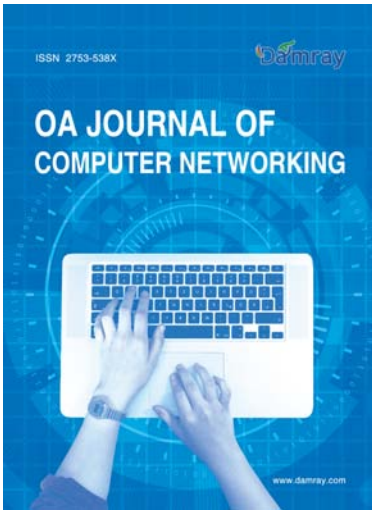
Damray | OA Journal of Computer Networking
2023, 2(1), 1-5
ISSN: 2753-538X



https://oajcn.damray.com/

OPEN ACCESS

DOI: 10.26855/oajcn.2023.06.001
Received: June 12, 2023
Accepted: July 9, 2023
Published: August 7, 2023

Stock Closing Price Prediction by ARIMA and SVM Models

Lisa Lan

Jinling High School, Nanjing, Jiangsu, China.

Abstract

Stock price prediction is a challenging task because stock market is dynamic, unpredictable, noisy and volatile in nature. To address these challenges, various machine learning algorithms and models are developed to identify patterns in the stock price movements. Stock forecasting is so appealing for both institutional and individual investors. For institutional investors, an increase of just a few percentage points can increase profit by millions of dollars. Individual investors also seek profits through market participation by investing spare money in market. For a long time, researchers have worked on advanced intelligent techniques based on either technical or fundamental analysis of stocks to identify patterns in the stock price movements using advanced data mining techniques. Different models, including linear regression, Autoregressive Integrated Moving Average (ARIMA), Random Walk Theory (RWT), Moving Average Convergence/Divergence (MACD), Support Vector Machine (SVM), Autoregressive Moving Average (ARMA), Random Forest (RF), neural networks such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and deep neural networks like Long Short Term Memory (LSTM) have been used for stock price prediction and shown promising results. In this paper, ARIMA model and SVM mode are applied to visualize, analyze and predict Apple Inc. closing stock price between Jan 01, 2012 and Jan 01, 17, 2023, I focus on two specific machine learning techniques, namely ARIMA mode and SVM model and applied them to visualize, analyze and predict Apple Inc. The results displayed that ARIMA model is better than SVM model with smaller MSE and high R-squared.

Keywords

Stock price, models, Apple Inc.

1. Data and Models

1.1 Data collection

The Apple Inc. data set used in this project is collected from yahoo finance. It covers daily price from 01-Jan-2012 to 17-Jan-2023.

1.2 The Autoregressive Integrated Moving Average (ARIMA) Model

Time series models are used to measure a set of variables x_t and put them in order at specific intervals $t_1, t_2, \ldots, t_n$ over a time period to identify the trend and correlation among variables [1]. The predicted data series with time lapse is

considered as a random sequence, and their dependent relation represents the time continuity of the original data. ARIMA is the most widely used model for stationary fitting sequences. The equation is given as below.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q} \quad (1)$$

When the time series is not stationary, it can be treated as stationary if the increase, namely, the first difference, is stabilized near zero. In practice, most non-stationary sequences can be transformed into stationary time series after one or more differences for the ARIMA model [2].

1.3 Support Vector Machine (SVM) Model

SVM can be used for classification as well as regression problems although it is primarily used for classification problems in Machine Learning (ML). Basically, the goal of the SVM algorithm is to find hyper-plane to create the best line in 2-dimensional space or decision boundary segregating n-dimensional space into classes [3]. Thus, the new data point can be easily put in the correct category in the future.

The support vectors or the extreme points/vectors are chosen to create hyperplane. Thus, the algorithm is termed as Support Vector Machine. The basic step of SVM model are: Data reading preparation, Define the explanatory and target variables, Split the data into train and test, Support Vector Classifer (SVC), Classifier accuracy and Strategy implementation.

