Stock Price Prediction Using Machine Learning and Deep Learning Techniques

by
Yu-Shan Zou

ADVISOR: Ying-Ho Liu

1. Introduction

Investment in the stock market has been regarded as having high risks and high gains, and it attracts a large number of investors. However, the prediction of stock prices is exceptionally challenging. In the past, people used to trade stocks based on feelings. As the number of investors grew, people started searching for tools and methods to maximize profit and minimize risk.

However, stock prices are considered to be volatile. Moreover, it is widely accepted that stock markets are affected by economic factors such as political events, the global economy, invertors' expectations, and companies' financial reports. Thus, researchers have spent much time and effort predicting stock prices to maximize profit and minimize risks.

In the past, researchers used to predict stock prices based on historical market data, analyzing the trend over past years. Later, researchers depended on Intelligent Trading System, which helps predict stock prices based on various situations and conditions, assisting investors in making more precise investment decisions.

Researchers have applied machine learning algorithms to predict stock prices in the literature. Nowadays, the advance in hardware makes performing many calculations reasonably feasible, which helps artificial intelligence solve complex problems. The latest emerging deep learning has achieved tremendous success in the past few years for many applications, including image classification, time series prediction, sound recognition, etc. Therefore, the main goal of this study is to assess the performance of using state-of-the-art machine learning algorithms and deep learning techniques to predict stock prices.

In this study, we use convolution neural network (CNN), long short-term memory (LSTM), Support Vector Regression (SVR), Classification and Regression Tree (CART), K-Nearest Neighbors (KNN), and Linear Regression to predict stock prices. We used four stocks, Yahoo, Google, Netflix, and Tesla, to evaluate the above algorithms. Our research surveys the feasibility of *algorithm trading*, which means trading stocks using computation algorithms. The main contributions of this study are listed as follows.

1. Compare the performance of different prediction models.
2. Categorize the differences between data sets in the different prediction models.

2. Related works

There are many related studies on the stock price prediction. Xia, Liu, and Chen [1] used support vector machines to create a regression model from historical stock market data and predict the performance of the stock markets. Zhang and Tan [2] used historical price data to predict future stock return rankings through a novel stock selection model based on a deep neural network. LSTM was combined with a naive Bayesian method to extract emotional factors from the market for improving prediction performance [3]. This method can forecast financial markets on different time scales with different variables. In [4], the emotion analysis model was integrated with the LSTM time series learning model to create a robust time series model to predict the opening price of a stock, and the results showed that this model could improve the accuracy of the prediction.

Many studies have shown that deep neural networks have a better ability to learn time-series data than other machine learning methods due to their memorability. Nelson, Pereira, and De Oliveira [5] predicted future stock trends in the Brazilian stock market based on historical prices and technical indicators. They showed that a deep neural network offers significant advantages in inventory forecasting accuracy over other machine learning methods. Jia [6] discussed the effectiveness of LSTM in predicting stock price, and the study found that LSTM is an effective method of predicting stock earnings. Real-time wavelet denoising and an LSTM network were coupled to predict the East Asian stock index [11]. Compared to the original LSTM, this combined model significantly improved with high prediction accuracy and small regression error. The evolutionary method predicted the stock price time series [7]. A deep belief network with inherent malleability was used to predict the stock price time series [8]. In addition, the convolutional neural network has been used to predict stock price trends [9]. A forward multi-layer neural network model was built to predict future stock prices [10]. The model used a hybrid method combining technical analysis variables, fundamental stock index analysis variables, and the BP algorithm.

Inspired by the above research, we conducted this study to examine further the feasibility and performance of selected machine learning algorithms and deep learning algorithms for predicting stock prices.

3. Method

Predicting stock prices is a complex problem because there are many factors to consider. However, with machine learning techniques, it is possible to relate previous data to current data and train prediction models.

For deep learning techniques, we use CNN to obtain the correlation of daily stock transaction data over a period and then adopt LSTM to make predictions. In this section, we introduce the algorithms used in the study, namely, CNN, LSTM, SVR, CART, KNN, and linear regression.

