

# Stock price prediction based on SVM, LSTM, ARIMA

Haoyu Ji\*

Department of science, McMaster University, Hamilton, Canada

\*Corresponding author: jih20@mcmaster.ca

**Abstract.** In general, forecasting on stock prices is a famous and interesting area that gathers many researchers in. Contemporarily, after the birth of AI, the number of the algorithms used in the prediction of equity market fluctuation are boomed rapidly. Applying the combination of statistics and algorithms can help researchers as well as investors learn about either short-term regulation (such as opening price) or the long-term market movement. This paper discusses three kinds of models which are used to predict the stock price for long or short term. Specifically, some empirical results are presented to prove the feasibility and significance of the models. By analyzing techniques used to predict stock prices and the limitation of these models, the discussion about the challenges and the outlook posed from the scope of future work in this filed are also shown and demonstrated. These results shed light on guiding further exploration of price forecasting for different kinds of underlying assets as well as portfolios.

**Keywords:** SVM; LSTM; ARIMA; Machine Learning.

## 1. Introduction

Stock price prediction is the attempt to predict the stock value and market sectors. Stock price movement is a concerning filed, for a huge amount of money invested in the stock market and many family's happiness bears directly upon the fluctration of this market. Many researches on stock price forecasting have been made every day, based on different data sources and many built-in models, such as newspaper, instagram, Baidu and Wikipedia data. All these external factors, combined with stock prices and stock technical indicators, show their influence on stock price changes. In today's society, how to calculate the accuracy of stock prices is a open question.

The factors that affect stock prices come from many aspects, including finicial factors, big political events, the recent behaviour of the company, etc. Whatever the factor is, the impact od the stock prices can be sum up as two shapes: linear or nonlinear. The complex relationship between stock prices and multiple factors arouses great interest for researchers to create responsible models to make prediction. One of the methods for forecasting stock prices is the big data method, which aims to gain insight from a large number of public data [1]. The meaning of the Machine learning is to solve the problem by using given data. The support vector machine (SVM) will be discussed in this study [2]. The main theory of deep learning method is to calculate based on neural network [3]. LSTM is a special kind of RNN, which is used to solve dependence of long-term [4]. Therefore, it is not easy to develop a consistent and accurate method. With these memory cells, the network can better link the memory with the time input remotely, which is why it is worth knowing that strong data can be formed over time, with great prediction ability. Changing with time is an important feature of stock price which makes researchers easily use time series approach to analyse this time-dependent data. Among various approaches in time series data, the ARIMA proposed in Ref. [5] is the most widely used one (e.g., in Ref. [6]).

In order to sum up the results of different kinds of models used in stock price prediction and analyse the limitation of each model, this paper analyses the representation of machine learning, deep learning and time series. Section 2 describes statistical results of SVM. Details of LSTM are shown in Section 3. ARIMA models are also presented by equations in Section 4. Section 5 is the limitation of stock prediction.

## 2. SVM

Support vector machine (SVM) is a way to classify, whose concept is to enlarge the interval to maximization. SVR is one of the most well performance type of it. In the calculation, a convex optimization problem with constraints replace th former peoblem. Result from introducing Lagrange multiplier method and dual learning method to simplify this problem, the constrained optimization problem is transformed into a Lagrangian multiplier optimization problem. In the whole procession, because of using dual learning, it is naturally to introduce kernel method. Kernel method can be used to map to high-dimensional space and solve nonlinear classification. Directly using the inner product function of space can effectively help dimensional space complexity. There are several normal kernel functions, such as Exponential Kernel, ANOVA Kernel, Multiquadric Kernel, etc. In this saction, a research made by Saahil is introduced. He chooses the radial kernel to solve a SVM predicting problem [7]. This function shows as

$$K(x_i, x_k) = \exp \left( -\frac{1}{\delta^2} \sum_{j=1}^n (x_{ij} - x_{kj})^2 \right) \quad (1)$$

Here,  $\delta$  is the bandwidth of the kernel function [8], which it can deal with different types of input sets [9]. Chen uses SVM model to analyse 4 stocks: AAPL, MAST, FORD, EXON [10]. An SVR algorithm is used on each stock dataset with a lookback window of 30 days. All variables will be included as inputs of the model. The results are shown in Fig. 1. 75% of each dataset will be trained by SVR model. The Fig. 4 show the prediction results. The author uses MAPE to evaluate the performance.

