Stock Price Forecast of Chinese Leading Liquor Stocks Using the Simple Moving Average

Yijia Pan*

Department of Economics and Business Administration Chongqing University Chongqing, China

*Corresponding author: 20190195@cqu.edu.cn

Abstract. Forecasting the trend of price movement of the Chinese leading liquor stocks could produce in significant profits. Traders use technical analysis to forecast the future price of the stock and the simple moving average is among the most important tools of technical analysis. This article aims at applying tow machine learning models (Linear Regression and Long Short-Term Memory) on simple moving averages to see which model performs better, and whether simple moving averages support the forecast by the Root Mean Square Error. Some further rethinks are proposed in the end. The investigation proves that linear regression model is the best one to forecast the tendency of Chinese top three liquor stocks. Instead of assistance, simple moving average firmly express as a hindrance of the prediction. While the selection of time periods of simple moving average presents an unexpected influence on the accuracy of forecast.

Keywords: Simple moving average; price prediction; Chinese leading liquor stocks; machine learning.

1. Introduction

In the modern market economy environment, financial market has attracted more and more attention as an important part of economic development. With the improvement of people's living standards and the increase of property savings after the opening up, the frequency and intensity of investment and people's transaction of financial products have increased a lot. The prosperity and stability of financial market play an increasingly significant role in promoting the rapid development of Chinese economy. As an important part of the capital market, the stock market acts an undoubted character in optimizing the allocation of resources for the country, raising funds for enterprises, and sharing the benefits of economic growth for investors. Charles Dow, founder of The Wall Street Journal, proposed in his book “The Stock Market Barometer” that the operation of the stock market is consistent with the operation of the macro economy, and its cyclical changes reflect and react on the economic changes, which is also a useful "barometer" and “weather observatory” [1].

At present, Chinese financial system is still undeveloped, and the domestic stock market has the characteristics of too frequent trading and high turnover rate. According to statistics, in many foreign mature stock markets, the average annual turnover rate is less than 40%, while Chinese annual turnover rate is more than 600%, which is 15 times of the former [2]. This phenomenon is mainly due to the excessive speculation of Chinese stock investors, who quickly buy and sell stocks when the stock price fluctuates slightly. However, excessive trading increases the cost of trading for retail investors. Because of the volatility of random, investors often cannot really detect the direction of the stock price, leading to greater losses. Therefore, it is of immense value to predict stock trends accurately. The fact that Chinese stock market is not entirely efficient also provides grounds for prediction [3]. Currently, common stock analysis methods mainly include fundamental analysis, technical analysis, and artificial intelligence analysis. However, the fundamental method is not precise enough and is biased towards long-term investors. The conclusion drawn by the technical analysis method can only continue the existed trend but cannot create a new one [2]. Now, the artificial intelligence method using machine learning and deep learning to complete the prediction provides new views for stock prediction.

Machine learning algorithm model can effectively capture the nonlinear relationship of the stock market, which the western countries have been widely used in financial market research, while domestic research on stock price and return prediction through machine learning is not yet mature

and perfect. This is different from the previous cross-boundary prediction method and present a better performance. Data and models provide an objective aspect to reduce the interference of subjective factors for rational investment. The most reasonable and effective prediction model for the stock market is being explored constantly. It effectively remedies the shortcomings of traditional forecast methods in the financial industry [4].

Moving average method is one of typical algorithm models processing time series data [4]. For the typical time series data of Chinese liquor stocks which are high complexity, nonlinear, high noise and share non-stationary, this paper uses the simple moving average of the shares price data as variables to train machine learning model and explore the best length of time periods for simple moving average. Three leading Chinese liquor stocks (Wuliangye, Guizhou Maotai, Yanghe Gufen) will be investigate. Also, in the article, the Long Short-Term Memory, and the linear regression, which are common methods for stock trends prediction, are used as a comparison to construct multiple model evaluation indicators.

Here are 3 points which are focused in this article:

First, between linear regression and Long Short-Term Memory (LSTM), which is better model.

Second, whether using moving average as a feature improves the accuracy of the prediction.

Third, whether it shows a similar conclusion between these three leading Chinese liquor stocks.

