Research on the Stock Price Prediction Using Machine Learning

Yang Shi[1,a,*]

[1]*Weatherhead School of Management, Case Western Reserve University, Cleveland OH 44106
United States*
a. yxs879@case.edu
corresponding author

Abstract: Stock price prediction is a complex and challenging problem that has attracted the attention of investors and researchers for decades. In recent years, machine learning algorithms have become powerful tools for predicting stock prices. This paper first introduces four popular machine learning algorithms used for stock price prediction which are linear regression, support vector machines, artificial neural networks and long short-term memory. In addition, applications and potential challenges of stock price prediction using machine learning are examined. Overall, this paper provides a comprehensive overview of ML-based models for stock price prediction and highlights the potential benefits and limitations of these models for financial researchers and artificial intelligence developers.

Keywords: stock price prediction, machine learning algorithms, linear regression, long short-term memory, model performance evaluation

1.   Introduction

The stock market fluctuates dramatically due to factors that are both within and beyond our control, which makes it a complicated and chaotic system. To an investor, predicting the future prices of stocks is one of the most important tasks that can assist in making sound investment decisions. As opposed to traditional technical and fundamental analysis methods for predicting stock prices, machine learning algorithms such as linear regression, autoregressive integrated moving average model and random forest can analyze massive amounts of historical data, identify market trends, and predict stock prices [1]. However, using machine learning to predict stock prices precisely and consistently requires strong financial and artificial intelligence knowledge basis and thorough understandings of various type of machine learning algorithms. In order to save investors and researchers time, four most popular and commonly used models are introduced in section 2 of this work. According to the majority of researchers, they all have significantly improved accuracy and efficiency of stock price prediction [2]. In section 3, this paper provides readers with a brief view of the main fields where these models can be applied. In section 4, some possible limitations of machine learning are given and improvements on further research are discussed for the purpose of advancing the field and improving the accuracy of stock price prediction.

2.   Popular Machine Learning Algorithms for Stock Price Prediction

Machine learning algorithm is a method of using mathematical rules and procedures to learn patterns and relationships from data. It's also a computer program that can enhance its performance on a specific task automatically by learning from different example data. Recently, machine learning has been researched and applied in different study fields. Financial researchers use machine learning algorithms to analyze historical data and predict stock market trends. Below are two recent studies on stock price prediction.

M Umer Ghani, M Awais and Muhammad Muzammul implemented Linear regression and predicted one-month stock market trend using data from Yahoo Finance (AMZN, APPL and GOOGL) in 2019 [3].

An algorithm model combined random forest regression and LSTM was chosen to predict the price of Bitcoin by Junwei Chen [4].

In this section, four machine learning algorithms which are Linear regression (LR) , Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Long Short-term Memory (LSTM) will be presented. They are commonly used by researchers for stock market analysis. During analysis, different factors such as the complexity of historical financial data, model performance, prediction time horizon and economic conditions need to be considered to ensure the accuracy and reliability of stock prediction. Therefore, it is very important to make detailed research and understand stock data thoroughly before selecting ML algorithms. And there is no guarantee that ML algorithm provides accurate predictions in all cases. Specific challenges and difficulties will be discussed in section 4.

2.1.   Linear Regression (LR)

Linear regression is a technical and quantitative measure for financial market analysis that examines two variables to determine an underlying relationship between them. The output variable in stock price prediction is the stock price, and the input variables vary depending on the research topics chosen. Linear regression can be used to forecast future stock prices by analyzing the relationship between these variables and the stock price.

Stock prices are predicted in regression by the below linear function:

P = a + bX + error

P = stock price

X = independent value

a,b = coefficients

One of the 2021 studies indicates that linear regression for stock market analysis is more accurate than the Support Vector Machine (SVM) algorithm [5]. To increase data accuracy and time efficiency, financial researchers usually use data analytics tools such as Excel and Python to apply the LR algorithm to stock price prediction. For example, Python was used by researchers C. C. Emioma and S. O. Edeki to train and validate stock price data for the purpose of estimating the closing price using the least-squares linear regression algorithm [6].

Four main steps on how to use linear regression algorithm to predict stock prices are suggested as below.

First, collect historical stock price data, including open, high, low, and close prices for a certain period, and then clean the data, which may include removing missing values or classifying the data set to increase accuracy.

Figure 1: Example chart for OHLC.

See figure 1 for an example of OHLC chat by Investopedia [7]. An OHLC chat shows the open, high low and close price for a given period. Vertical lines represent high and low prices for the period, the left line represents the open price, and the right line represents the close price. The entire structure is called a bar.

Secondly, select input variables that have impacted stock price during chosen periods and sort out only numerical variables due to the limitation of Linear Regression model.