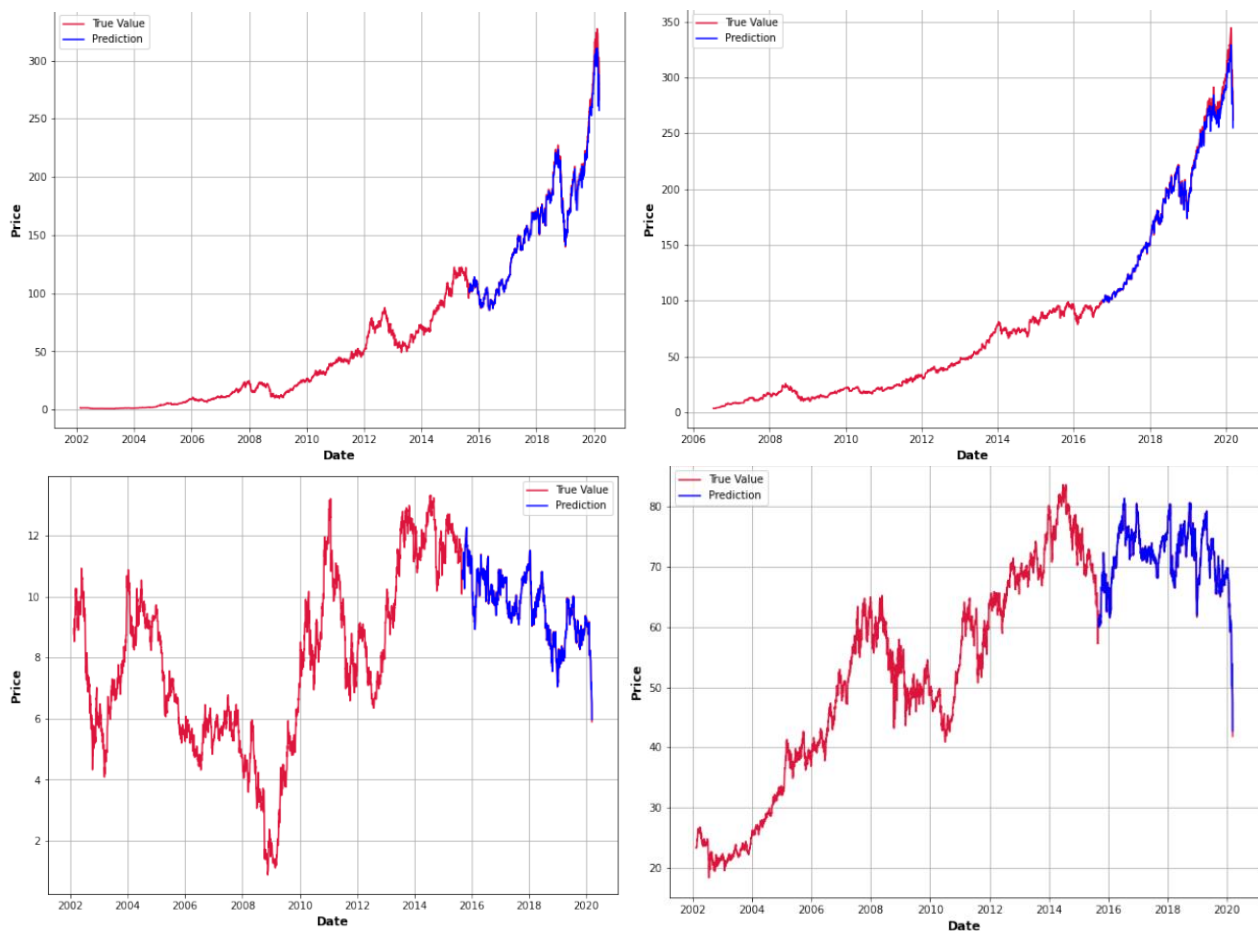


Figure 1. SVR used on AAPL, MAST

To investigate the accuracy of successfully using an SVR model which is trained on a given stock to get predictions on another stock, the author uses the trained SVR output based on Apple dataset and then apply them to forecast prices [10]. Analysis results are shown in Fig. 2. It is observed that the accuracy of an SVR model trained on the Apple stock is useful to make predictions about other stock prices. The performance of the model is judged by the MAPE, which shows good performance as reported in Ref. [10].