As a traditional unique liquor in China, the liquor has become a part of people's social culture. Treating guests with high-grade liquor is regarded as a kind of respect. Mao Tai, the top-notch liquor in China, has gradually become a symbol of status. Therefore, leading liquor stocks attract great attention and own high investment value as well. However, liquor stocks are high-risk stocks. In recent years, although the rapid growth of Chinese liquor manufacturing industry has made the stock prices of leading liquor stocks show an overall upward trend, but in the short term, the price fluctuations are large and obvious, and it is difficult to predict. At present, there are various basic research on the financial status of liquor listed companies, but few research on its changing trend using machine learning model [4]. This paper can make up for this gap.

2. Data and Methods

2.1 Data

In order to get a more accurate forecast of recent years, download the stock price data of these three leading liquor stocks respectively for a total of six years (2016-03-02-2022-03-02) which have 1461 trading days in total. All the data are from “Xinlang Finance” (an official stock information website in China).

Table I. The first five rows of MT database

	Date	Close	High	Low	Open
0	2016/3/2	224.22	225.50	218.67	219.00
1	2016/3/3	227.81	231.98	221.66	225.00
2	2016/3/4	238.778	240.16	225.01	226.88
3	2016/3/7	233.64	240.00	231.05	238.11
4	2016/3/8	229.21	233.30	227.00	233.30

a. the figure is from Xinlang Finance: https://finance.sina.com.cn/

Take MT’s database as an example, there are 5 columns for the dataset. The opening and closing prices of a stock on a given day are represented by the columns Open and Close. The greatest, minimum, and final prices of the share for the day are symbolized by columns High, Low, and Last. All five columns are used as feature in prediction because they are all directly related to the stock price.

2.2 Method

The study of previous market data, especially price and volume, with the purpose of analyzing future price movement is known as technical analysis [5]. Prediction in technical analysis refers to determining the best timing to join or exit the market so that stock purchases and sales can take place. It has been demonstrated that stock exchanges using forecasting signals of technical analysis have much greater stock buy and sale rates than stock exchanges without such signals [6]. The moving averages (MA) is an important instrument in technical analysis. Therefore, it is worthy to try it as a feature in the machine learning to see whether it still works well.

The moving averages is a basic technical indicator that smoothens price data by generating an average price continually updating [7]. They are very versatile and allow you to easily read trends and predict changes in trends. In addition, this provides a different way to view relevant information.

There are many kinds of moving average. Among them, the simple moving average (SMA) is the most used moving average, which sums up the price over a specified time periods and then divides, finding the geometric average [8]. The formula is as follows:

$$SMA = (A_I + A_{II} + \cdots + A_n) / n \quad (1)$$

Where:
A = average in period n
n = number of A

The selection of the time periods of the SMA remains a question. “For instance, Samadi S, et al9. used 10-day and 20-day moving averages to predict stock indicators in Tehran Stock Exchange. Further, Chashami SAN, et al.10 used 30-, 60- and 90-day moving averages in their research. In another study, Sajjadi SH, et al.11 employed 3-, 7- and 10-day moving averages” [6]. The usual durations are 5, 10, 20, 50, 100, 150 and 200 days. Long-term SMA are noted above 100 days, whereas short-term SMA are shown between 5 and 100 days. A distance value is also produced for each combination by subtracting long-term from short-term value [7]. Shorter-term moving averages are seen to yield higher profits according to certain investors, while longer-term moving averages are regarded to offer more accurate indications by academics [6]. Therefore, whether longer-term moving averages or shorter-term moving averages is better for prediction in the article will be discussed in the article.

3. Results and Discussion

It can be roughly divided into 4 steps.

3.1 Extract Data

Guizhou Maotai (MT), Wuliangye (WLY) and Yanghe Gufen (YHGF) all belong to China A-shares that are specifically referred to RMB ordinary shares. These stocks are sued by companies registered in China and listed in China with nominal value indicated in RMB for domestic institutions, organizations, or individuals to subscribe and trade in RMB. A-shares are not physical stocks. They are made of paperless electronic bookkeeping and carry out the "T+1" delivery system with a limit of up or down (10%). The participating investors are institutions or individuals in mainland China.

MT, YHGF and WLY are three largest and most famous liquor stocks in China. Till April 17, 2022, the market caps of these three stocks are $2.27 trillion (MT), $659.4 billion (WLY) and $208.0 billion (YHGF) respectively [10]. It is over 10 years since they issued. All stocks have seen huge gains in recent years, while have experienced violent fluctuations at the same time.

Here is the stock price trend of these stocks. Clearly, in the last five years, the closed stock price of MT has increased by about 437%.

b. the figure is from Xinlang Finance: https://finance.sina.com.cn/

Fig 1. The stock price trend of MT in last 5 years.