2. Experimental Results and Discussion

2.1 Modelling of Apple Inc. Closing Price by ARIMA model

First, I visualize the Timing diagram of original AAPL Inc. stock closing price from 2012 to 2022, and also the data in series through a probability distribution. Then I analyze the APPL Inc. closing price by dickey fuller test to check whether the series is stationary since ARIMA model only works with stationary data. In the below Figure 1, the increasing rolling mean and standard deviation indicate that the series is not stationary [4]. In addition as shown in Table 1, p-value is greater than 0.05 and the test statistics is greater than the critical values. Therefore, the data is non-stationary.

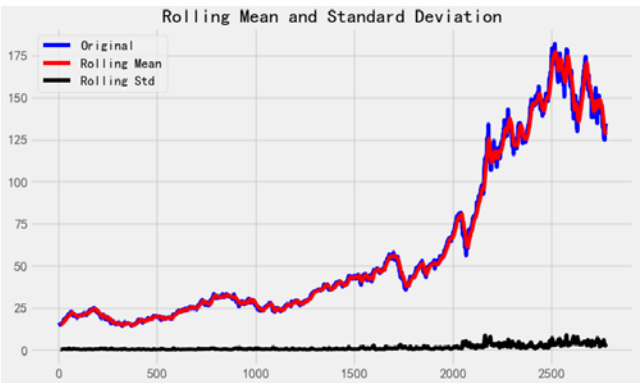


Figure 1. Rolling mean and standard deviation of AAPL Inc. stock closing price for dickey fuller test.

Table 1. Results of dickey fuller test

	Values
Test Statistics	-0.250778
p-value	0.932192
No. of lags used	22.000000
Number of observations used	754.000000
Critical value (1%)	-3.432727
Critical value (5%)	-2.862590
Critical value (10%)	-2.567329

To perform ARIMA modelling, stationary series is generated through separating seasonality (the repeating short-term cycle in the series) and trend (the increasing or decreasing value in the series) from our series. The log value of the series is taken to reduce the magnitude of the values and the rising trend in the series [5]. The rolling average of the log of the series is shown in Figure 2. In addition, as seen in Figure 3, the standard deviation is flat and the series is stationary after separation of seasonality and trend.

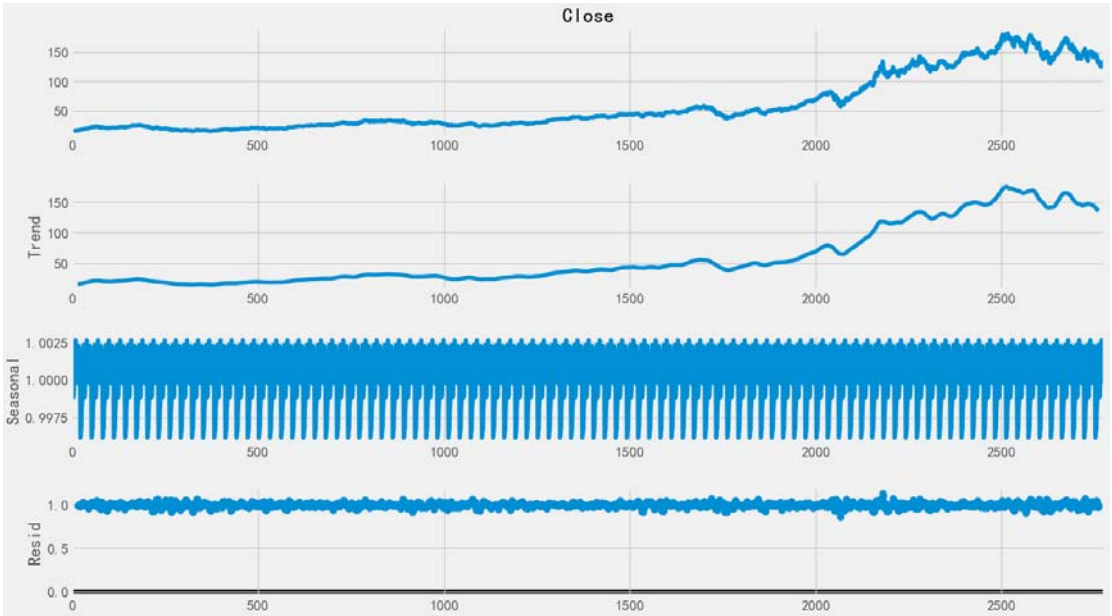


Figure 2. Separation of seasonality and trend from AAPL Inc. stock closing prices series.