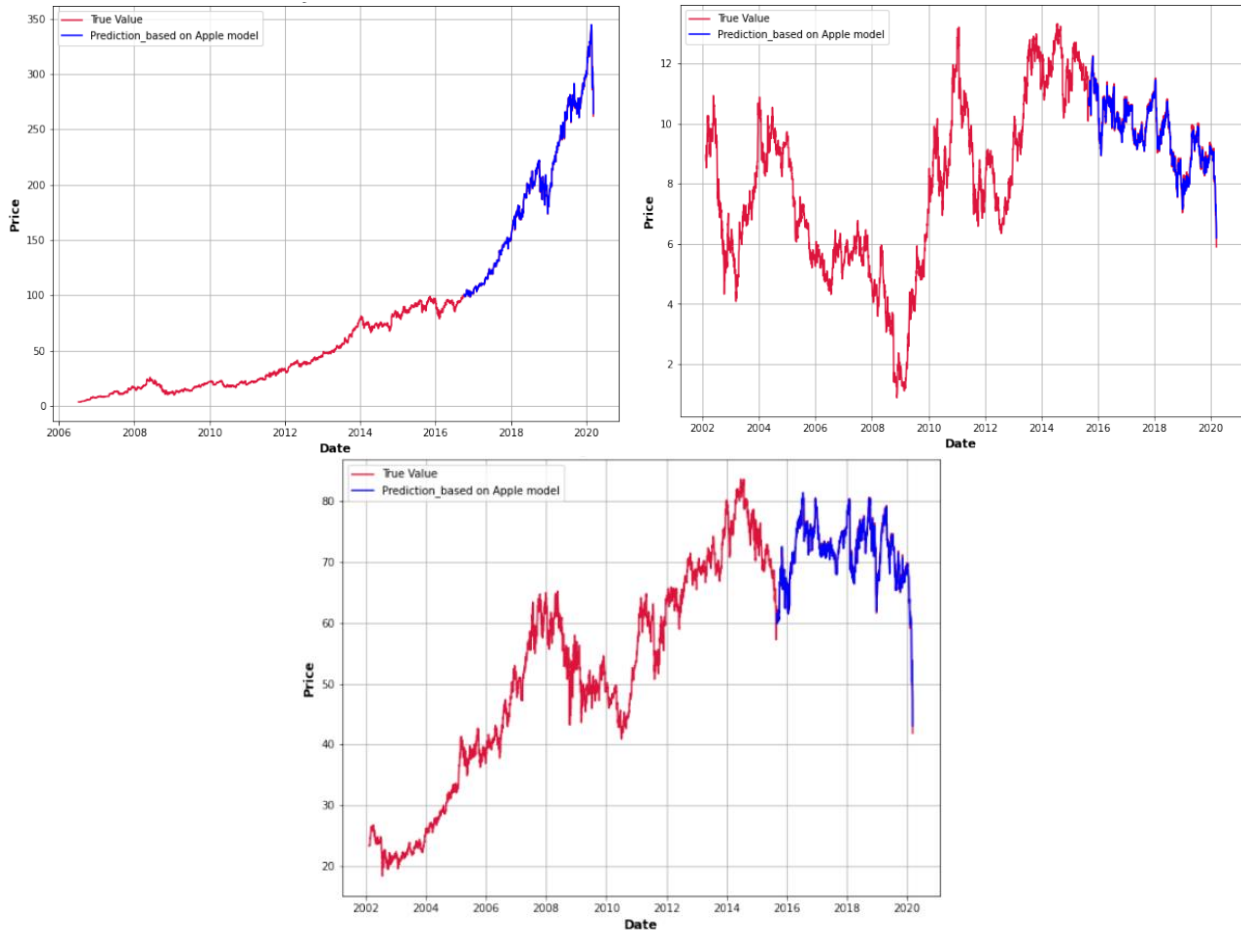


Figure 2. Trained SVR used on MAST, FORD and EXON.

