Research on Microsoft Stock Price Prediction Based on Various Models

Yuanhao Fu
School of Finance, Inner Mongolia University of Finance and Economics, Inner Mongolia, China

Keywords: Machine Learning Models, Linear Regression, Microsoft Stock Price, Time Series Models, LSTM.

Abstract: With the development of social economy, stock investment is more and more popular. In the process of investing in stocks, people execute investment strategies in a quantitative trading manner, hoping to obtain the highest return with the least risk. To be successful in quantitative investing, the key is to build excellent mathematical models and grasp the accurate trading time node. The paper uses the dataset of Microsoft stock prices from April 2015 to April 2021 to build Machine learning models such as Linear regression, Time series models such as ARIMA, and LSTM model are used to fotcast the Microsoft's stock price. This thesis provides the theoretical knowledge of LSTM neural model and time series model, selects the actual stocks in the stock market, conducts modeling analysis and predicts the stock price, and then uses RMSE to compare the prediction results of several models. Since the time series model cannot get the utmost out of the non-linear part of the data and cannot carry out long-term memory, the LSTM neural network can make full use of it and long-term memory to obtain useful information in the stock data. In terms of root-mean-square error, LSTM neural network is smaller than the time series model, which indicates that LSTM neural network is a better method for prediction.

1 INTRODUCTION

Nowadays, people try to use computers to manage investment transactions, and add trading strategies to the instructions of computers quantitatively, which is called quantitative investment trading. To be successful in quantitative investing, the key is to build excellent mathematical models and grasp the accurate trading time node. One idea is to build mathematical models to predict the rise and fall of stock prices, timing them to buy on the rise and sell on the fall.

Financial data are affected by many factors and are characterized by high complexity. With the development of artificial intelligence, more people have applied machine learning to the study of financial stocks.

Stock volatility is influenced by many elements, such as historic stock price data, social media opinion, investor sentiment, etc. Stock text fusion is a high-efficiency method for forecast, but there are still some problems such as poor time dependence of historical information, low availability of experiment and insufficient validity of fusion features. The noise database, low quality of it and incomplete abnormal one in the existing situation leads to the inaccuracy of the learned characteristics and the poor prediction performance of the model. In addition, most of the subsistent sets improve the usability of it by changing the network structure of outcome, and lack in-depth research on the uncertainty factors of datum.

Chowdhury et al. in 2020 adopted machine learning methods to verify stock prediction based on improved Black code option pricing model. Akhtar et al used support vector to predict stocks in 2022. Saranya et al. in 2019 compared the results of various machine learning algorithms to predict stock price volatility and found the best means to predict stock price. Maqbool et al. used three ways to compute multifarious view scores and used them in disparate groups to comprehend the incidence of news on stock prices and the impact of each sentiment scoring method.

Keren et al. studied the virtues of CNN and LSTM in improving the accuracy of stock prediction, used the convolution idea of CNN to build a feature extraction layer to extract features, and input the extracted features into LSTM to better study the time information of features. Akshit et al. integrated ANN to establish ANN 's-MLP, GACH-MLP hybrid

Fu, Y.
Research on Microsoft Stock Price Prediction Based on Various Models.
DOI: 10.5220/0012807400004547
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 11-16
ISBN: 978-989-758-690-3
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

model, and used the new method and new technology of combining BP algorithm with multi-layer feedforward network, which achieved good results in stock prediction.

Ankit et al proposed that fusion can be considered as a method to integrate data or features and enhance prediction based on combination methods that can help each other. Their fusion applied in various stock markets is divided into information fusion, feature fusion and model fusion. Nine new intraday stock price synthesis models are proposed by Kumar to improve the accuracy of intraday stock price prediction (Chandar 2021).

2 METHODS

2.1 Data Source

The paper using the dataset of Microsoft stock prices from April 2015 to April 2021 from Kaggle, selected with 1511 observations.

2.2 Data Visualization

The on and off columns indicate the initiate and eventual price at which the stock will trade on a given day.