Fig 1 shows that the price of WLY has raised by nearly 4 times compared to five years ago.

c. the figure is from Xinlang Finance: https://finance.sina.com.cn/

Fig 2. The stock price trend of WLY in last 5 years.

Fig 2 illustrates that the price of YHGF has also grown by 158% than five years ago.

d. the figure is from Xinlang Finance: https://finance.sina.com.cn/.

Fig 3. The stock price trend of YHGF in last 5 years.

Fig 3 provides firm evidence that the price of YHGF has also grown by 158% than five years ago.

It is this huge growth on price and enormous market caps of these three stocks that is worthy studying to find which model can better predict the trend.

3.2 Data Preparation

After the check of missing data of these three databases, it is good to see there is no missing one, and all of them have 1461 rows which is big enough to run the models on them.

From Fig.1 to Fig 3, stock prices of MT, WLY and YHGF all fluctuate a lot in last 6 years. To avoid the effect of the noise and overfitting, normalization for all the data is needed. After the treatment of the minmax scaler, all the data is in the range from -1 to 1. After this, both datasets are divided into training set and test set in a 7:3 ratio.

3.3 Add SMA

According to Stock Market Value Prediction Using Neural Networks by Naeini, M. P., Taremian, H., & Hashemi, H. B. [10], these three datasets are copied, and calculate the SMA to all the four original features (close, high, low, open) to create SMA features in duplicate datasets. These datasets with SMA features will serve as control groups to test the effect of SMA on the overall prediction. To adjust the length of time period of SMA in a convenient way, the sliding window is also set, which can be really helpful to compare the accuracy of the forecast between the shorter and longer term SMA.

3.4 Set and Operate Models

Before the model training, the adaptive moment estimation (Adam) is used to adjust the LSTM model to generate optimal parameters in a default learning rate.

Then the forecasts of Linear Regression and LSTM for MT, WLY and YHGF are tested respectively. Also, the sliding window is adjusted to find the best-performance periods of SMA. After several attempt, the performance of the two models with the SMA and the performance without SMA are compared to show the influence of SMA.

3.5 Results

The forecasts of Linear Regression and LSTM for MT, WLY and YHGF are tested respectively. Also, the sliding window is adjusted to find the best-performance periods of SMA. As the usual range of periods of SMA is between 10 and 200 days, 10 days, 30 days, 50 days, 100 days and 200 days are used as different sizes of window of moving average to investigate whether the longer-term moving averages or shorter-term moving averages better predict the stock prices. After several attempt, it comes that in MT and YHGF the shortest term and the longest-term lead to an unprecise forecast, when the period comes to 100 days (a middle-length term), the closest distance between the prediction curve and the actual one appears. But for WLY, the result become completely opposite. When the period comes to 100 days, the most unprecise prediction appears. Therefore, for the SMA of different terms, there are two inverse results on different Chinese leading liquor stocks. For MT and YHGF, the middle-term SMA are most accurate, while it is the most inaccurate for YHGF.

After this, SAM 10 is chosen to display the performance of the two models with and without the influence of SMA.

Below are 3 figures showing the prediction visually:

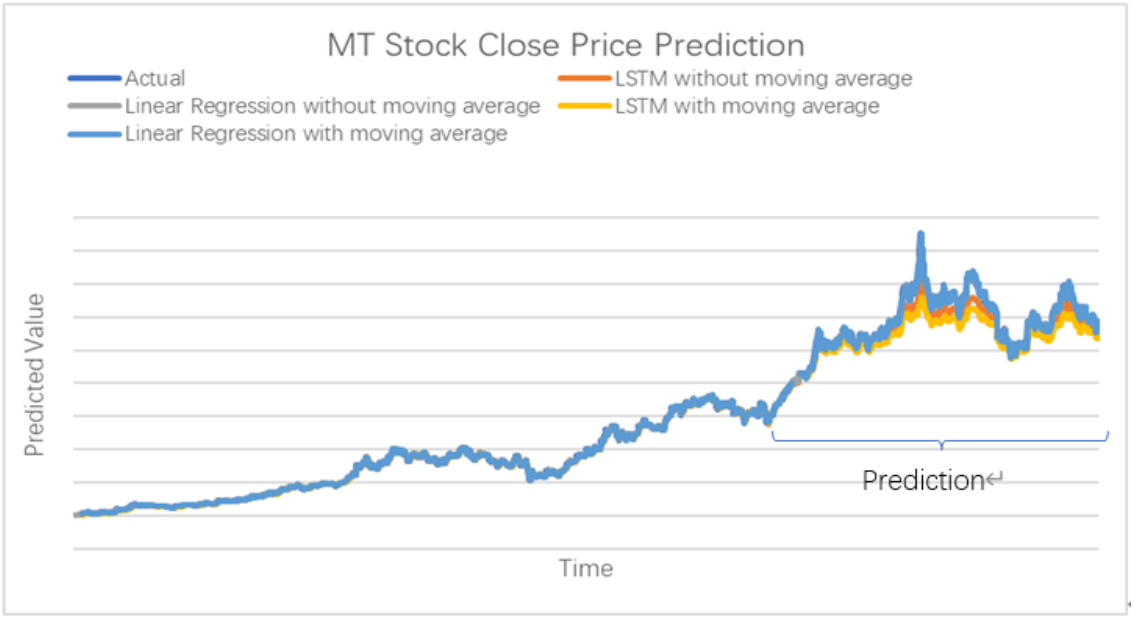

Fig 4. Stock close price prediction of MT.

In Fig 4, the prediction of MT demonstrates in 5 lines, which represents four prediction results and an actual data. For datasets of stock prices are divided into training set and test set in a 7:3 ratio, the prediction lines only occupy the lower half of the real curve. Though they are very close to each other, there is still a short distance from the real value to the curves standing for LSTM prediction results while the linear regression curves are too close to the actual curve to observe. And the curve of LSTM without moving average is a bit closer to actual curve than the curve of LSTM with moving average. Therefore, it seems that the performance of LSTM is worse than the linear regression. Only for LSTM, the prediction without moving average is better. But there is little information for linear regression results.

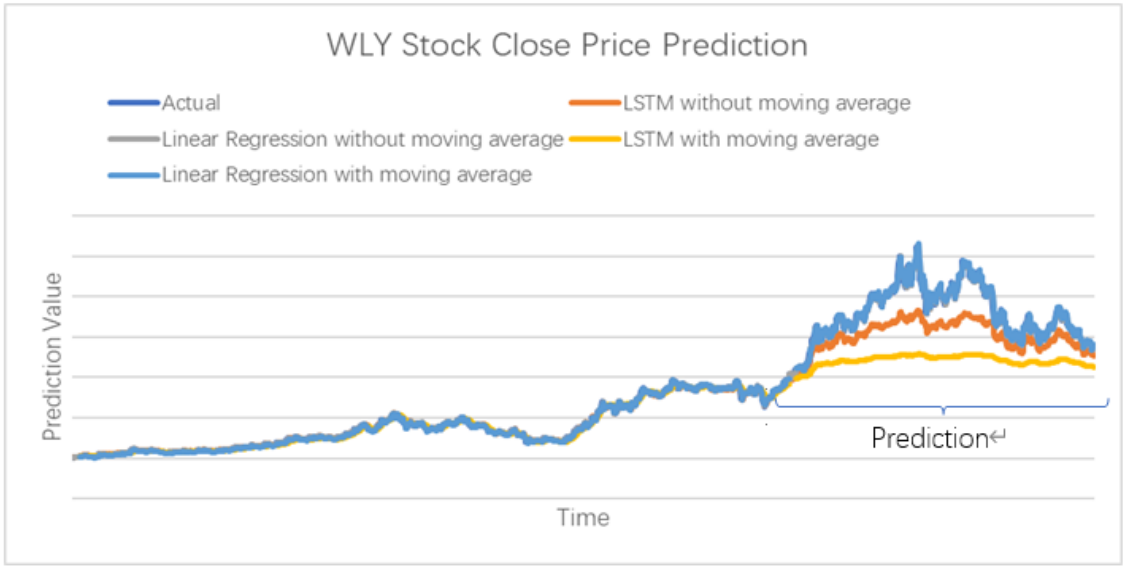

Fig 5. Stock close price prediction of WLY.

In Fig 5, same to the Fig 4, the prediction of WLY displays in 5 lines, representing four prediction results and an actual data. For WLY, the same results as in Fig 4 are shown in a more clearly (the distance between curves getting larger): there is little information for linear regression results, the performance of LSTM is worse than the linear regression and only for LSTM the prediction without moving average is better.

