A Comparative Study of LSTM Variants in Prediction for Tesla's Stock Price

Jianyao Li*

Department of Management, Shandong University, Jinan, China

*Corresponding author: 202000271040@mail.sdu.edu.cn

Abstract. Long short-term memory (LSTM) is widely used in the stock market to train the prediction model and forecast future stock prices. Applying the LSTM method to research may incur some problems and facilitate the improvement of the method. Therefore, many LSTM variants are put forward under different circumstances. This paper surveys four LSTM variants, including Vanilla, Stacked, Bi-directional, and CNN LSTM on two different data sets regarding Tesla's stock price. Two data sets mentioned in this paper represent different stock types. To be more specific, data set 1 refers to stocks with a single long-term trend, while data set 2 can be seen as an example of stocks with more complexity. The result shows that the Vanilla LSTM reaches the highest prediction accuracy on the data set without any irregular shift in the long-term trend. CNN LSTM also provides decent predictions for the stock price. Otherwise, the Stacked LSTM performs the best for stock prediction. Bi-LSTM and CNN LSTM are also suitable for stock forecasting in more complicated situations. The change in preference for model selection proves that a company's operation situation and market circumstances also influence the prediction performance of LSTM variants.

Keywords: Vanilla LSTM; Stacked LSTM; Bi-LSTM; CNN LSTM; Tesla.

1. Introduction

Stock is viewed as an important component of the virtual economy. The variation in stock prices can reflect people's attitudes towards the company and related markets. Moreover, the prediction for the price is closely connected with people's expectations for the future. An accurate prediction of the stock price will help both the company to improve strategic decisions and the investors to purchase and sell the stock.

However, forecasting the stock price is never an easy task. Firstly, the stock price depends on various factors, including numerical variables (close price, volume, turnover, etc.) and qualitative features such as policy and financial news. Additionally, a certain company's stock is in a dynamic system that never changes alone. A particular stock's movement can interact with other correlated stocks. Even though the price can be accurately predicted, this prediction will influence people's investing behavior. One of the most demanding tasks in the field of machine learning is still the prediction of stock prices.

The R&D of new energy has been a prevalent topic and an increasing number of people are contributing to this field. Substituting fossil fuels with eco-friendly new energy tends to be an inevitable trend, which can benefit the sustainable economy. Nowadays, Tesla plays a leading role in the new energy stock market. Therefore, it's of great significance to predict Tesla's stock price.

Fig.1 depicts the close price of Tesla from January 2018 to August 2022. Before 2020, the close price was a stationary time series. According to Fig.2, the close price varies marginally around the mean value ($59) during this period. And the close price information from 2020 to 2022 is shown in Fig.3. It can be concluded that the price shows an overall rapid increase trend, even though the close price has experienced a fluctuation since 2022. The shift in the long-term trend may bring more challenges to making predictions, resulting in the loss of predictability with the same model. So, Tesla's stock is quite representative and deserves to be researched.