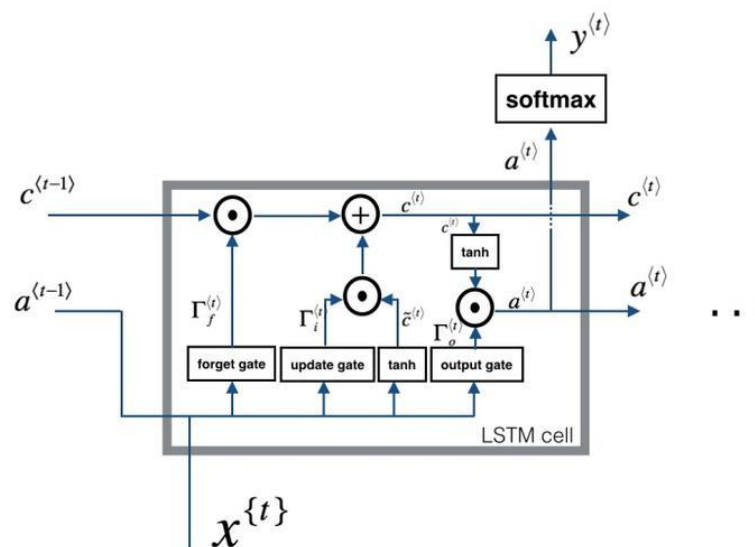


Figure 3. A sketch of the LSTM.

### 3. LSTM

In 21<sup>st</sup> century, ANN model has become a popular method for the stock price prediction and many researches are based on it. With the analyzation of Tokyo stock exchange, Kimoto et al. [11] used the ANN to predict the stock prices. To overcome the gradient vanishing problem, LSTM, a variation of RNN, is designed as sketched in Fig. 3. Chen et al. choose LSTM-based method on prediction of the Chinese stock return [12]. The results shows that the accuracy has the same movement with input data volume. The prediction of the price for AAPL is shown in Fig. 4. To investigate whether the accuracy of using a model trained on one stock to make predictions on another stocks, the author uses LSTM to train Apple dataset and then predict the other three stock prices, MAST, FORD, EXON. As shown in Fig. 5. As a result, when using a trained model based on the Apple dataset, it's accurate, reflected on the evaluation of the performance reflected on the MAPE.