Observe Figure 1. Volume is the number of shares bought or sold that day. Another vital thing to note the market is closed on weekends and public holidays. The count of profit or loss is usually decided by the closing price of a stock for the day, so think about it as the target variable (Cao et al 2022 & Huang and Fang 2021).

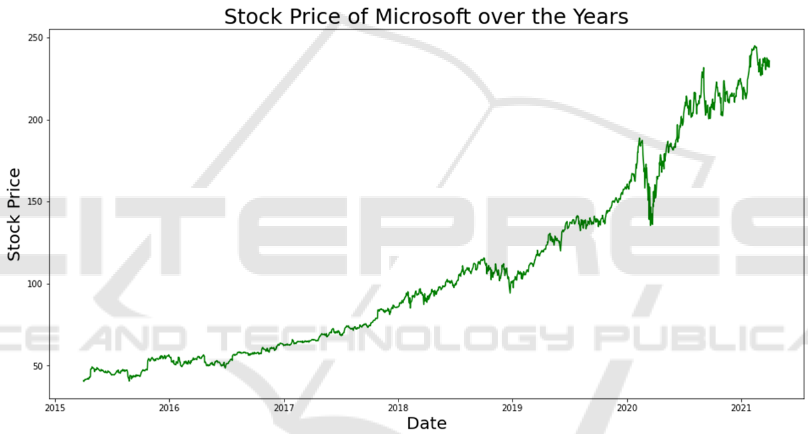


Figure 1: Stock price of Microsoft over the years (Picture credit: Original).

2.3 Model Selection

2.3.1 Moving Average

In the financial world, the moving average (MA) is a common stock indicator used in artistical analysis. The reason for calculating it is to help smooth the price data by creating a constantly updated mean price.

By calculating it, it is possible to mitigate the effects of tatted short-dated waves in stock prices over a given time.

2.3.2 Linear Regression

The analysis mainly studies the relationship between variables, using lines to fit all the data points, and then studies how to minimize the distance difference between the line and all data points. Linear regression is our most common regression analysis algorithm, it also has different names in different scenarios, such as weighted average, multifactor model, and so on.

Linear regression processes the observed data to obtain a mathematical model expression that is relatively consistent with the law, it means that the law between the independent variable data and the dependent variable data can be found, so that unknown data can be simulated.

2.3.3 KNN

k-nearest neighbour (KNN) is an elementary sort management and regression methodology, a model based on labeled drilling database. It's a surveillant learning means of count.

2.3.4 ARIMA

ARIMA: Autoregressive Moving Average Model. The model used for time series forecasting is usually suitable for single-column time series data analysis, provided that the time series data is stable and there is no obvious upward/downward trend, and the stability can be tested using ADF.

2.3.5 Prophet

Prophet is an opensource time series prediction algorithm from Facebook that can availably process holiday message and fit the changing trend of data by week, month, and year. According to the official website, Prophet has a good fitting effect on historical data with strong cyclical characteristics, which can not only deal with some outliers in the time series, but also deal with some missing values. The algorithm provides two implementations based on Python and R.

Prophet applies to business behaviour data with obvious inherent laws, such as business problems with the following characteristics:

2.3.6 LSTM

Recurrent Neural Network (RNN) is a sort of time sequence that can be preserved. The neural network structure of column state, which can use previous data to process current data, has the ability to capture useful information in the time series during the cycle. Parameters in RNN are learned by Back Propagation algorithm. Nevertheless, as the number of network layers and iterations increases during the learning process, subsequent nodes of the RNN will gradually forget the message before, leading to gradient elimination loss or gradient explosion problem. LSTM neural network introduces cell state and gating mechanism inside, which can effectively solve the problem of gradient disappearance or explosion. Its unit structure is shown in Figure 2.

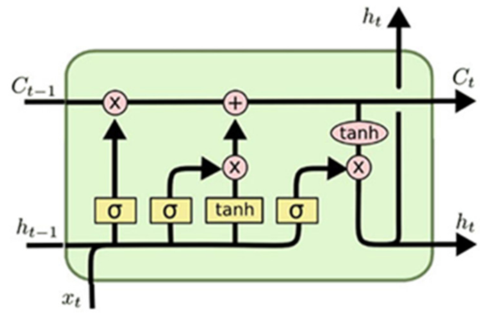


Figure 2: LSTM cell structure (Picture credit: Original).