Figure 1. The close price information from 2018 to 2022

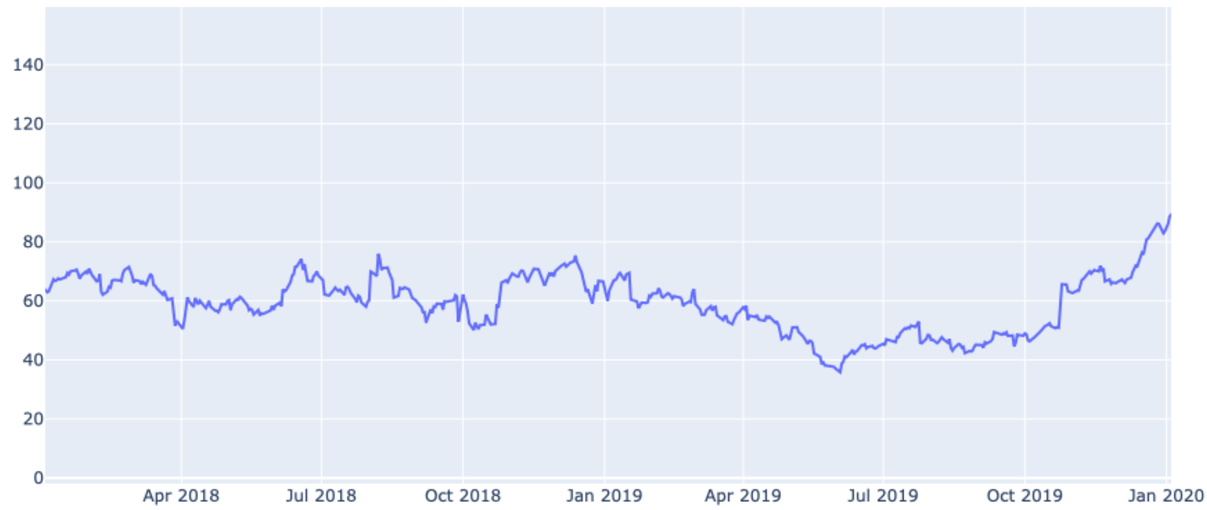

Figure 2. The close price information from 2018 to 2019

Figure 3. The close price information from 2020 to 2022

In this paper, an empirical study was conducted on four LSTM variants, including the Vanilla LSTM, Stacked LSTM, Bi-LSTM, and CNN LSTM. The close price of Tesla from Jan 2018 to Aug 2022 was downloaded from the Straight Flush. This study sets all four LSTM variants as univariate models. Two data sets are used to train and validate the models under different circumstances. After a comparison of the prediction accuracy, the best strategy for model selection can be concluded.

This study trains the LSTM variants with the same data set, so the accuracy can serve as a reliable criterion for the model assessment. Additionally, with data set 1 and data set 2 representing the simple and complicated conditions respectively, the performance of the same method is also comparable. The research gives the best prediction model for Tesla among the four LSTM variants according to the experiment results. Moreover, it also provides suggestions for model selection when considering the operation condition of the company.

The remaining part of this paper is organized as follows. In the next section, some basic machine learning and deep learning models in the stock market are introduced. Section 3 is about the structure of the data sets and the LSTM variants model as well as the evaluating indicators. The results and analysis are given out in section 4. In the final section, conclusions are drawn according to the motivation. And after analyzing some deficiencies in this study, the direction for future work is listed.

2. Literature review

2.1 Traditional machine learning method

The linear regression algorithm defines a linear relationship between the predictor variables and the response. The least-square approach is widely used in estimating the coefficients. To solve the overfitting problem, dimensionality reduction together with cross-validation are implemented. However, the non-linear feature of the stock market limits the performance of linear regression models on stock price forecasting. Also, various factors affecting stock prices are hard to quantify and they cannot be included in the model.

With the advent of Support Vector Machines (SVM), the model has become widely available due to the ability to extract non-linear features. The training process of SVM is similar to solving a quadratic programming problem with linear constraints, so the solution obtained is bound to be the optimal solution. Therefore, SVM can avoid the problems of overfitting and the curse of dimensionality [1]. The application of SVM in stock is known as Support Vector Regression (SVR). In applications, SVR still suffers from problems such as kernel function selection, difficulty in tuning too many parameters, and the extraction of shallow features [2]. With the stock price being pretty stochastic, time series influenced by multi-variables will contain much noise and cause SVM to underperform.

A popular supervised machine learning algorithm for classification is K-Nearest Neighbor (KNN). A new case can be categorized into the majority of the K closest observations. In practice, the Euclidean distance is the most frequently used indicator to select the K-closest points to the given data [3]. KNN can reach a decent accuracy without paying too much attention to tuning parameters. But KNN cannot respond rapidly to the sudden change in price.

2.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) is a specific neural network. With the help of feedback connections, the current state can be updated according to previous states and current inputs [4]. The stock price forecast can benefit from the feedback mechanism of RNNs. Only a single feedback connection still benefits RNNs in learning underlying patter from a time sequence. But the prediction emphasizes more on recent information rather than historic data. This can result in short-term memory [5].

2.3 Long short-term memory

Long short-term memory (LSTM) is a sub-type of RNNs where prior-state information can be retained [6]. LSTM was proposed in 1997, and this method introduced gates into the cell structure to avoid vanishing gradients and exploding gradients. During the training process, LSTM can learn the long-term dependency which is of great importance in the stock market. Fig.4 illustrates LSTM structures with three gates.