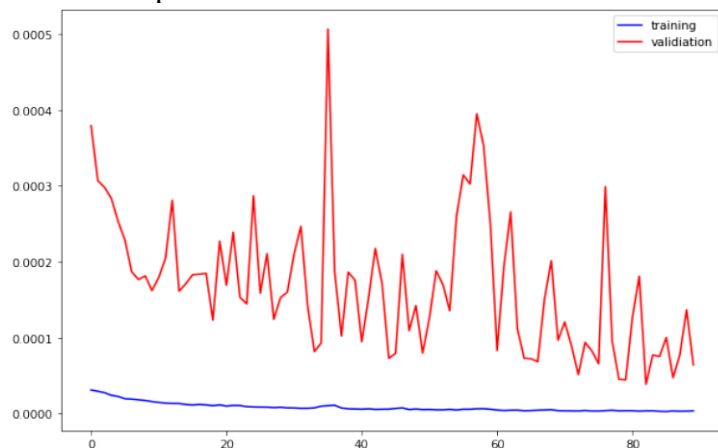


Figure 4. Convergence rate of LSTM for AAPL dataset

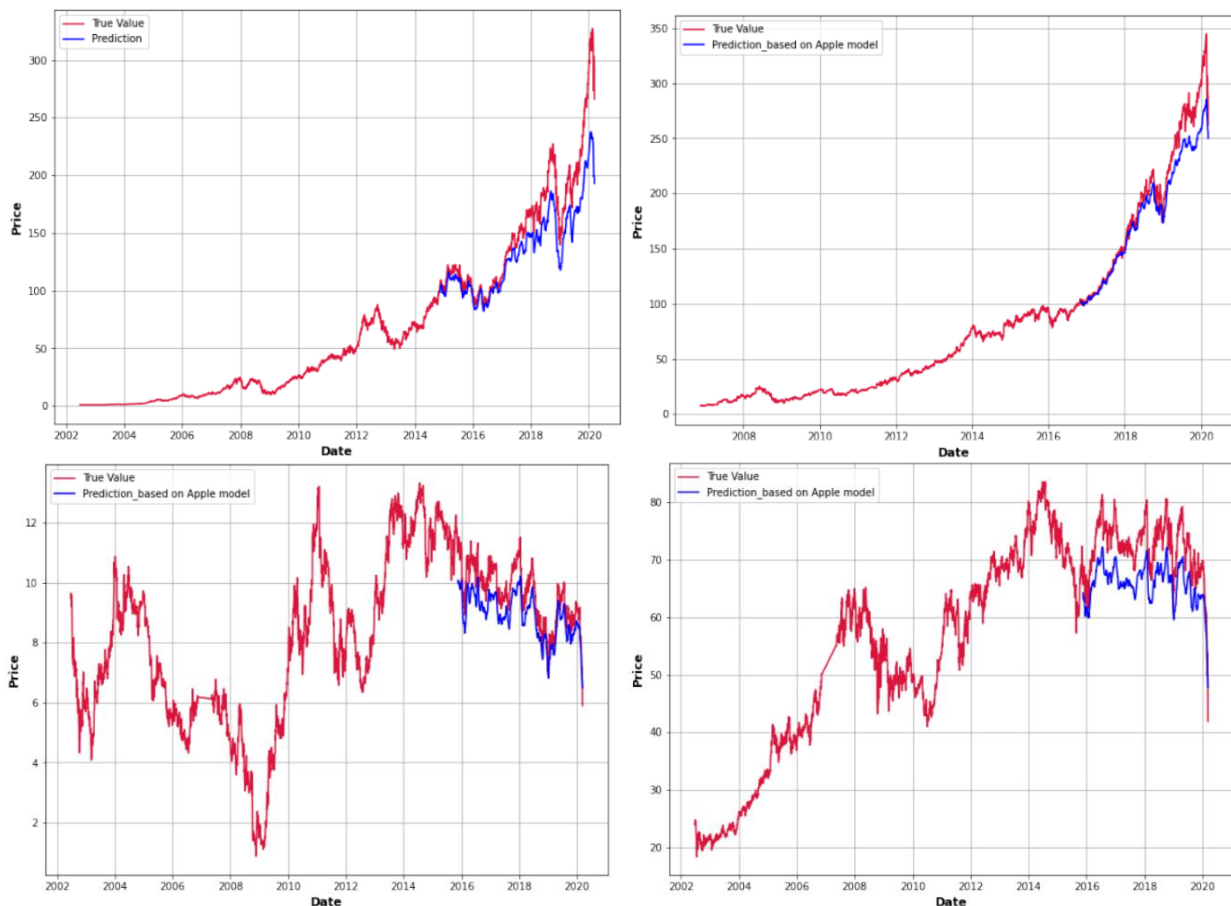


Figure 5. LSTM used on AAPL, MAST, FORD and EXON.

#### 4. ARIMA

Prediction is always an exciting area of experimentation. Improving existing predictive models to help investors to create effective arrangement and strategy in the stock market is a strong temptation for researchers. Generally, the prediction is done from 2 perspectives: statistical and AI techniques. ARIMA models belong to the statistical models. ARIMA models are strong and efficient in short-run prediction than even the fore most wide spread ANNs techniques. It's been widely utilized in the area of economics and finance. An ARIMA model is a vibrant univariate forecasting method to project the future values of a time-series. The equation of ARIMA(p,d,q) is presented as follows:

$$(1 - \sum_{i=1}^p \varphi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t \tag{2}$$

Here, L is lag operator,  $d \in Z, d > 0$ . A. Adebisi uses ARIMA model to predict NOKIA stock price [13]. The data composed by open price, low price, high price and close price respectively. According to the prediction results of a specific example given in Fig. 6, the performance is almost perfect.

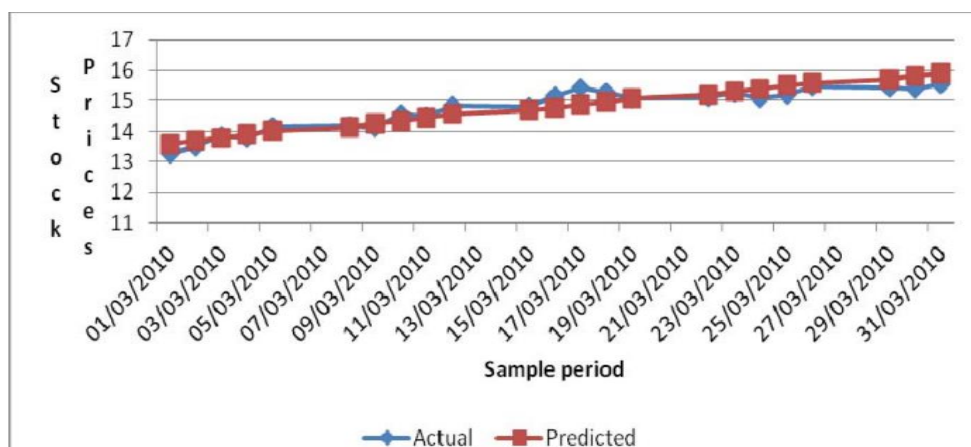


Figure 6. Actual Stock Price vs Predicted values of Nokia Stock Index.

