Research on the Google Stock Price Prediction Based on SVR, Random Forest, and KNN Models

Yuping Huang *

Department of Mathematics, University College London, London, Untied Kingdom

* Corresponding Author Email: zcahuaj@ucl.ac.uk

Abstract. The prediction of stock prices is an intriguing and difficult study issue. Due to its frequent offers of quick rewards and low risk of loss, the stock market is currently regarded as a prestigious trading sector. Data mining and business researchers find the stock market to be an ideal environment due to its extensive and dynamic information sources. In this paper, the data of google stock price from 2018-1-2 to 2023-3-31 is imported. As our classification algorithms, the authors independently selected and contrasted super vector regression, random forest, and K-nearest neighbor. Following test, analysis, and comparison of these classifiers and calculate the mean square error of test data, the final result was reached. According to the result, the random forest has the smallest MSE value, so it is the best prediction and uses this method to predict the stock price of next ten days. After using random forest to do the stock price prediction, we find that the close price has the increasing trend for next ten days, which is good for the investors.

Keywords: Stock price prediction; super vector regression; random forest; K-nearest neighbor; MSE value.

1. Introduction

Google is a global technology business with a focus on internet-related services and goods. A crucial and essential component of today's financial markets is Google stock. Recent business study interests have been focused on areas of future stock price movement predictions, which makes it difficult and demanding. Researchers, corporate communities, and interested individuals are eager to find the stock price forecast of movements in stock markets because they believe that future events will depend on historical data [1].

Companies that provide financial services are introducing new tools to help with future forecasting. Stock market mining, which is another name for stock prediction, is one of several areas of financial information sources that can be useful research topics. Stock forecasting becomes more crucial, especially if several guidelines can be developed to aid in improving investing choices across various stock markets.

To forecast stock market prices, Kazema uses support vector regression with chaos-based fireflies. They discover that, when compared to other algorithms, the 3-fold model performs best based on two error measures [2]. Presenting Ullah's Prediction on Google Stock Prices They forecasted the stock price for the first two months of 2017 using deep learning, using the Recurrent Neural Network (RNN) as a deep learning model, and they were able to do so with an accuracy of 87.32% [3, 4]. For the purpose of classifying texts, Shah compares the performance of the Logistic Regression, Random Forest, and KNN models. They discovered that the Logistic Regression classifier achieves the maximum accuracy when compared to the KNN and Random Forest classifiers [5, 6].

On the basis of the previous literature, this paper is not only an important supplement and enrichment to the literature in the field of stock price forecasting, but also provides some method guidance for investors to forecast stock price and make investment decisions in practice.

This study uses Support vector regression for Google stock price prediction, the result is showed in figure2. The predicted price is almost similar to the original close price, expect that train predicted price is lower than original price at 2022 when close price reaches the peak and test predicted is higher than original price at 2023.

2. Data and Methods

2.1. Data

2.1.1. Data Sample

This study uses the close price of Google from 2018-1-2 to 2023-3-31 as the data set. The data was collected from Kaggle and the stock price trend was shown in figure 1. The close price increased from 60 to 140 between 2018 and 2022. After reaching a peak at 140 in 2022, the close price then declines to around 80 in 2023.

