# Analysis of the Different Models for Stock Price Prediction

Zhehao Lu *

Suzhou Foreign Language School, Suzhou, China

* Corresponding Author Email: sflsdzb@sfls.com.cn

**Abstract.** As a matter of fact, the stock price prediction is sustained for many years when stocks appeared. People are finding ways to predict the stock prices and find the best price for earning most profits. On this basis, this study will discuss about the stock price prediction in terms of various models including time series models as well as machine learning and deep learning schemes, there are different types of models, and these models have different uses and methods. According to the analysis, this paper find ARIMA model, recurrent neural model, and logistic regression model are available for calculating the stock prices. This is proved by using the data before and calculate the data afterwards. The prices calculated is similar to the real prices. Which means these methods or models are useful for predicting. Overall, these results shed light on guiding further exploration of stock price prediction based on the state-of-art machine learning scenarios.

**Keywords:** Stock price prediction; ARIMA; recurrent neural network; Logistic regression model.

## 1. Introduction

The stock price prediction has been favored all over the world, everyone wants to have the ability to predict the stock price and then earn money. The problem has been for 400 years [1], as the stock has been given, and this will improve the concentration of capitals and improve the organic composition of enterprise capital [2], this will then have economic growth, and benefit the whole society. The stock market been made for the society to trade and get investments [1]. The trading price can then be predicted to find the best price for trading, and then earn more profits, the risks will not be danger as usual. Scientists then find ways to seek more methods for prediction. Human can plan about the future and motivate by imagine to be rich [3].

Human now develop more ways and more models to predict the stock price. Time Series Relational Models [4], a model that according to the time, the time changes will affect the prices. There might be a cycle between the times, as one figures out the loop, then one can find the stock price immediately. The other finding is using reinforcement learning, since Markov process which can be used in the stock prediction. One can then do the experiment to prove the functions, finding the relations is correct. This will be a more effective way then the other indicators [5]. GWO-ENN [6], is saying that can predict the price that is closed, it has the hidden, input and output layers. Grey wolves' algorithm is better than other algorithms, the feedback connection shows time delay between input and output pattern [6]. The other findings are that using LSTM-based deep learning regression models is superior to predicted, this is a type of model that uses fifty different stock prices, and then put the movements into the computer, and tried to let the computer be trained. Afterwards, let the computer to predict the movement in the short run and long run [7]. The different models are needed to take different data.

As interested in predicting for the future, and the stock price is a chance for anyone to become rich in the future. Everyone thinks the lucky guy will be them, how to prevent the risks and taking the most profit is the problem. Also, how the probability is different, is the rise and fall will be the same probability? Or the probability has one side higher. In the next few parts, this study would like to talk about the three different models that been used for predicting the stock price. There will be ARIMA, RNN and Logistic regression model. Then I will find the how the data are used in different models, and how these models work. Then give out the formula need for every model and the different applications of each model can use, not only about the stock price predicting, then give some of the graphs using the models and see the differences of that with the originals. At last, this study will find whether the model is available for predicting stock prices.

## 2. Basic Description

There are many ways to predict the stock price. The two of the ways are reinforcement learning, and long short-term memory. The reinforcement learning has two actions, which is state and action [8]. When there are problems of the data and results, the computer will change the results and the strategy, when there are new data. They will get the results after the strategy. The reward of learning will be the profits and losses, the factors will be states, trading is the action [8]. The reinforcement learning need to use RNN-LSTM model for predicting the stock price [9]. It uses TD algorithms, the computer will learn the experience directly, then one uses the price of the first day, highest price during the day, lowest price in the day, the last price traded, and the numbers of shares or contracts [5]. It can then use the average root mean-square to find the errors of the real and predicted values. The difference means there will be an error in it and can't be extreme [5]. The another is using the experiment data that is not invalid and then set the parameter, this will not have the error during the process, then select ten stocks and let the computer predict these. The distribution of the results will find the data accuracy, findings were when the time is too short or too long, the output isn't accurate [5]. The ways of finding the results aren't correct needs to compare with the data of originals, the data used for machine learning shouldn't be all, just take a big part.