The LSTM neural network proceeds from the received sequence data by continuously repeating the above data processing process extract useful information and output the extracted information. LSTM neural network will produce an output at each moment. For the prediction task in this paper, only the output at the last moment is used for prediction.

3 RESULTS AND DISCUSSION

3.1 Results

Looking at Figure 3, the value is about 76.62, but the results are not hopeful (as can be gleaned from the figure). The values that have been predicted have the identical scope as the conscious values in the training collection (up first and then down slowly).

The data set in Figure 4 is arranged in climbing sequence, and then a separate one is invented so that any new traits created do not influence the primordial information.

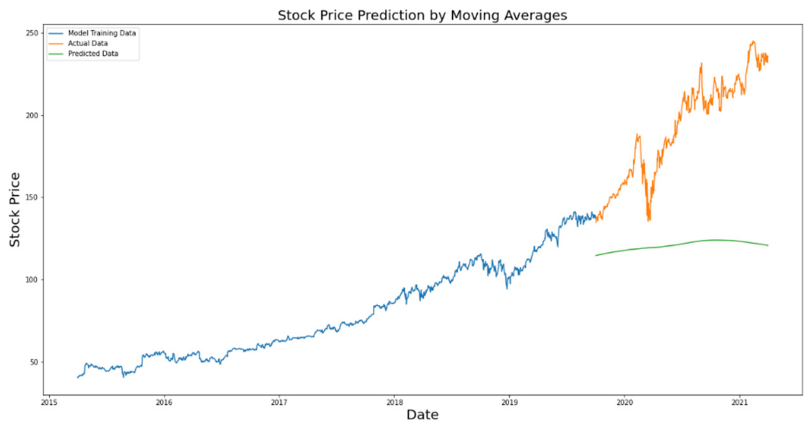


Figure 3: Moving Averages (Picture credit: Original).

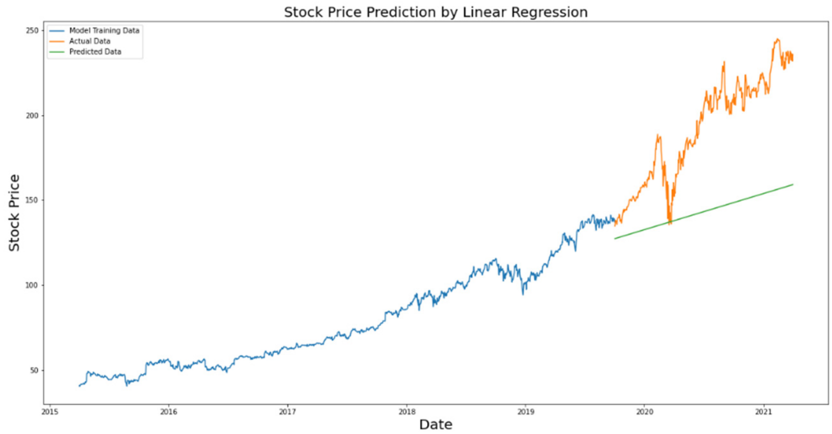
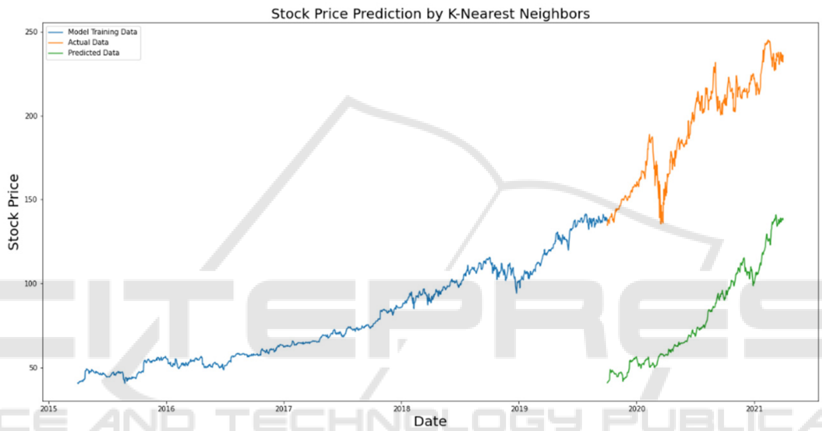


Figure 4: Linear Regression (Picture credit: Original).



SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 5: KNN (Picture credit: Original)

Observing Fig 5, The RMSE values are nearly alike to the linear regression model. As this has been the pattern for the past few years. It can be surely said this way does not show good quality on this database.

Have a look at some time series forecasting techniques and look forward how they represent when in the face of the challenge.

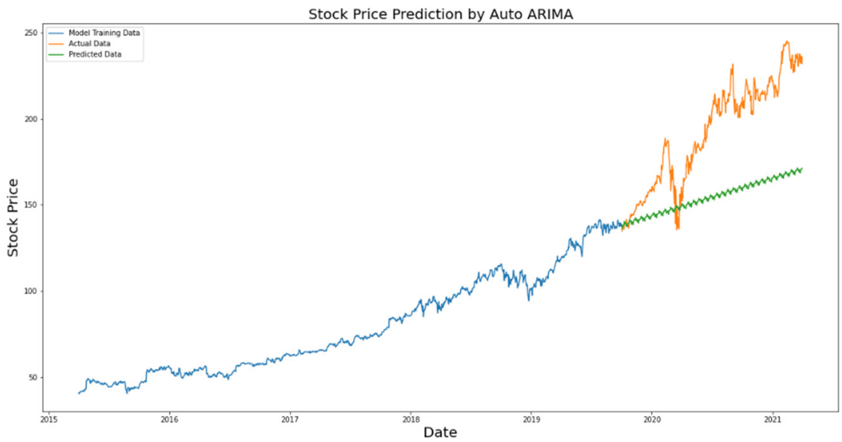


Figure 6 : Auto ARIMA (Picture credit: Original)

Observing Fig 6, the model uses lapsed data to make the pattern in the time series clear. Using these numbers, it captured an incremental tendency in the series. Although the predictions are much better, it is still not close to the real one. Seeing it, the model has exhibited a trend in the catena, but does not pay close attention on seasonal sectors.

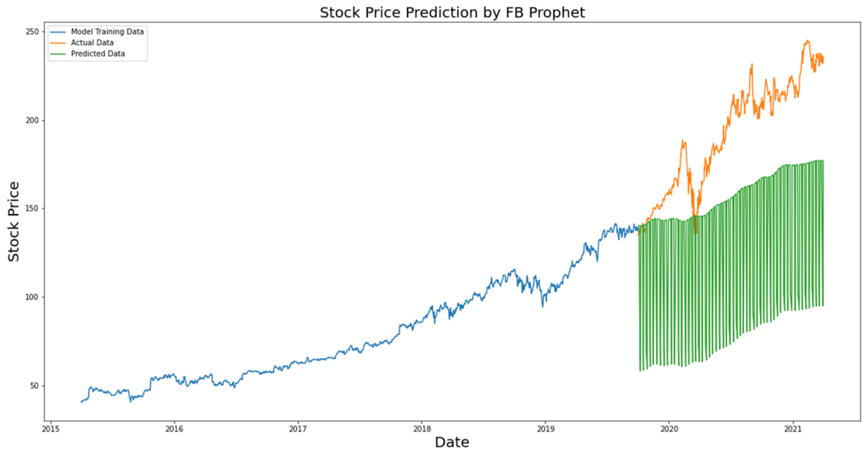


Figure 7: Prophet (Picture credit: Original).

Observing Fig 7, prophet (like others attempt to obtain the seasonal characteristic of the past. In this situation, it failed, which turns out that there is no specific trend or seasonal characteristic to stock prices. Much depends on what is happening in the market at the moment, causing prices to fluctuate. So, mantic techniques do not show good results for this given question. Attempt to another advanced technology next.