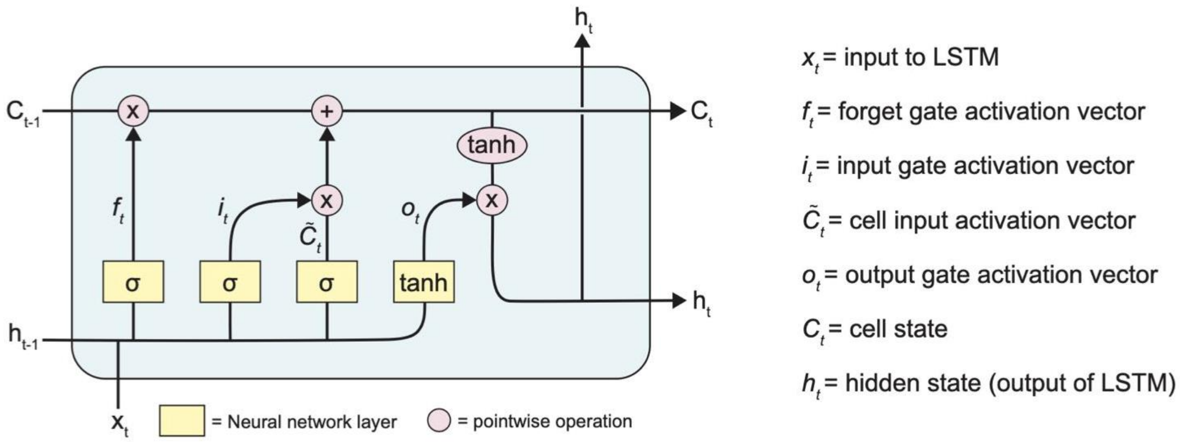

Figure 4. LSTM structure, downloaded from [7]

Forget gate decides on the memory to be discarded from the previous state. Updating the input information relies on the input gate. These two gates determine the current cell state. As for the output gate, it controls the number of outputs from the current cell state. Sigmoid function plays an important role in three gates. The sigmoid function gives out the value between 0 and 1. The mechanism is that 0 means the information will be completely discarded and 1 means to be completely retained.

Researchers [8] mentioned that the LSTM method can realize all satisfying results derived from RNNs. Therefore, LSTM is believed to have a promising future in the deep learning field. In reality, LSTM models are used to deal with an expanding number of machine learning tasks with time series data [4].

2.3.1 Vanilla LSTM

Vanilla LSTM is an LSTM model which has a single hidden layer of memory cells and predicts with an output layer [9]. In 2017, researchers presented the investigation of eight LSTM variants on tasks in three main fields and emphasized the distinguished performance over a range of data sets [10]. Ngoc Hai, Pham, et al. studied the first paper to predict the stock price with this model and found the results to be promising [11]. The result illustrated Vanilla LSTM’s outstanding capacity to analyze long-term dependency.

2.3.2 Stacked LSTM

Stacked LSTM identifies the model where various hidden LSTM layers can be stacked one by one [9]. Stacked LSTM has a more complicated structure than Vanilla LSTM. However, according to Zou and Qu, this doesn’t mean that the Stacked LSTM model is bound to overperform LSTM [13].

2.3.3 Bi-directional LSTM

Bi-directional LSTM (Bi-LSTM) operates with two layers, processing the data sequence in two directions at the same time [12]. Bi-LSTM learns the input sequence both forward and backward, and the information is concatenated for interpretation. J. Shah, R. Jain, V. Jolly, and A. Godbole tried to predict the close price for various stocks with the Bi-LSTM model [3]. They found the predicted values to be pretty accurate and close to the true values.

2.3.4 CNN LSTM

Convolutional Neural Networks (CNNs) are popular in extracting useful and deep features from raw data. According to Lu, et al., the CNN LSTM model uses CNN to capture useful features from the input, and LSTM is for predicting [14]. The result proves that the combination of the strengths of the two methods can provide a reliable stock price prediction.