## 3. ARIMA model

The ARIMA model is a model that has many factors that can influence the results. There will be 3 influences, the p, which is number of autoregressive terms, d is the number of differences, and q is the number of moving averages [10]. A small BIC, standard error, high R square is the criteria used for study each stock index [3]. The application of the ARIMA model is used in predicting the price, the price of the Nokia has been predicted by using the ARIMA model. The left panel of Fig. 1 is the real Nokia stock price and drawn in a graph, and then the stock price is predicted by the ARIMA model by giving the values of variables. Then the computer uses the formula and predict the prices. After that, the right panel of Fig. 1 is drawn, which is the graph of the differences of the price of usual Nokia stocks and the price of predicted Nokia stocks.

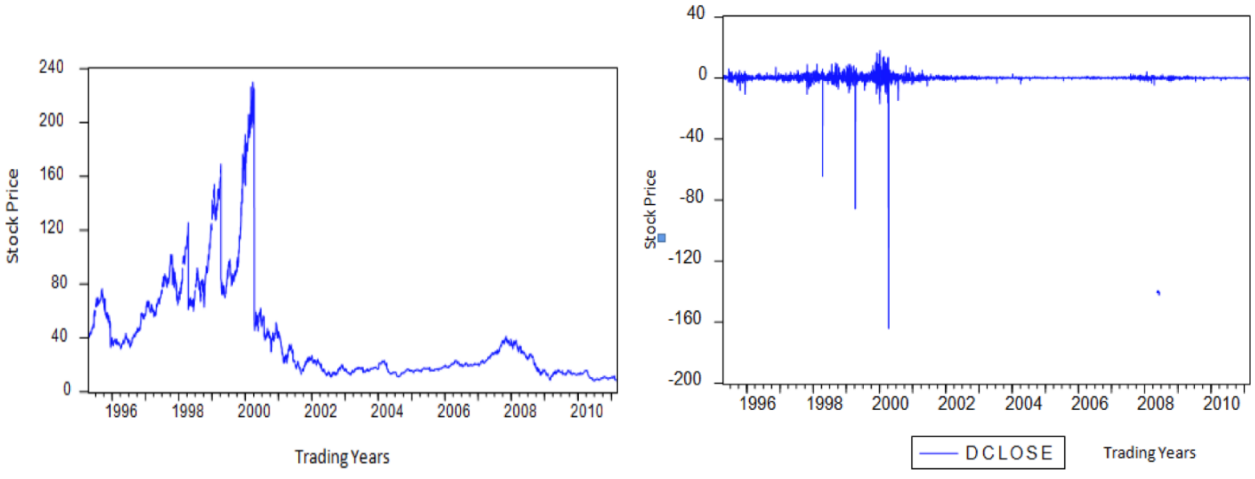


**Fig. 1** An example of price prediction.

The data should be stationary, for example the banking stock market data predicting, there are 2000 people, and the number of banking is given [11]. The prediction is only for short term, the long term will not be so accurate. The ARIMA (1,1,2) is the most accurate as it is smallest [12]. The another proving which is the most accurate is ARIMA (4,1,4) ARIMA (1,1,33) ARIMA (1,2,33), it uses 50% of the real data and it is shown by figure [13]. It is found that the ARIMA (1,1,33) has a better accuracy, so one can still using this model. The smaller RMSE means the accuracy is higher.