#### 5. Limitations and Future outlook

There are several limitation of SVM. The main limitation is lack of effective technologies and methods to deal with non quantitative factors. The factors that affect stock price fluctuation are very complex. In addition to quantitative factors such as fundamentals and technical indicators, it is also affected by non quantitative factors such as policies, psychological fluctuations and international emergencies. At present, most of our stock price volatility prediction models are pure quantitative models, i.e., one only brings quantitative information into the model for calculation, and then one can get the prediction results. However, the non quantitative information that has a great impact on the volatility of stock returns cannot be added to the quantitative model due to the limitations of the model's own processing methods and processing technologies, so it cannot achieve accurate prediction that is very consistent with the actual situation. Besides, the market fluctuations made by those companies which do not quote should not be ignored, for example, though valve, an American gaming company doesn't quote, it still makes strong influence to other gaming stocks.

There are three disadvantages should be concerned while using the LSTM model, the first is that if the data is sequence related, it is not easy to use the time series data that is not formed completely according to the time. The second is that it doesn't support parallel operation. The third is that when the sequence length exceeds a certain limit, the gradient will disappear.

The limitation of ARIMA is that essentially ARIMA is a time series algorithm, time series data is required to be stationary or stable after differentiation, in this area, ARCH has advantages on ARIMA. In addition, only linear relationship can be captured, but not nonlinear relationship.

Looking at the development process of stock price forecasting models, it is not difficult to see that this process is a process in which new forecasting theories and new forecasting technologies are constantly introduced, absorbed, digested, applied, and improved. In the future, exploring of training algorithms supporting vector machines will be an accessible direction to develop, for example, integration with data pre-processing (sample importance, attribute importance, feature selection, etc.). The fluctuation of the stock price reflects the emotional changing, and the model can be combined more tightly with factors about emotion in the future. In addition, the models used on stock price prediction can be also used in other areas, such as forecasting the price of commodities such as gold.

## 6. Conclusion

In summary, this paper discusses stock price prediction using three different models: SVM, LSTM, ARIMA. The SVM model is one of the machine learning methods, it shows high accuracy in prediction of four stock prices in short term. When talking about long term, LSTM has advantage than SVM model, LSTM is one of deep learning model. The ARIMA model has the lowest MAPE, but only for linear dataset, which is caused by its essential that it is a time series model. In the future, more indicators from many areas will be used to construct feature factors. Thus, more models will also be used to create investment strategies, such as GARCH, multi-factor regression model. Based on these models, researchers can gain higher return effectively. Stock market is an important part of the world economy, the researches using different models will reinforce motivation of economic growth.

## References

- [1] Jose J, Mana S, Samhitha B K. An efficient system to predict and analyze stock data using Hadoop techniques. *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, 8 (2): 2277 - 3878.
- [2] Choudhry R, Garg K. A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 2008, 39 (3): 315 - 318.
- [3] Hu Z, Zhao Y, Khushi M. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 2021, 4 (1): 9.
- [4] Obthong M, Tantisantiwong N, Jeamwathanachai W, et al. A survey on machine learning for stock price prediction: algorithms and technique. *Online Prediction*, 2020.
- [5] Box G, Jenkins G M, Reinsel G C, Ljung G M. *Time series analysis: forecasting and control*. New York: Wiley, 2015.
- [6] Chen L P, Tian W T, Yeh H C. A study of the birth rate in Taiwan. *Journal of Data Analysis*. 2015, 10(6): 141- 190.
- [7] Madge S, Bhatt S. Predicting stock price direction using support vector machines. *Independent work report spring*, 2015, 45.
- [8] Kjae K. Financial time series forecasting using support vector machines. *Neurocomputing*, 2003, 55, 1.
- [9] Buhmann M. *Radial basis function*. Mathematisches Institut, Justus-Liebig-Universität Giessen, 2010.
- [10] Chen L P. Using machine learning algorithms on prediction of stock price. *Journal of Modeling and Optimization*, 2020, 12 (2): 84 - 99.
- [11] Kimoto T, Asakawa K, Yoda M, et al. Stock market prediction system with modular neural networks. 1990 IJCNN international joint conference on neural networks. IEEE, 1990: 1 - 6.
- [12] Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: A case study of China stock market. 2015 IEEE international conference on big data (big data). IEEE, 2015: 2823 - 2824.
- [13] 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.