3. Methodology

3.1 Data structure

The close price of Tesla from January 2018 to August 2022 was downloaded from Straight Flush. Data set 1 is the subset containing the close prices from Jan 2020 to Aug 2022. Data set 1 represents stocks with a single trend. The entire data set from Jan 2018 to Aug 2022 is used as data set 2. Data set 2 experiences an irregular shift in the long-term trend (from a stable one to a rapidly increasing one), and this can be seen as an example of stocks with more complexity.

Table 1 shows related information about two data sets for empirical research. In this study, the past 10 days' close prices are used to predict the next day's price, so all data sets only include one feature. Data set 1 contains 666 samples, and data set 2 has 1169. Both data sets are split into a training set and a test set in the ratio of 4:1. The prediction results on the test set are subsequently used to measure the performance of the applied models on different data sets.

Table 1. Data structure

	Time Range	Data size	Training set	Test set
Data set 1	Jan 2020 — Aug 2022	(666,1)	(532,1)	(134,1)
Data set 2	Jan 2018 — Aug 2022	(1169,1)	(935,1)	(234,1)

3.2 Model architecture

Fig.5, Fig.6, Fig.7, and Fig.8 illustrate the architectures of four LSTM variants in the study.

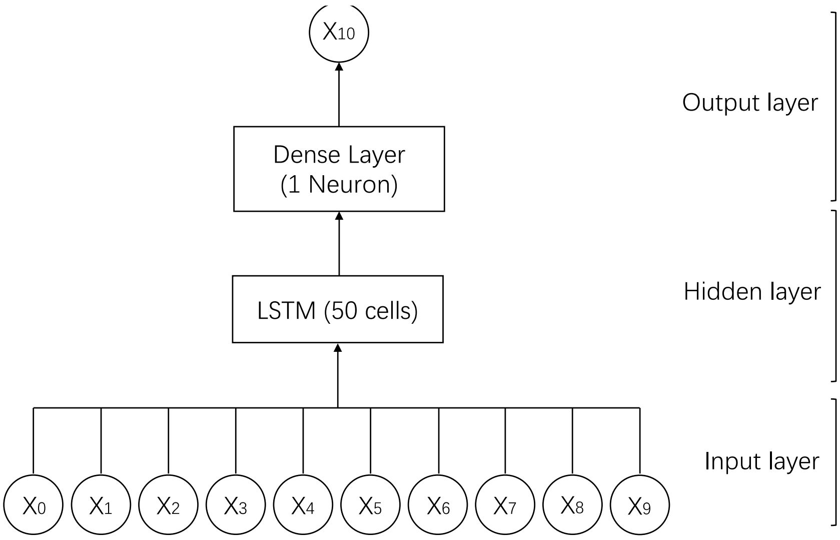

Figure 5. Vanilla LSTM architecture

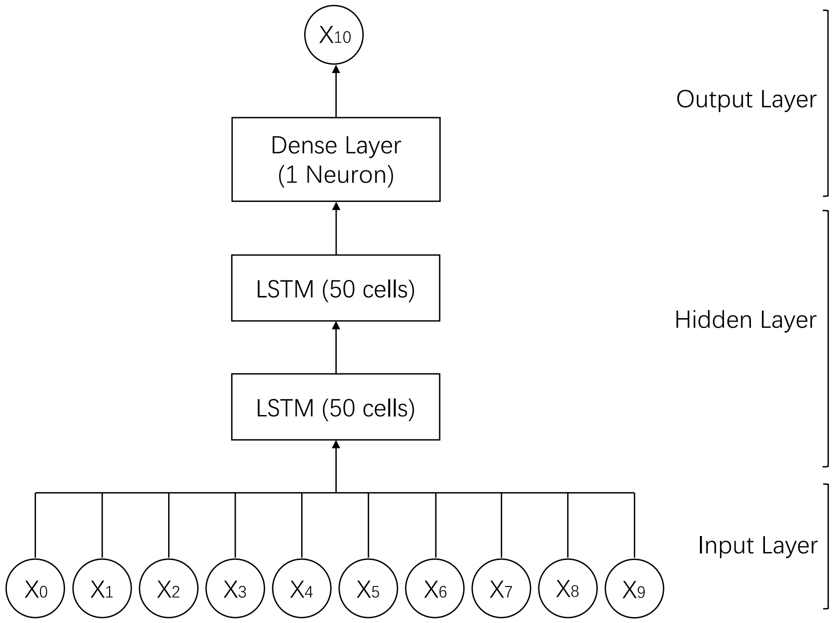

Figure 6. Stacked LSTM architecture

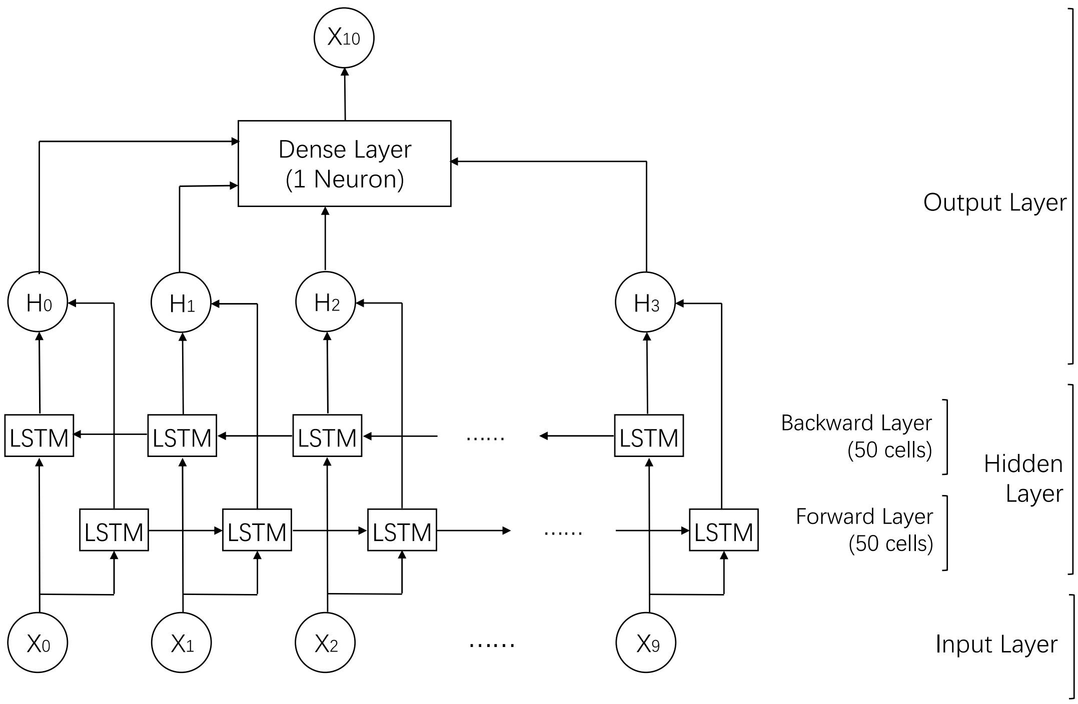

Figure 7. Bi-LSTM architecture

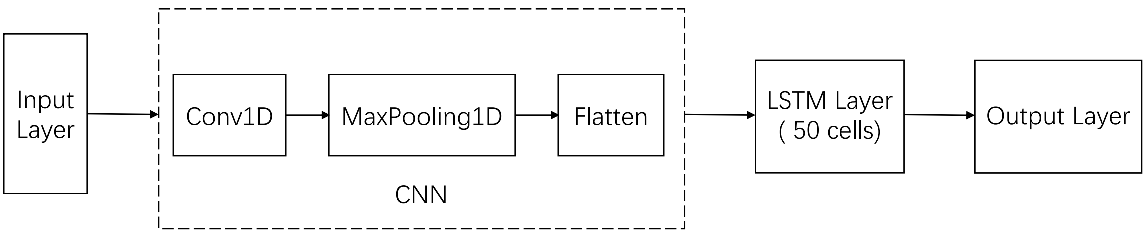

Figure 8. CNN LSTM architecture

3.3 Model evaluation

The prediction performance is evaluated by the Root Mean Square Error (RMSE) and accuracy. Generally speaking, the smaller the RMSE, the closer the predicted value is to the true value. However, it is not easy to interpret the difference caused by different models and types of data sets. Therefore, accuracy can be a better indicator for model evaluation.