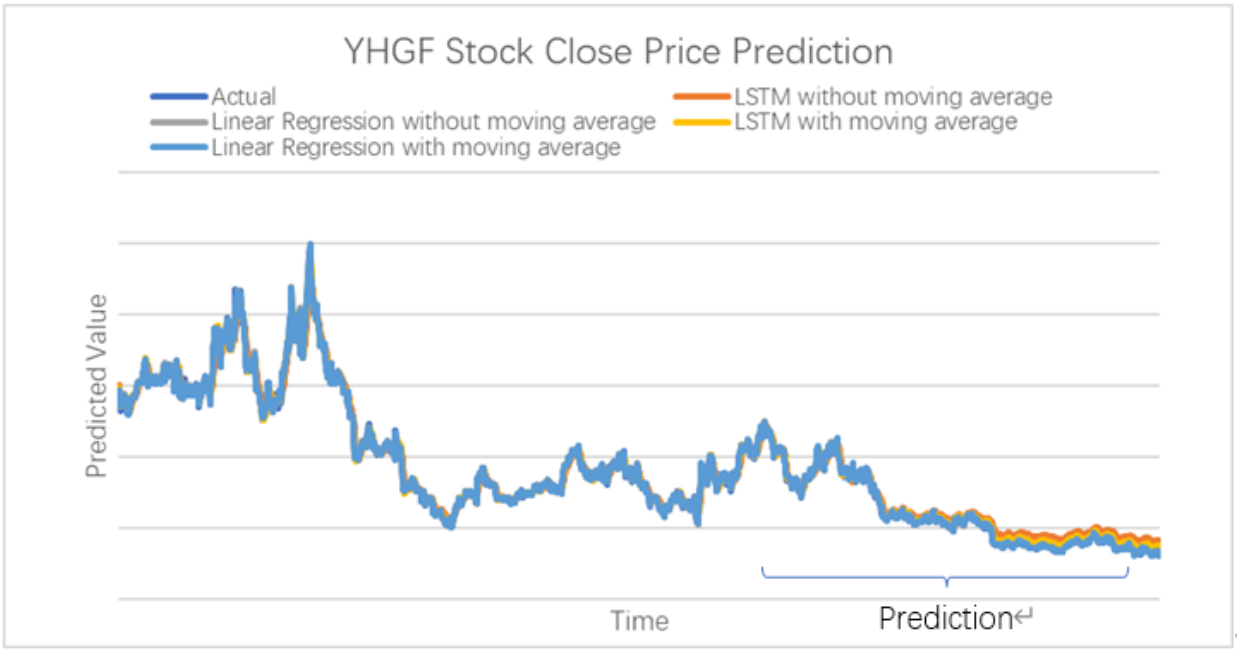

Fig 6. Stock close price prediction of YHGF.

In Fig 6, there shows a difference compared to Fig 4 and 5, the prices have decreased during the period. However, it still conveys the same results as the previous 2 figures. For further results, mathematical measurable indicator is strongly needed.

Root mean squared error (RMSE) considered an excellent general-purpose error metric for numerical predictions. Lower the value of RMSE, more accurate the forecast is. Use RMSE to evaluate the forecast outcome, and the clear outcome is as follows:

Table II. RMSE results of MT.

RMSE on test dataset	LSTM	Linear Regression
With SMA	0.242276	0.076052
Without SMA	0.153752	0.074522

From table II, RMSE results of MT are illustrated. The most precise model is linear regression model without SMA which developed with a RMSE of only 0.0745 on the MT stock, while the most inaccurate one is LSTM model without SMA which has a RMSE of 0.153. In addition, the RMSE of models with SMA are all higher than that without SMA. Especially for LSTM, the RMSE value of LSTM without SMA is nearly a half of that with SMA. Meanwhile, though the RMSE values of linear regression with and without SMA are close which respectively are 0.076 and 0.075, the result of linear regression indicates a higher accuracy then results of LSTM.

Table III. RMSE results of wly.

RMSE on test dataset	LSTM	Linear Regression
With SMA	1.23994377	0.116236582
Without SMA	0.58434798	0.110386614

From table III, RMSE results of WLY are illustrated. The basic data presentation is exactly alike that of MT. The linear regression model without SMA is the most precise model, while the LSTM model without SMA is the most unprecise one. The RMSE of models with SMA are all higher than that without SMA, and that of linear regression are lower than LSTM. The differences between data are even greater, so the result also becomes more obvious and convictive.

Table IV. RMSE results of HYGF.