1. Convolution neural network (CNN)

   CNN was widely used in image recognition because of its powerful pattern recognition ability. In recent years, it has been increasingly applied in stock prediction. CNN has shown promising results in capturing the complex patterns and dependencies in financial time series data.

   Suppose the characteristics of the stock transactions at a specific time point are regarded as a feature graph. In that case, CNN has the potential to extract the characteristics of the stock transactions at the corresponding period from these feature graphs. Therefore, CNN can be used to build a timing-selection model and can ultimately be used to complete the construction of the timing-selection strategy.

2. Long short-term memory (LSTM)

   The LSTM model is a type of RNN that will read each time step of the financial time series one step at a time. Its core contribution is to introduce the design of a self-loop to generate the path of a gradient that could continuously flow for an extended period. LSTM has an internal memory allowing it to accumulate an internal state as it reads across the steps of a given input sequence.

   The advantage of LSTM networks lies in the fact that short-term and long-term values can be remembered. Therefore, LSTM networks are mainly used for sequential data analysis.

3. Support Vector Machines / Regression

   Support Vector Machine is one of the most popular supervised learning techniques, which can be applied not only to classification problems but also to problems with time series analysis and regression.

Recently, it has become a hot topic of intensive study due to its successful application in classification and regression tasks, especially in time series prediction and financial-related applications. The most crucial feature of SVR is the principle of the maximum margin algorithm.

4. Classification and Regression Tree (CART)

The decision trees method has been one of the most critical classification and prediction methods in recent years. CART is one of the most commonly used decision tree methods selected within the framework of predicting a quantitative trait. The CART method has advantages such as detecting interactions in data sets. Also, CART has a built-in algorithm to deal with the missing values of a variable for a case.

5. K-Nearest Neighbors (KNN)

K-Nearest-Neighbors (KNN), one of the most essential and practical algorithms for data segregation, can become the primary choice for implementation, especially when the given data is quite ambiguous.

The K-NN algorithm only applies to a few adjacent samples in category decision-making. Since the K-NN algorithm mainly relies on the surrounding limited adjacent samples, rather than relying on the discriminant domain method to determine the category, thus the K-NN algorithm is more suitable than other methods for the pending sample sets where the class domain crosses or overlaps more.

6. Linear Regression

Linear regression is a supervised machine learning technique that aims to build a learning model and establish a relationship between two variables by best fitting a linear line to input data.

Linear regression can be helpful for stock price prediction when combined with appropriate variables and data sources.

4. Experiment results

In this section, the performance of different algorithms on each dataset is presented in terms of mean square error, mean absolute error, and mean absolute percentage error. The three metrics are briefed as follows.

- The mean square error (MSE) represents the average of the residual sum of squares of the actual and predicted values. It intends to measure the degree of variation between the actual and predicted values.
  The smaller the MSE value, the better the effect and the closer the prediction ability is to the actual data.

- The mean absolute error (MAE) is the average absolute value of the error and intends to measure the mean of the sum of the predicted errors.
  The smaller the value of MAE, the smaller the total value of the prediction error.

- The mean absolute percentage error (MAPE) is the average absolute value of the percentage error. The smaller the value of MAPE, the smaller the prediction error.

The stock data were retrieved from: https://www.kaggle.com/. The daily open price, close price, highest price, lowest price, and transaction volume are retrieved. The data period of each stock is as follows.

- Yahoo stock from 2015/11/23 to 2020/11/20
- Google stock from 2020/3/11 to 2022/9/12
- Netflix stock from 2018/2/5 to 2022/2/4
- Tesla stock from 2019/9/30 to 2022/4/11

The programs used for prediction are written in python, and the code is available on GitHub. Links: https://github.com/yushan21/Code.git

The following tables show the experiment results.