Thirdly, separate the dataset into a training and testing set and use training set to fit LR algorithm to the data.

Finally, calculate data metrics such as mean squared error (MSE) or R-squared to evaluate the LR algorithm model performance on the testing set. After that, the model can be used to forecast stock prices and adjusted if needed.

In general, the LR algorithm has high performance for linearly separable data, which limits its ability to analyze complicated and nonlinear variables.

2.2. Support Vector Machines (SVM)

As a supervised ML binary classification algorithm, the Support Vector Machine (SVM) can be used for analyzing both linear and nonlinear stock data, classifying stock peaks and troughs as well as predicting future stock prices. Using SVM, two types of dimensional points are divided into two groups by an N-1 dimensional hyperplane. The optimized hyperplane is the one that has the maximum distance from both classes.

There are various SVM Kernel functions that can help change data dimension and allow researchers to transform linearly inseparable data into a separate dataset. The most popular and widely used function is Radial Basis Function (RBF). It can create the most complex and curve-fitting dataset and identify nonlinear relationships between stock prices and input variables.

Several research studies show that SVM can be powerful and effective for stock predictions. According to the analysis done by Jingdong Yang 2023, Kernel functions used in SVM provide high accuracy in price forecasting and ensure a global minimum analysis result [8]. Another study

back in 2006 indicates that SVM outperformed other ML algorithms such as random forest, neural network and traditional models [9].

2.3. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) or Neural Networks (NN) are computer systems inspired by biological neural networks. It consists of artificial neurons that have inputs and generate an output connected with other neurons in multiple layers. In the past couple of years, ANN has been used increasingly in stock forecasts. It helps detect nonlinear patterns that other technical methods are unable to reach.

Various types of ANNs have been discovered and applied in prediction such as Recurrent Neural Networks (RNN), Convolutional Neural Network (CNN) and other neural networks. In 2016, several ANN methods have been applied and compared in price movement forecast by L Di Persio and O Honchar. Results show that CNN performed better on modeling financial time series [10]. Another relevant work done by Milad Shahvaroughi Farahani and Seyed Hossein Razavi Hajiagha shows that ANN can predict stock price more accurately due to nonlinearity and asymmetric qualification of data, and hybrid model performs better with low error [11].

2.4. Long Short-Term Memory (LSTM)

Many financial experts consider Long Short-term Memory (LSTM) as one of the most promising ML Algorithms. The ability of processing entire sequences of data makes LSTM stand out in stock price prediction. LSTM is a type of Recurrent Neural Networks that can efficiently draw out useful information from long period of time series data which assists in presenting long term price movements.

A method called Gated Recurrent Units (GRUs) is used in LSTM to learn long term dependencies. By using GRUs, information from former timestamps can be selectively forgotten and data from present timestep can be passed on.

Numerous studies have been conducted by financial researchers in the past a few years to test the performance of LSTM in stock price prediction. Below three ones are selected to demonstrate high accuracy and effectiveness of LSTM.

JT Firouzjaee and P Khaliliyan used RNNs and LSTM to predict the stock prices of oil companies and found LSTM presents a reasonably accurate prediction and interprets system's surprising well [12].

Mohammad Diqi used LSTM to predict stock price of the Indofood Sukses Makmur Tbk (Indonesia Listed Company in IHSG) and found LSTM can have good accuracy with a small error rate and can be an ideal tool to deal with dynamic prices [13].

Md. Ebtidaul Karim, Md. Foysal and Sunanda Das used Bi-LSTM and GRU based hybrid ML method to predict stock prices, achieving higher performance than individual models [14].

3. Applications of Stock Price Prediction Using Machine Learning

Stock price prediction using machine learning is a hot topic in the financial world. Researchers have been investigating, developing, and comparing different ML algorithms to increase the efficiency and accuracy of trend analysis and price forecasting. As technology develops and market grows, model users anticipate seeing more creative ML algorithms and wider application in the financial field. Three main popular applications of stock price prediction using ML are included in this section.

3.1. Risk Identification and Management

Using ML Algorithms to forecast huge changes in stock price movements and market trend can assist investors and financial organizations to identify and reduce the risk of financial investments.

3.2. Investment Analysis and Market Evaluation

ML predictive models can provide helpful insights on evaluating financial market tendencies and recognizing reasons behind them, as well as learning relationships and correlations between economic, political, and media factors that affect stock prices for financial workers.

3.3. Investment Portfolio Optimization and Trading Techniques

ML can be applied in stock price prediction in order to build buying and selling techniques for financial traders and optimize stock portfolio for investors so that they can maximize their profit and make wise and sound financial decisions.