By comparing today’s real close price and yesterday’s real close price, a day can be labeled as a 'rise' or 'fall'. By comparing today's predicted close price and yesterday's real close price, a day can also be labeled as a 'rise' or 'fall'. Two corresponding labels prove that the model has made an accurate prediction. And the accuracy is the ratio of accurate prediction to total prediction on the test set. The higher the accuracy is, the better the model can predict.

4. Results

Due to the setting of random seeds in Python, repeating the experiment can produce different results that sometimes may be misleading. To improve the authenticity, each data set is used to fit the four LSTM variants 10 times respectively. All results are derived from the performance on test sets.

Table 2 shows the final result for data set 1. Vanilla LSTM reaches the highest mean accuracy at 71.63%, and CNN LSTM is 0.09% lower than that. The mean accuracy for the Stacked LSTM model is 70.70% and Bi-LSTM (67.48%) is the only model with a mean accuracy lower than 70%.

Table 2. Result (data set 1: 2020-2022)

	Vanilla LSTM		Stacked LSTM		Bi-LSTM		CNN LSTM	
Experiment	RMSE	Accuracy	RMSE	Accuracy	RMSE	Accuracy	RMSE	Accuracy
1	36.68	72.36%	44.64	69.92%	35.98	69.92%	40.11	73.98%
2	35.99	73.17%	42.13	70.73%	37.46	73.17%	37.60	73.98%
3	37.92	72.36%	48.14	69.92%	38.52	72.36%	41.09	69.11%
4	39.23	69.92%	36.88	71.21%	50.58	65.85%	39.93	73.17%
5	35.88	73.17%	43.12	69.11%	36.82	70.73%	36.18	71.54%
6	40.64	69.92%	37.44	71.54%	43.39	68.29%	36.11	72.36%
7	36.12	70.73%	36.56	70.73%	79.67	57.72%	36.43	70.73%
8	36.95	70.73%	36.85	70.73%	62.60	62.60%	39.88	69.11%
9	35.72	72.36%	38.50	71.54%	38.21	71.54%	43.20	69.11%
10	36.67	71.54%	37.44	71.54%	59.71	62.60%	36.28	72.36%
Average	37.18	71.63%	40.17	70.70%	48.29	67.48%	38.68	71.54%

Table 3 shows the final result for data set 2. This time, the Stacked LSTM makes the best prediction, with a mean accuracy of 71.26%. Bi-LSTM (70.13%) and CNN LSTM (70.22%) also provide good predictions for the close price. The mean accuracy for the Vanilla LSTM is 67.76%, which is 3.87% lower than before.

Table 3. Result (data set 2: 2018-2022)

	Vanilla LSTM		Stacked LSTM		Bi-LSTM		CNN LSTM	
Experiment	RMSE	Accuracy	RMSE	Accuracy	RMSE	Accuracy	RMSE	Accuracy
1	44.71	69.51%	44.32	71.75%	41.24	73.09%	55.47	69.51%
2	43.95	67.71%	44.72	74.44%	42.37	73.09%	42.63	70.85%
3	46.70	69.51%	46.33	73.09%	50.15	66.37%	43.65	72.65%
4	58.99	65.92%	56.56	69.06%	44.56	71.30%	46.68	71.30%
5	53.36	68.61%	69.48	62.78%	49.25	69.51%	42.85	71.75%
6	67.76	62.78%	42.94	73.54%	74.00	60.54%	45.86	70.85%
7	60.77	65.47%	42.54	73.99%	41.37	74.44%	51.97	66.37%
8	48.89	69.06%	44.27	73.09%	40.79	74.89%	43.62	73.54%
9	48.43	69.06%	53.61	65.92%	60.04	63.23%	70.64	62.78%
10	43.36	69.96%	42.25	74.89%	40.34	74.89%	45.50	72.56%
Average	51.69	67.76%	48.70	71.26%	48.41	70.13%	48.89	70.22%

It's noticed that the features of the data set have an impact on the performance of models. With a single long-term trend data sequence, Vanilla LSTM outperformed the other three models. There isn't a significant difference between the performance of Vanilla and CNN LSTM. According to Hai, P. N., et al. [11], a too complex model may cause overfitting on the training set and generalization problems on the test set. This theory can explain the poorer performance on the test set 1 with Stacked LSTM and Bi-LSTM methods. However, when the historic data shows a more complex trend, such as data set 2, which has experienced an irregular change in the long-term trend, the structure of Vanilla LSTM is too simple to satisfy the complicated situation. And the other three variants are more suitable for forecasting, especially the Stacked LSTM.