## 4. RNN

RNN is connections of the computational units form a directed circle. The equation 2 is used to define the hidden units as illustrated in Fig. 2 [14]. The LSTM is also a kind of RNN [15], the LSTM cells have input gate, cell state, forget gate, output gate, igmoid layer, and tanh layer. The deep RNN training scheme can be introduced by the 3 lines, the first line is about the whole sequences input, and the second line is input long sequence into small clips, the third line is overlapping coherence training scheme [16]. This scheme can be seen in Fig. 3 The gates are controlled by forget, memory, and output gate. The output gate will only show 0 and 1 as ignore and keeps on. Memory is for choosing which to store. Output gate will decide which to output. Previous study uses the Apples stocks as an example, it has the trading time, opening price, highest price, lowest price, and closing price of the stock as the data. The training only consists 65% of the data, the others is tested. It uses different time steps to show the performances. Both the predicting is effective, but if the time steps is larger, there will be more mistakes (seen from Fig. 4) There are a larger gap when time steps is 10 from around 700 and so on. The whole model uses 80% of data for training and the other 20% of data for testing [17]. Also uses different Epochs for training data and structuring. The [18] uses T trading days is time window size, which the input of the RNN is a feature vector of T timestamps [18]. The RNN has a big mistake in predicting the share price start from 9000 minutes of the INFOSYS. This is because RNN did not observe the change in the behavior of stock pattern. The RNN change its behavior during the 2000 to 6000 minutes. The reason is the RNN did not matches the pattern of original data.