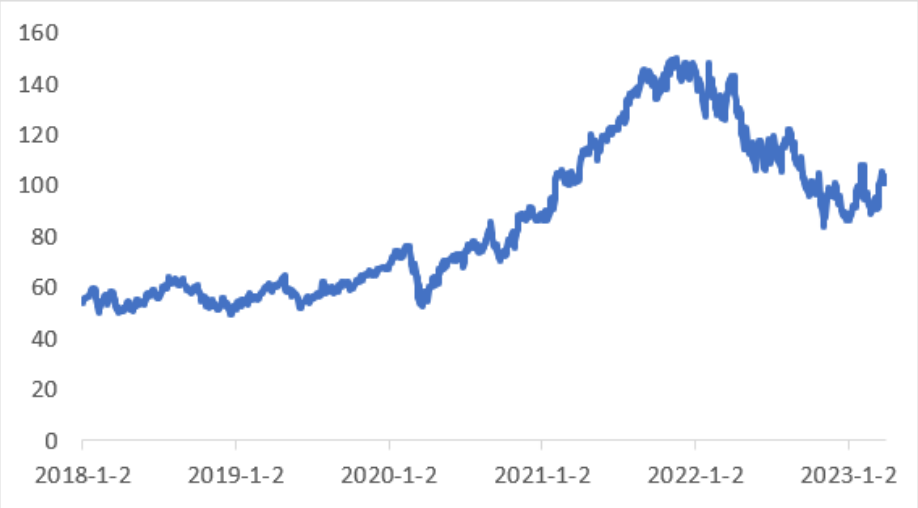


Fig. 1 Data imported

2.1.2. Normalization

This study uses MinMaxScalar to normalize the close-set data between 0 and 1, and then split the data set into training data and testing data. Training data is from 2018-1-2 to 2022-3-24 and testing data is from 2022-3-24 to 2023-3-31.

2.2. Methods

2.2.1. Support Vector Regression

A series of training patterns(x_1, y_1)... (x_n, y_n) where x_n is inR^n, n = 1..., i, and y_n is in R, are provided to us in a regression issue. The input vector$x_n$intended target, or output value, is represented by eachy_n. These patterns are utilized to train a regression model, which is then applied to forecast the target values of unknown input vectors. SVR looks for a regression hyperplane in a high-dimensional feature space that has a low risk. It has strong generalization and function approximation skills [2].

2.2.2. Random Forest

The training phase of a random forest uses a number of decision-tree classifiers, sometimes referred to as regressors. Each decision-tree classifier in RF is produced by randomly choosing the input variables and training data. Every RF's decision tree is robust enough to handle a large number of variables without overfitting [7, 8].

2.2.3. K-Nearest Neighbor

K-nearest neighbor technique is also one of machine learning approaches. A classification system based on similarity can be used to map the stock prediction problem. The calculation of a similarity measure is necessary to make a choice.

The training data set's k records with the most similar k values to the test are made using kNN. A majority vote is used to decide the class label, and it is then applied to the query record. kNN is used to calculate the following closing price prediction for the stock market:

a) Determine the k nearest neighbors.
b) Calculate the separation between the query record and the training samples.
c) Arrange all training data records by distance values.
d) Assign the class labels of the k nearest neighbors that received the most votes as the forecast value for the query record [1].

3. Results

3.1. Results of Support Vector Regression

This study uses SVR for Google stock price prediction, the result is showed in figure 2. The predicted price is almost similar to the original close price, expect that train predicted price is lower than original price in 2022 when close price reaches the peak and test predicted is higher than original price at 2023.

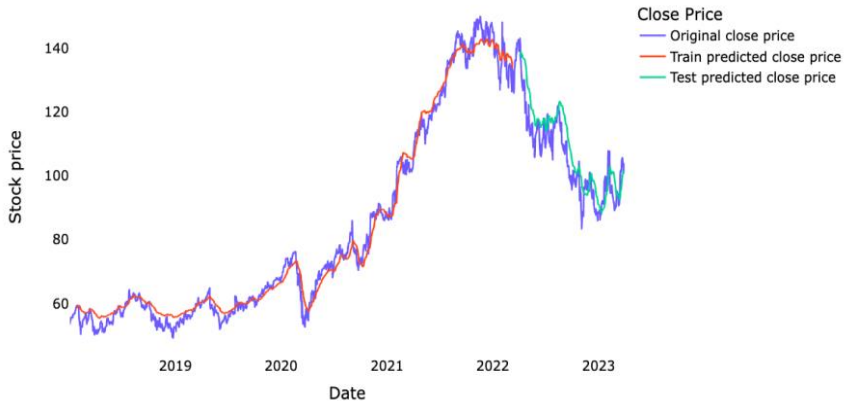


Fig. 2 Comparison between original close price vs predicted close price for SVR.