Table 1 The performance on Yahoo stock

	MSE	MAE	MAPE
CNN-LSTM	492192.09	654.71	0.2
SVR	2426.58	31.35	0.01
CART	42702.79	161.70	0.05
KNN	49866.05	190.63	0.06
Linear regression	2572.77	29.51	0.01

Table 2 The performance on Google stock

	MSE	MAE	MAPE
CNN-LSTM	17.08	3.53	0.03
SVR	8.21	2.31	0.02
CART	23.70	23.70	0.03
KNN	24.68	4.14	0.04
Linear regression	8.04	2.26	0.02

Table 3 The performance on Netflix stock

	MSE	MAE	MAPE
CNN-LSTM	26582.58	77.0	0.12
SVR	189.02	8.54	0.02
CART	1684.12	29.20	0.05
KNN	3252.03	40.47	0.07
Linear regression	180.45	8.30	0.02

Table 4 The performance on Tesla stock

	MSE	MAE	MAPE
CNN-LSTM	27792.37	135.46	0.13
SVR	2245.63	36.64	0.04
CART	36417.65	160.71	0.15
KNN	32930.02	151.47	0.14
Linear regression	1930.52	33.35	0.03

As the tables above show, the linear regression prediction method achieved the lowest MSE, MAE, and MAPE among the five methods. Low MSE, MAE, and MAPE indicate that the model accurately predicted future prices. The performance demonstrates that the linear regression model predicted the stock prices better than other algorithms. Besides linear regression, SVR also had similar prediction results. Thus, we can conclude that these two prediction methods have better prediction ability.

5. Conclusion

Prediction is about predicting an unseen future. Prediction errors must be within acceptable limits to create a reliable prediction. This study conducts experiments on the performance of different prediction models in stock market prediction. Of all the algorithms we applied, we saw that support vector regression and linear regression prediction methods have better prediction ability. Using different models to predict stock prices can give people a more accurate understanding of future stock prices. It can also be a useful tool for investors to make stock market decisions.

In the future, we plan to combine those techniques with different feature selection techniques and use different kinds of models to predict stock prices.

References

1. Xia Y, Liu Y, Chen Z (2013) Support vector regression for prediction of stock trend. 6th Int Conf Inf Manag Innov Manag Ind Eng (ICIII) 2:123–126
2. Zhang, X., & Tan, Y. (2018). Deep stock ranker: A LSTM neural network model for stock selection. International conference on data mining and big data. Springer614–623.
3. Li J, Bu H, Wu J (2017) Sentiment-aware stock market prediction: a deep learning method. In: IEEE: 2017 international conference on service systems and service management, pp 1–6
4. Zhuge Q, Xu L, Zhang G (2017) LSTM neural network with emotional analysis for prediction of stock price. Eng Lett 2(25):167–175
5. Nelson, D. M., Pereira, A. C., & De Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. Proceedings of the

international joint conference on neural networks IEEE1419–1426. https://doi.org/10.1109/IJCNN.2017. 7966019.

6. Jia H (2016) Investigation into the efectiveness of long short term memory networks for stock price prediction. arXiv [cs.NE], pp 1–6
7. Tsai Y-C, Hong C-Y (2017) The application of evolutionary approach for stock trend awareness. In: IEEE: IEEE 8th international conference on awareness science and technology (iCAST), pp 306–311
8. Li X, Yang L, Xue F, Zhou H (2017) Time series prediction of stock price using deep belief networks with Intrinsic plasticity. In: IEEE: 2017 29th Chinese Control and Decision Conference (CCDC), pp 1237–1242
9. Gudelek MU, Boluk SA, Ozbayoglu M (2017) A deep learning based stock trading model with 2-D CNN trend detection. In: IEEE: 2017 IEEE symposium series on computational intelligence (SSCI), pp 1–8
10. Adebiyi AA, Ayo KC, Adebiyi MO, Otokiti SO (2012) Stock price prediction using neural network with hybridized market indicators. J Emerg Trends Comput Inf Sci 1(3):1–9
11. Li Z, Tam V (2017) Combining the real-time wavelet denoising and long- short-term-memory neural network for predicting stock indexes. In: IEEE: 2017 IEEE symposium series on computational intelligence (SSCI), pp 1–8