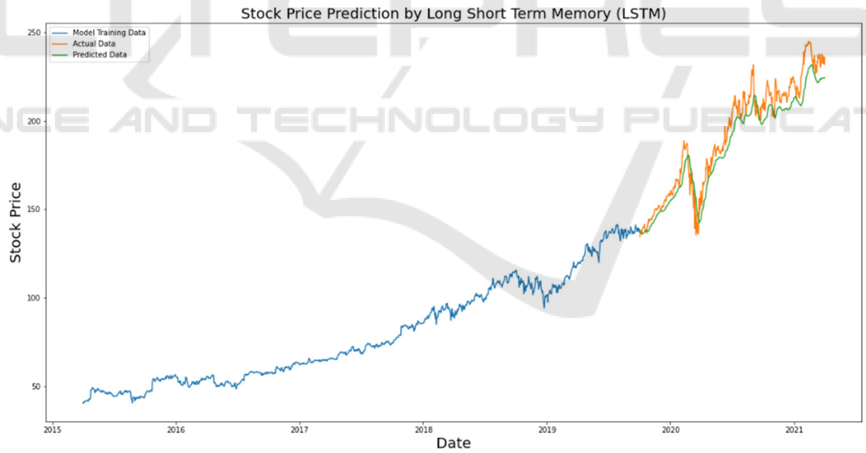


Figure 8: LSTM (Picture credit: Original).

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

3.2 Discussion

Looking at Figure 8, the LSTM model can be adjusted by increasing the dropout value, or the epoch. But are LSTM's forecasts sufficient to determine whether the share price will rise or fall? Of course not! As talked by the author at the beginning of the paper, stock prices can be impressed by company journalism and other factors such as demonetisation or merger/break-up of a company. There are also intangible factors that usually cannot be predicted in advance. The model evaluation results can be seen as Table 1, and the LSTM method has a better performance.

Table 1: Model Evaluation (RMSE).

Linear Regression	KNN	Auto ARIMA	Prophet	LSTM
58.366	112.947	43.470	69.194	9.465

4 CONCLUSION

To sum up, it can be seen that LSTM prediction model's RMSE is smaller and more accurate than other models. The results show that the LSTM is better. However, In the real market, the elements that can impress the movement of stock prices are numerous and intricate. And these factors can also be related, so try through the essay relatively simple assumptions about trading conditions predict stock prices and spot speculative opportunities seems impossible. Stock price prediction model can be proposed for stock market investors for reference, improve the rationality of investors and increase the effectiveness of the stock market, which can improve the efficiency and dimension of stock market and protect the stability of the stock market. State also avoid stock market fluctuates abnormally through rational policy, based on this.

REFERENCES

C. Reaz, M. R. C. Mahdy, T. N. Alam, G. D. Quaderi, M. A. Rahman, Stat. Mech. Appl., 555, 124444 (2020).

M. D. Akhtar, Z. A. Sarwar, K. Shakir, S. A. S. Ali, D. Sara, S. Faizan, J. King Saud Univer. Sci., 34(4), 101940 (2022).

A. Saranya, R. Anandan, Inter. J. Rec. Tech. Eng., 8, 280-283 (2019).

M. Junaid, A. Preeti, K. Ravreet, M. Ajay, G. I. Ali, Pro. Comp. Sci., 218, 1067-1078 (2023).

K. He, Q. Jiang, Acad. J. Comp. Inf. Sci., 5(12), 98-106 (2022).

A. Kurani, P. Doshi, A. Vakharia, M. Shah, Ann. Data Sci., 10, 1-26 (2021).

A. Thakkar, K. Chaudhari, Inf. Fus., 65, 95-107 (2021).

S. K. Chandar, Pat. Rec. Let., 147(3), 124-133 (2021).

C. F. Cao, Z. N. Luo, J. X. Xie, L. Li, Comp. Eng. Appl., 58 (5), 280-286 (2022).

Y. C. Huang, W. W. Fang, Mod. Comp. Sci., 27 (34), 6 (2021).

SCIENCE AND TECHNOLOGY PUBLICATIONS