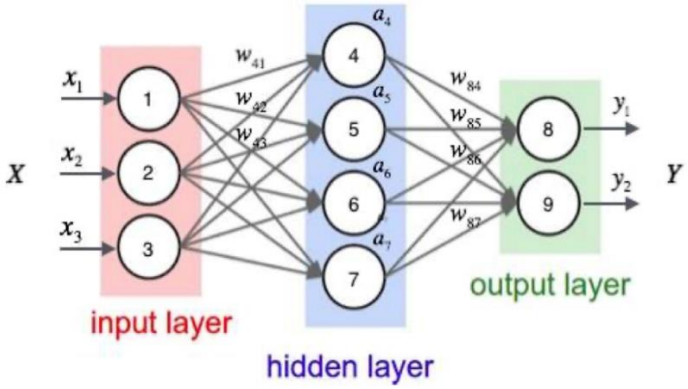


**Fig. 2** A sketch of RNN.

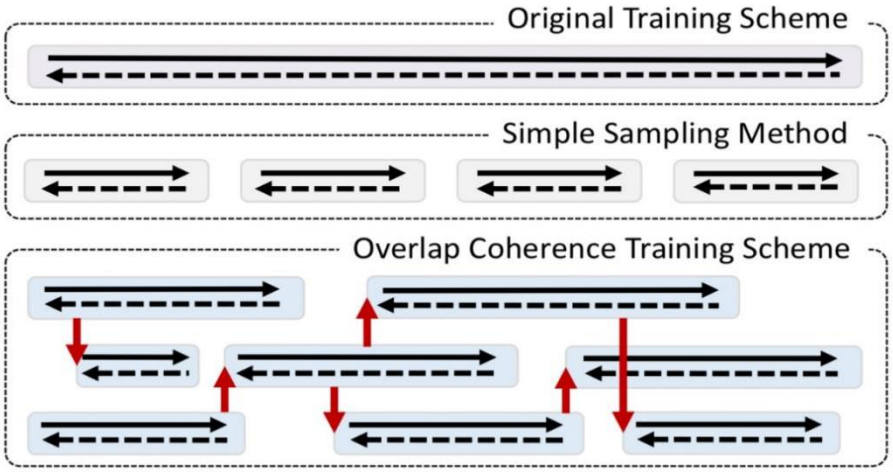


**Fig. 3** The deep RNN training.

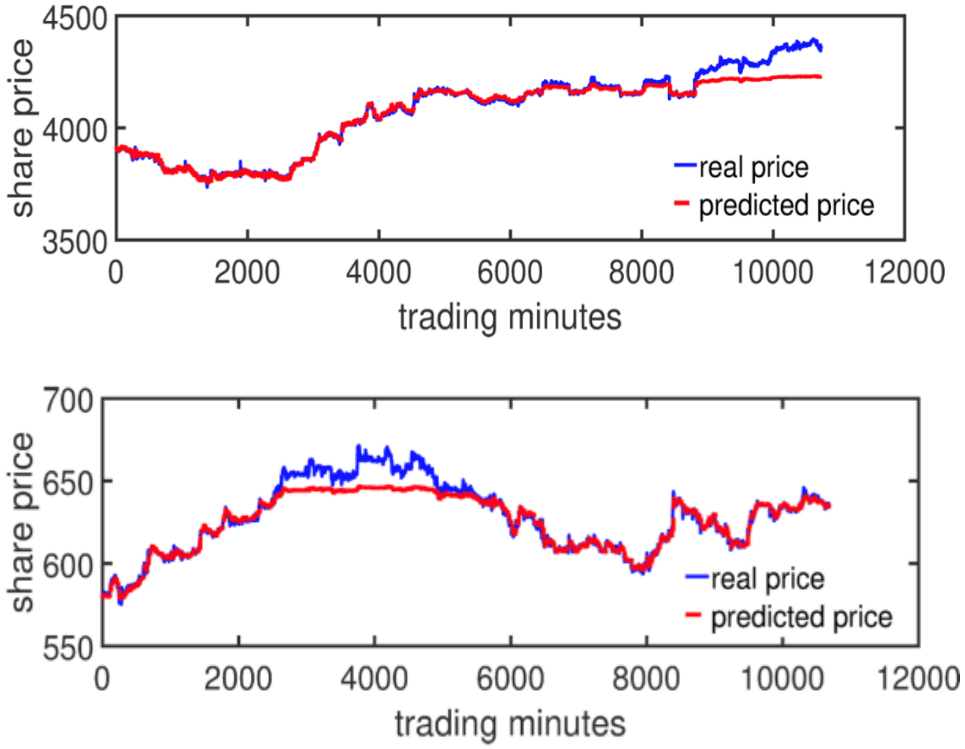


**Fig. 4** An example of the prediction results for RNN.

## 5. Logistic Regression

The logistic regression model is binomials, it only has "1" or "0". "1" represents the trends average price is higher or equal to the current month and the "0" means the average price trends is decreasing [19]. Previous close price, opening price, highest price, lowest price, closing price, Composite weight price, Daily Turnover, Amount traded, and Number traded into consideration. These are the index variables that are taken in the stock price trend predicting [20]. The predicting of monthly stock trend, it needs Stock Integrated Index for training and testing. For choosing the valid month, the month shouldn't be linear dependent and needs 9 or more to run the logistic regression model. For the month remains, it needs 1 for testing, and the others are for training.

Scholars use python to predict the stock prices, it first collects the data of news information by crawler in different websites, and tweets are collected in twitter. Lemmatization is used for making the words readable for the computer. The last is to analyze the sentiment, it is calculating the polarity. The polarity is by adding the polarity shown in the information's. Then it defines the pattern is consist for 5 days, the trend is calculated through the last 3 days data. For monthly prediction, the correlation is the measure of the differences and finding the pattern through months.

Other scholars choose the non-financial firms from 2012 to 2015, and then divide the 109 firms into poor and good. The equation is to calculate the logistic regression by the performance, and the p will be the probability is good. The other variables are independent, for example, Earnings per share, Price to Book value, Return of Equity, Current Ratio, Debt to Equity, Percentage changes of Net Sales [21]. The results of Ref. [21] is the overall prediction rate is 88.4%, the good is 87.2% and poor is 89.7%. Which means the model can be used in the prediction.

## 6. Limitations and Prospects

The stock price can't be predicted accurately in future for ARIMA model. It is because if the data is predicted wrong in the present or next few years, so the next few years will also be predicted wrong, it is not so convenient. It can be improved by being able for predicting in long terms. The recurrent

neural network will not observe the changing pattern and matches the changes. The logistic regression model has an 88.4% accurate rate. The models cannot predict the long-term prices of stocks, if there are some pattern errors, the future predicted prices will be influenced, which means the predicted data needs being independent so will not be influenced by the data predicted before.