RMSE on test dataset	LSTM	Linear Regression
With SMA	0.02357	0.00004151
Without SMA	0.021809	0.00003814

From table IV, RMSE results of YHGF are illustrated. The basic data presentation is the same as the MT and WLY. The only caveat here is that the linear regression with and without SMA displays extraordinarily well on YHGF, the RMSE of it is less than one in ten thousand, respectively are 0.000042 and 0.000038.

3.6 Discussion

This article proposes the use of machine learning on SMA in Chinese leading liquor stocks. This identifies the inaccuracy of moving averages as a feature and recognize a linear regression is a better model for prediction of Chinese leading liquor stocks. It is noteworthy that for the SMA of different terms, there are two inverse results on different Chinese leading liquor stocks. For MT and YHGF, the middle-term SMA are most accurate, while it is the most inaccurate for YHGF.

The outcome shows three points: first, no matter whether include the SMA features, linear regression has a much more accurate prediction than LSTM model; second, though the moving average is one of the most important tools of technical analysis, it only makes the prediction more inaccurate as a feature in these two machine learning models; third, MT, WLY and YHGF show the alike prediction results.

While being a trend indicator, the moving average is also a lagging indicator. A lagging indication is a financial sign that appears only after a significant price change has occurred. And this might lead to the phenomenon that SMA doesn't work well in machine learning models. For linear regression model and LSTM model, maybe a longer time prediction and a larger size of dataset will give an opportunity for LSTM to show its power.

4. Conclusion

At present, Chinese financial system is still undeveloped, and the domestic stock market has the characteristics of too frequent trading and high turnover rate. Many factors interact in the stock market including political events, general economic conditions, and traders' expectations. Therefore, accurate forecast of the stock price is extraordinarily challenging and is of immense value.

As a symbol of status, the high-grade liquor plays an important role in people's work and social life. Leading liquor stocks own great popularity as well as high investment value in China. But the high-risk stocks fluctuate a lot, resulting in difficulties to predict. This article investigates three leading Chinese liquor stocks (MT, WLY, YHGF) with a total of six years' data (2016-03-02-2022-03-02), which has 1461 trading days respectively. The simple moving average of the shares price data are used as variables to train machine learning model and explore the best intervals' length of the simple moving average. LSTM, and linear regression, which are common methods for stock trends prediction, are used as a comparison to construct multiple model evaluation indicators. The accuracy of forecast between two machine learning models, the accuracy between results with or without MA and the similarity of the performance of these three leading liquor stocks have been emphatically investigated on the article.

Experimental results show that moving averages are barriers to prediction as features of machine learning models. In this research, linear regression unexpectedly has a much more accurate prediction than LSTM model. Though performance of the prediction results is similar between three stocks, they demonstrate 2 completely opposite influence of SMA duration selection on the accuracy of prediction results.

This article enriches the domestic research on liquor stock prediction by machine learning, and innovatively introduces stock price data processed by SMA as variable to train model. In the future

study, it can be investigated the prediction results between leading liquor stocks and normal liquor stocks or the rule behind the influence of SMA duration selection on the accuracy of prediction results of Chinese liquor stocks.

References

[1] Ding Shujin (2010). An empirical study on the "barometer" effect of China's stock market.Shandong University.

[2] Li Zhongchen (2020). Research on Stock Trend Prediction based on machine learning.

[3] Luo Chaoyu, & Yang Qifan (2022). Research on the effectiveness of Chinese stock market. Finance,12(2),195-201.

[4] Yang Yi (2021). Research on liquor stock prediction based on machine learning algorithm model.Shandong University.

[5] Ihab A. El-Khodary (2009). A Decision Support System for Technical Analysis of FinancialMarkets Based on the Moving Average Crossover. World Applied Sciences Journal 6 (11).

[6] Khorram, M., & Sheshmani, M. (2015). What period of time is better to use for moving average to predict stock price in Tehran stock exchange. Indian Journal of Science and Technology, 8(27), 1-8.

[7] Dinesh, S., Rao, N., Anusha, S. P., & Samhitha, R. (2021, April). Prediction of Trends in Stock Market using Moving Averages and Machine Learning. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[8] Kirkpatrick II, F. C. D., & Julie, R. (2017). Moving averages. CMT Level II 2017: Theory and Analysis.

[9] Xinlang Finance: https://finance.sina.com.cn/

[10] Naeini, M. P., Taremian, H., & Hashemi, H. B. (2010, October). Stock market value prediction using neural networks. In 2010 international conference on computer information systems and industrial management applications (CISIM) (pp. 132-136). IEEE.