3.2. Results of Random forest

This study uses Random Forest for Google stock price prediction, the result is showed in figure 3. The predicted price is almost same as the original close price, but only except the lowest point at 2023.1.

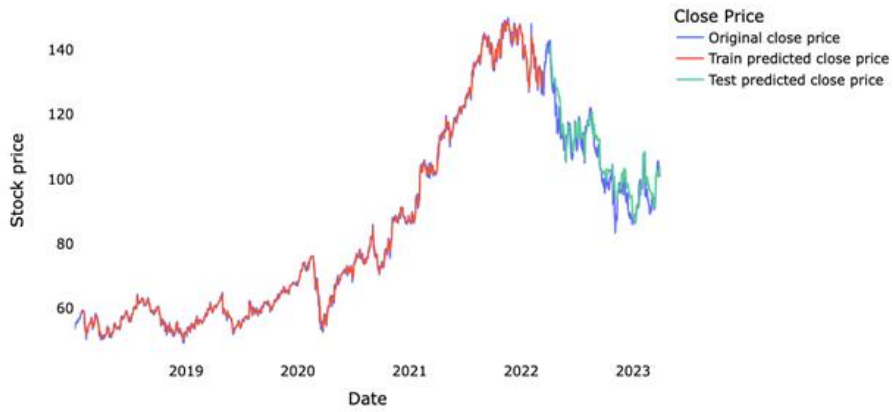


Fig. 3 Comparison between original close price vs predicted close price for RF.

3.3. Results of K-Nearest Neighbor

This study uses Random Forest for Google stock price prediction, the result is showed in figure4. The train predicted price is almost same as the original close price. However, there is slight difference between test predicted price and original close price. There was a significant decrease for original price, but the test predicted price is still steady at that period.

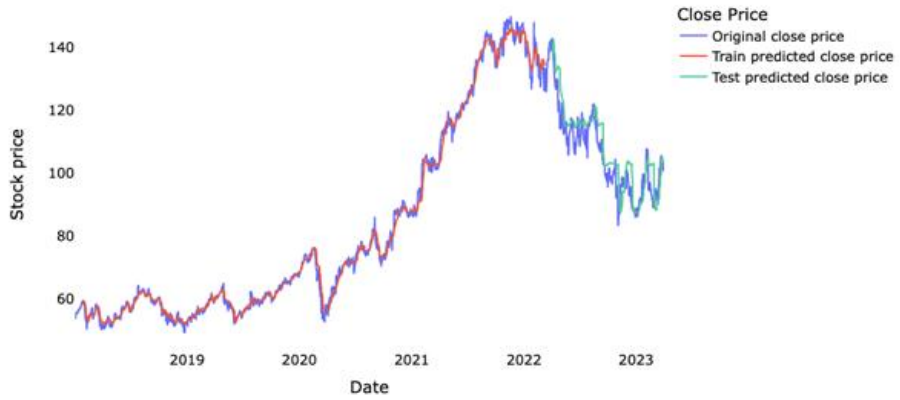


Fig. 4 Comparison between original close price vs predicted close price for kNN.

3.4. Comparison of Three Methods

This study uses MSE to analysis the accuracy of this three-prediction model, the result is showed in table 1. Random forest has the smallest MSE value which is 13.4 and MSE value of SVR is the largest. Comparing with the result from the research of Adlakha et al., the result is similar [9, 10].

Table 1. MSE values for SVR, RF and kNN.

Method	MSE value
Support Vector Regression	43.8
Random Forest	13.4
K-nearest neighbour	37.7

3.5. Comparison of Three Methods

Since Random Forest has the smallest MSE values, so this study uses RF to predict the close price for next ten days, the result is showed in figure 5.