In the future, it is better to be more accurate, there will be more models that can predict the stock price. Not only stock prices, others that related to probability will also be solved. The equation might use a few days of data, and they can predict the data afterwards, because the probability will also relate to other exterior effects, and they can exchange the results and effect into data. Since the data will be related to the probability, by using the variables related to this, the probability will be calculated more precisely. If the data can be predicted individually, the factors of data predicted before will not affect the solution of the next prices. So, the method needs to improve to make the calculating method will not be easily influenced by the wrong data, or the model can recheck the data before with another method.

## 7. Conclusion

The whole passage is talking about the models of the stock price prediction, there are ARIMA, recurrent neural network, logistic regression model. All of them needs the variables needs the stock price before and only uses most of the data for the devices deep learning, the others is computer testing. The ARIMA model has a different number, so it is important to find the number that has the biggest accuracy. The recurrent neural work has forgotten, memory, and output gate. These gates will show the prediction of the stocks by finding and replacing data. The logistic regression model is the model that needs to find all prices and then calculate the logistic regression. For the future, there will be more effective way calculating the stock prices for future, even using the present data to calculate the data before. This paper is to study the different models that can be used for predicting the stock prices, and find these models have different calculating methods and results and accuracy.

## References

[1] Jin Z, Yang Y, Liu Y. Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 2020, 32: 9713-9729.

[2] Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications, 2021, 33: 4741-4753.

[3] Ariyo A A, Adewumi A O, Ayo C K. Stock price prediction using the ARIMA model. 2014 UKSim-AMSS 16th international conference on computer modelling and simulation. IEEE, 2014: 106-112.

[4] Zhao C, Hu P, Liu X, et al. Stock market analysis using time series relational models for stock price prediction. Mathematics, 2023, 11(5): 1130.

[5] Lee J W. Stock price prediction using reinforcement learning. ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570). IEEE, 2001, 1: 690-695.

[6] Kumar Chandar S. Grey Wolf optimization-Elman neural network model for stock price prediction. Soft Computing, 2021, 25: 649-658.

[7] Mehtab S, Sen J, Dutta A. Stock price prediction using machine learning and LSTM-based deep learning models. Machine Learning and Metaheuristics Algorithms, and Applications: Second Symposium, SoMMA 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 2. Springer Singapore, 2021: 88-106.

[8] Hu Z, Zhao Y, Khushi M. A survey of forex and stock price prediction using deep learning. Applied System Innovation, 2021, 4(1): 9.

[9] Li Y, Ni P, Chang V. Application of deep reinforcement learning in stock trading strategies and stock forecasting. Computing, 2020, 102(6): 1305-1322.

[10] Fattah J, Ezzine L, Aman Z, et al. Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 2018, 10: 1847979018808673.

[11] Kong-lai Z, Jing-Jing L. Studies of Discriminant analysis and Logistic regression model application in Credit Risk for China's Listed Companies. Management Science and Engineering, 2010, 4(4): 24.

[12] Almasarweh M, Alwadi S. ARIMA model in predicting banking stock market data. Modern Applied Science, 2018, 12(11): 309.

[13] Khan S, Alghulaiakh H. ARIMA model for accurate time series stocks forecasting. International Journal of Advanced Computer Science and Applications, 2020, 11(7).

[14] Zhu Y. Stock price prediction using the RNN model. Journal of Physics: Conference Series. IOP Publishing, 2020, 1650(3): 032103

[15] Selvin S, Vinayakumar R, Gopalakrishnan E A, et al. Stock price prediction using LSTM, RNN and CNN-sliding window model. 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 2017: 1643-1647.

[16] Pang B, Zha K, Cao H, et al. Deep rnn framework for visual sequential applications. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 423-432.

[17] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network. Procedia Computer Science, 2020, 170: 1168-1173.

[18] 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: 745-753.

[19] Gong J, Sun S. A new approach of stock price prediction based on logistic regression model. 2009 International Conference on New Trends in Information and Service Science. IEEE, 2009: 1366-1371.

[20] Nayak A, Pai M M M, Pai R M. Prediction models for Indian stock market. Procedia Computer Science, 2016, 89: 441-449.

[21] Ali S S, Mubeen M, Hussain A. Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX). Patron of the Conference. 2018, 15.