some special cases.

3.3. XGBoost

The boosting model, also known as XGBoost and GBDT, is a type of ensemble learning technique that improves classification and regression outcomes by using weak decision trees as classifiers. The structure of XGBoost is shown below in Fig. 2.



$$F_n(x) = F_{n-1}(x) + a_n h_n(X, r_{n-1})$$

Fig. 2 The structure of XGBoost

3.4. ARIMA

An autoregressive integral moving average model (ARIMA) is a sort of regression analysis. ARIMA is a popular time series analysis model. When compared to other models like as AR, RA, and so on, ARIMA adds a specific difference system, which is a typical stabilizing mechanism that assists ARIMA in dealing with non-stationary time series and improves ARIMA's applicability.

3.5. Prophet

Prophet is a method for predicting time series data that employs additive models to take into account seasonal non-linear patterns that appear every year, every day, and on special occasions. It works best when historical data covers multiple seasons and time series have substantial seasonal impacts (see Fig. 3). Prophet can often accept abnormalities, missing data, and trend changes [7].

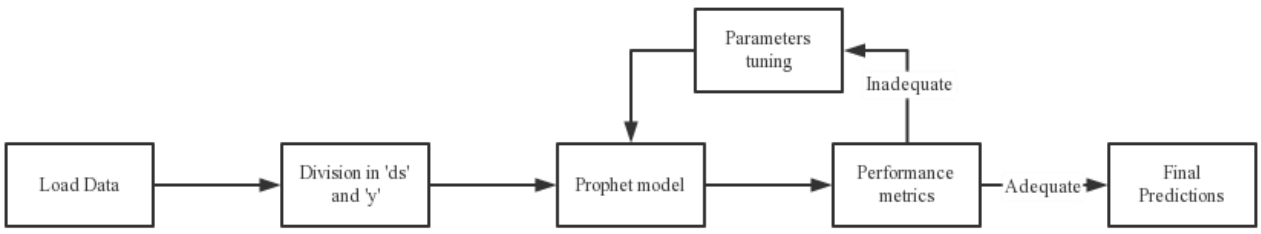


Fig. 3 Predict Stock Prices Using Prophet

3.6. LSTM

A recurrent neural network called the long short-term memory network was developed to address the vanishing gradient issue that plagues conventional RNNs. Being largely insensitive to gap length gives it an advantage over other RNNs, hidden Markov Models, and other sequence learning methods. The term "long short-term memory" refers to the attempt to give RNN a short-term memory that can endure thousands of timesteps.It can be used for time series-based data classification, processing, and prediction [8]. LSTM with a forget gate is used in this paper.

4. Criteria

In order to better judge the prediction accuracy of each model, this paper introduced four evaluation criteria, namely MSE, RMSE, MAE and Accuracy of trend predict.

4.1. MSE

MSE is a metric used to show how much the estimator and the estimated amount differ. The square of the Euclidean distance provides MSE.

4.2. RMSE

RMSE is the square root of MSE. The results of RMSE are comparable to those of MSE. RMSE may efficiently reduce the quantity of calculation and allow the conclusion that is simpler to understand when the data in the data set is too huge.

4.3. MAE

MAE is the mean absolute error. It can more accurately portray the predicted error's actual situation.In time series analysis, MAE often performs well.

4.4. Accuracy

When predicting stocks, researchers are not only concerned about the accuracy of the value, but also need to predict the trend of the stock. Therefore, this paper introduces a standard similar to the classifier to judge whether the model accurately predicts the stock [9].

5. Experimental Process and Results

5.1. Data Cleaning and Normalization

Before the experiment, the original data was first cleaned and the missing values were filled. Data cleansing is essential in quantitative research to guarantee that the findings generated from the data are as broadly applicable as possible [10].This experiment uses the normalization method MinMaxScaler that comes with python to normalize the original data

5.2. Data Seperation

There are 1258 pieces of stock data. In this experiment, all data will be divided into training set and test set in a specific way, of which 986 data are training sets, and the remaining data are test sets In order to avoid the problem of not correctly reflecting the trend of stocks that may be caused by random sampling, this experiment adopts the method of distinguishing training sets and test sets in time series, which is also a common method in stock forecasting [11].

5.3. Experimental Results

In this study, the above-mentioned training set was used to train six machine learning models, and tested with the test set. The final results are shown in the Fig. 4 below:

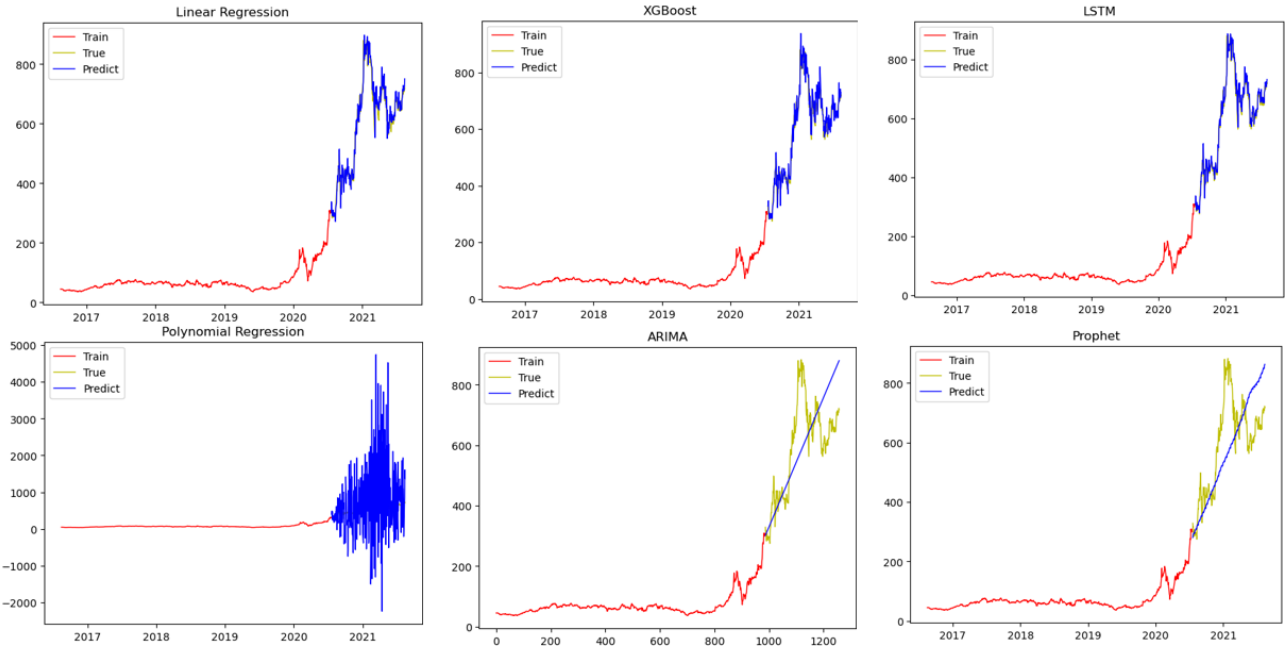


Fig. 4 Model training result

As can be seen from the above Fig. 4, the predictions of Linear Regression, XGBoost, and LSTM are relatively accurate, while the other three predictions are less effective.

To more accurately compare the six models' accuracy, this study will use the four evaluation criteria mentioned in Section 4 for further analysis. The analysis results are shown in the Table 2 below:

Table 2. Model training result

	MSE	RMSE	MAE	Accuracy of trend predict
Linear Regression	730.2	27.0	20.7	0.504
Polynomial Regression	971471.0	985.6	704.8	0.444
XGBoost	920.1	30.3	22.8	0.541
ARIMA	16561.4	128.7	99.7	0.530
LSTM	591.5	24.3	18.3	0.478
Prophet	17627.4	132.8	103.2	0.559

6. Conclusion

This experiment analyzes and predicts Tesla's stock by using 6 machine learning models, including the relatively basic Linear Regression and the currently used LSTM model. This experiment not only provides investment data support for investors interested in the new energy market, but also facilitates further stock analysis in this field in the future. In terms of models, some of the models used in this experiment do not belong to the current mainstream models. The final selected LSTM still has a ton of path for growth in the accuracy of trend prediction. Therefore, there may be some errors in the final conclusions. In addition, due to the uncertainty of the stock market itself, the current stock value predicted by the machine learning model can only be used as a reference and cannot be used as an actual accurate value.

References

[1] Shah, D., Isah, H., & Zulkernine, F. Stock Market Analysis: A Review and Classification of Forecasting Techniques. Journal of International Financial Studies, 2019, 7(2): 26.

[2] Zhong, X., & Enke, D. Forecasting daily stock market return using dimensionality reduction. Expert systems with applications, 2017, 67: 126-139.

[3] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. Predict stock closing prices using machine learning techniques. Procedia Computer Science, 2020, 167: 599-606.

[4] Jorgenson, D. W., Weitzman, M. L., ZXhang, Y. X., Haxo, Y. M., & Mat, Y. X. Can Neural Networks Predict Stock Market?. AC Investment Research Journal, 2023, 220(44).

[5] Dong, J., Chen, Y., Yao, B., Zhang, X., & Zeng, N. A neural network boosting regression model based on XGBoost. Applied Soft Computing, 2022, 125: 109067.

[6] Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 2022, 197: 116659.

[7] Bashir, T., Haoyong, C., Tahir, M. F. and Liqiang, Z. Short-term Electric Load Forecasting Using a BPNN-Optimized Hybrid Prophet-LSTM Model. Energy Report, 2022, 8: 1678-1686.

[8] Shesansky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) networks. Physics D: Nonlinear Phenomena, 2020, 404: 132306.

[9] Lu, R., & Lu, M. Stock trend prediction algorithm based on deep recurrent neural network. Wireless Communications and Mobile Computing, 2021: 1-10.

[10] Osborne, J.W. Data Cleaning Basics: Best Practices for Handling Extreme Fractions. Neonatal and Infant Care Reviews, 2010, 10(1): 37-43.

[11] Lim, B. & Zohren, S. Time Series Forecasting Using Deep Learning: A Survey. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209.