4. Challenges in Stock Price Prediction Using Machine Learning

When applying ML algorithms into stock price forecast, a growing number of researchers have recognized there are several major challenges that need to be addressed in further studies and development.

One of the main challenges is that ML models are not capable of predicting everything. Due to the tremendous instability and high volatility of stock market, it makes prediction extremely difficult with the assumption that ML models analyzes historical data. Only existing factors and previous events are considered. ML may not be able to build the connection between old data and future trends such as stock market crash due to COVID-19, natural disasters and wars. With market sensitivity, there is always a challenge for ML to conform to future changes and develop new models.

Another challenge researchers may face is the high complexity and time-consuming use of ML models. Applying and developing a ML algorithm needs high functionality of computational systems to process great amount of stock data for a long period of time, which can be high-priced and time-consuming. It sometimes takes weeks to train and test data so that the system can produce useful and relatively accurate results.

Limited data availability is another main difficulty that model users and developers experience. They struggle to obtain high-quality and big-quantity stock data, which hugely affects analysis accuracy and effectiveness. Lack of data availability will lead to partial and false predictions. To overcome these challenges, financial researchers and ML developers need to focus on closely monitoring price movements and adjust their systems accordingly. They should also invest in computer system upgrades and feature development to improve the efficiency and accuracy of ML algorithms. In addition, careful selection and preprocessing of stock price data can help improve the quality and quantity of available data, thereby improving overall model performance.

5. Conclusion

In summary, machine learning in stock price prediction has drawn huge attention to financial and artificial intelligence researchers as a result of significant benefits in different fields of applications. In addition to the above four popular ML algorithms (LR, SVM, ANN and LSTM), various types of ML-based models have been developed and applied in the stock market successfully. Despite the significant benefit, several challenges need to be addressed while using ML algorithms for stock price forecast, such as the inability to forecast everything, high complexity and time-consuming use,

and limited data availability. Further research is required to focus on overcoming these challenges and develop more accurate and reliable stock price prediction models using ML.

References

[1] Nti, IK. Adekoya, AF., and Weyori, BA. (2020) A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 53: 3007-3057.

[2] Obthong, M. Tantisantiwong, N. Jeamwatthanachai, W., and Wills, G. (2020) A survey on machine learning for stock price prediction: algorithms and technique. 2nd International Conference on Finance, Economics, Management and IT Business, 63-71.

[3] Umer, M. Awais, M. and Muzammul, M. (2019) Stock market prediction using machine learning (ML) algorithms. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 8(4): 97-116.

[4] Chen, J. (2023) Analysis of Bitcoin Price Prediction Using Machine Learning. Journal of Risk and Financial Management, 16(1): 51.

[5] Panwar, B., Dhuriya, G., Johri, P., Yadav, S. S., and Gaur, N. (2021) Stock market prediction using linear regression and svm. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 629-631.

[6] Emioma, C. C. and Edeki, S. O. (2021) Stock price prediction using machine learning on least-squares linear regression basis. In Journal of Physics: Conference Series, 1734 (1): 012058.

[7] Cory, Mitchell (Investopedia), 2021. Understanding an OHLC Chart and How to Interpret It. www.investopedia.com/terms/o/ohlcchart.asp#:~:text=An%20OHLC%20chart%20shows%20the,structure%20is %20called%20a%20bar.

[8] Yang, J. (2023) Support Vector Machine-based Stock Prediction Analysis. Highlights in Business, Economics and Management, 3: 12-18.

[9] Kumar, M. and Thenmozhi, M. (2006) Forecasting stock index movement: A comparison of support vector machines and random forest. In Indian institute of capital markets 9th capital markets conference paper.

[10] Di Persio, L. and Honchar, O. (2016) Artificial neural networks architectures for stock price prediction: Comparisons and applications. International journal of circuits, systems and signal processing, 10(2016): 403-413.

[11] Shahvaroughi Farahani, M., and Razavi Hajiagha, S. H. (2021) Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. Soft computing, 25(13): 8483-8513.

[12] Taghizadeh Firouzjaee, J. and Khaliliyan, P. (2022) Considering Interpretability of the LSTM Architecture for Oil Stocks Prices Prediction. Available at SSRN 4178888.

[13] Karim, M. E., Foysal, M., and Das, S. (2022) Stock Price Prediction Using Bi-LSTM and GRU-Based Hybrid Deep Learning Approach. In Proceedings of Third Doctoral Symposium on Computational Intelligence: DoSCI 2022: 701-711.

[14] Diqi, M. (2022) StockTM: Accurate Stock Price Prediction Model Using LSTM. International Journal of Informatics and Computation, 4(1): 1-10.