5. Conclusion

This paper conducted an empirical study of four LSTM variants on two data sets about Tesla's close price. Overall, all models' accuracies are higher than 50%, which demonstrates the good prediction ability of four LSTM variants. According to the findings of this study, the Vanilla LSTM may be the best method for making predictions on a data set without an irregular shift in the long-term trend. CNN LSTM also provides a decent prediction for the stock price. Otherwise, the Stacked LSTM can be the best choice. Bi-LSTM and CNN LSTM are also suitable for forecast in more complex situations. The change of preference for model selection proves that the data structure also influences the prediction performance of LSTM variants. Therefore, an optimal method fitting all situations does not exist.

There are several deficiencies in this work. Firstly, compared with other related papers about Tesla’s stock price prediction, the accuracy in this study is relatively lower. Besides, since the next day’s close price may not vary too much compared with today’s price, it seems to be reasonable that all models can make a satisfying prediction for the next day’s price. Moreover, as this study only focuses on the univariate LSTM models, many features which are of great significance in the stock market cannot be included in the model. This may limit the predictive ability of the methods.

As for the direction for future work, the integration of LSTM and sentiment analysis can be used to predict the next seven day’s stock price or even for a longer period. This paper proves that policies to promote green energy have changed Tesla’s operation condition and further influenced the model selection. With sentiment analysis, a proper LSTM variant is more likely to be chosen. And information extracted from non-numeric factors can also be included in the model to improve the prediction accuracy.

References

[1] Aggarwal, Divya, Shabana Chandrasekaran, and Balamurugan Annamalai. "A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices." Journal of behavioral and experimental finance 27 (2020): 100335.Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[2] Zhang, Lei. Research on stock market trend analysis and prediction based on support vector machine.2020. Nanjing University of Posts and Telecommunications, MA thesis.

[3] Shah, Jaiwin, et al. "Stock Market Prediction using Bi-Directional LSTM." 2021 International Conference on Communication Information and Computing Technology (ICCICT). IEEE, 2021.

[4] Leevy, Joffrey L., Taghi M. Khoshgoftaar, and Flavio Villanustre. "Survey on RNN and CRF models for de-identification of medical free text." Journal of Big Data 7.1 (2020): 1 - 22.

[5] K. D. Gunawardana, An Introduction to Artificial Neural Networks for Accounting and Finance Modelling, Colombo: Author Publication, 2009.

[6] Smagulova, Kamilya, and Alex Pappachen James. "A survey on LSTM memristive neural network architectures and applications." The European Physical Journal Special Topics228.10 (2019): 2313 - 2324.

[7] Orsel, Ogulcan E., and Sasha S. Yamada. "Comparative Study of Machine Learning Models for Stock Price Prediction." arXiv preprint arXiv:2202.03156 (2022).

[8] Yu, Yong, et al. "A review of recurrent neural networks: LSTM cells and network architectures." Neural computation 31.7 (2019): 1235 - 1270.

[9] Brownlee, J. Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow Using Keras; Machine Learning Mastery: Melbourne, Australia, 2017.

[10] Greff, Klaus, et al. "LSTM: A search space odyssey." IEEE transactions on neural networks and learning systems 28.10 (2016): 2222 - 2232.

[11] Ngoc Hai, Pham, et al. "An Empirical Research on the Effectiveness of Different LSTM Architectures on Vietnamese Stock Market." 2020 International Conference on Control, Robotics and Intelligent System. 2020.

[12] Althelaya, Khaled A., El-Sayed M. El-Alfy, and Salahadin Mohammed. "Evaluation of bidirectional LSTM for short-and long-term stock market prediction." 2018 9th international conference on information and communication systems (ICICS). IEEE, 2018.

[13] Zou, Zhichao, and Zihao Qu. "Using LSTM in Stock prediction and Quantitative Trading." CS230: Deep Learning, Winter (2020): 1 - 6.

[14] Lu, Wenjie, et al. "A CNN-LSTM-based model to forecast stock prices." Complexity 2020 (2020).