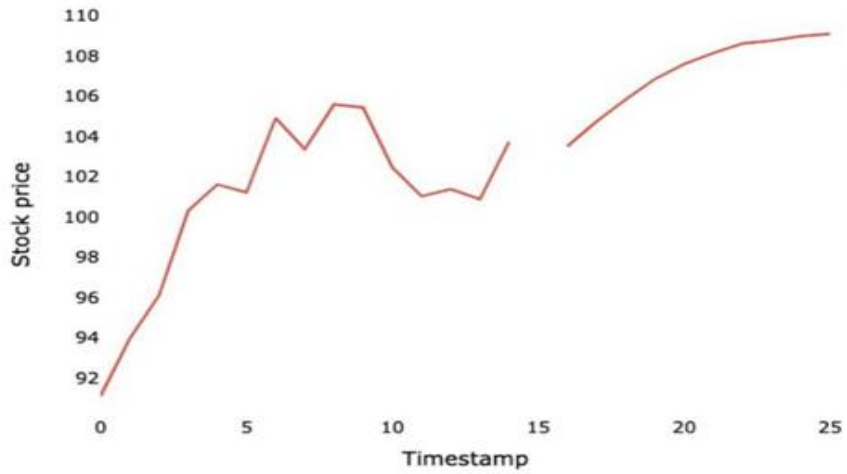


Fig. 5 Compare last 15 days vs next 10 days.

4. Conclusion

In order to predict rising stock prices, this study examined SVR, RF, and kNN modeling methodologies along with numerous variable selection strategies. Given that RF's MSE is 13.4, it fared better than SVR and kNN models. These findings imply that RF are more effective at determining Google's close price.

This paper has some implications. Firstly, the close price can be used as an important indicator of Google's future close price, and the use of Random Forest in machine learning has a better prediction

effect compared to SVR and kNN methods, which reveals that investors can refer to this model when analysing Google's close price and help them to make better investment decisions. On the other hand, this paper uses Random Forest to predict the stock price in the next ten days which shows that Google has more upward trend in the coming period, which has better investment value for investors.

However, this study hasn't use LSTM or other machine learning methods to do the price prediction. In the future, more machine learning methods can be further analysed and more comprehensive comparisons can be made to select better predictive models.

References

[1] Khalid Alkhatib, et al. Stock Price Prediction Using K-Nearest Neighbor (KNN) Algorithm. International Journal of Business, 2013, 3(3).

[2] Ahmad Kazem, et al. Support Vector Regression with Chaos-Based Firefly Algorithm for Stock Market Price Forecasting. Applied Soft Computing, 2013, 13(2):947–958.

[3] Kifayat Ullahand, Muhammad Qasim. Google Stock Prices Prediction Using Deep Learning.2020.

[4] Ishita Parmar, Ridam Arora, Lokesh Chouhan, Navanshu Agarwa, Shikhin Gupta, Sheirsh Saxena, et al. Stock Market Prediction Using Machine Learning. First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018.

[5] Kanish Shah, et al. A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. Augmented Human Research, 2020, 5(1).

[6] Kumar R, Kaur J. Random forest-based sarcastic tweet classification using multiple feature collection. Tanwar S, Tyagi S, Kumar N (eds) Multimedia big data computing for IoT applications. Intelligent systems reference library, 2020,163.

[7] Park, Hyun Jun, et al. Stock Market Forecasting Using a Multi-Task Approach Integrating Long Short-Term Memory and the Random Forest Framework. Applied Soft Computing, 2021, 1: 106-108.

[8] Y. Ma, R. Han, X. Fu. Stock prediction based on random forest and LSTM neural network. 19th international conference on control automation and systems, ICCAS, Jeju, Korea (South), 2019, 2019(10):126–130.

[9] Adlakha, N., Ridhima and Katal, A.Real Time Stock Market Analysis. 2021 International Conference on System, Computation, Automation and Networking (ICSCAN). 2021.

[10] M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar. Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science, 2020, 167:599-606.