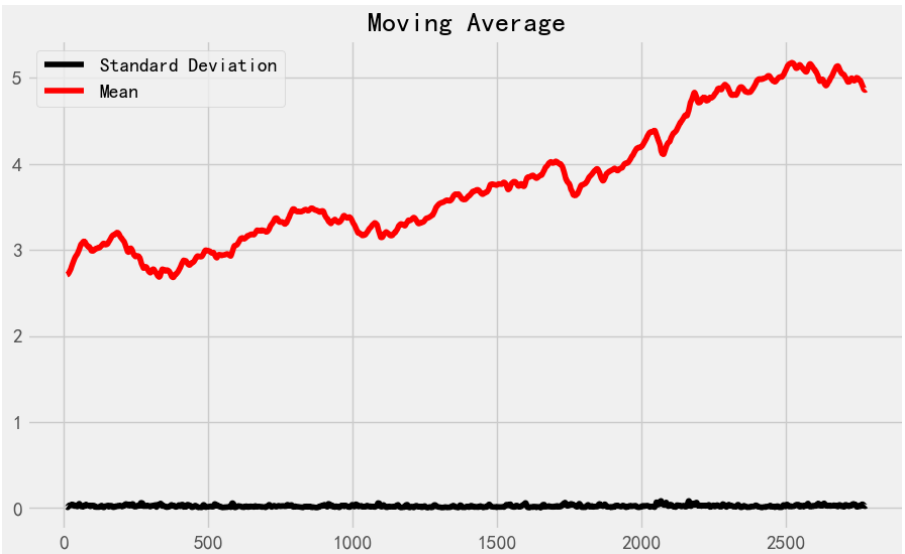


Figure 3. Moving mean and average after separation of seasonality and trend from AAPL Inc. stock closing prices series.

The auto arima function in python is applied to fit an ARIMA model. It can automatically discover the most optimal parameters for an ARIMA model and return a fitted ARIMA model.

In Fig 4, the top left panel shows that residual errors fluctuate around a mean of zero and have a uniform variance. The top right density plot suggests normal distribution with mean zero. In the bottom left theoretical quantile plot, there is not significant deviations [6]. In addition, in the bottom right ACF plot, the residual errors are not autocorrelated. Thus, it seems to be a good fit overall.

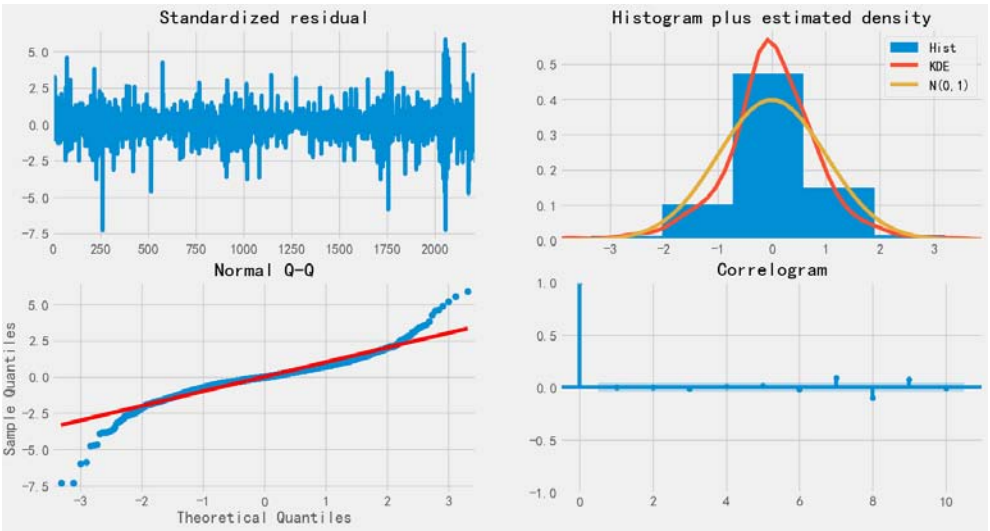


Figure 4. Residual errors, density, sample quintiles and Correlogram, aka, ACF plot of AUTO ARIMA model.

Finally, the values of three order parameters, namely, p (AR Autoregression), d (Integration), q (Moving Average) values from auto ARIMA model are applied for modelling. The prediction figure is shown as below. The R-squared valued is greater than 0.99, indicating that the model is a good fit.

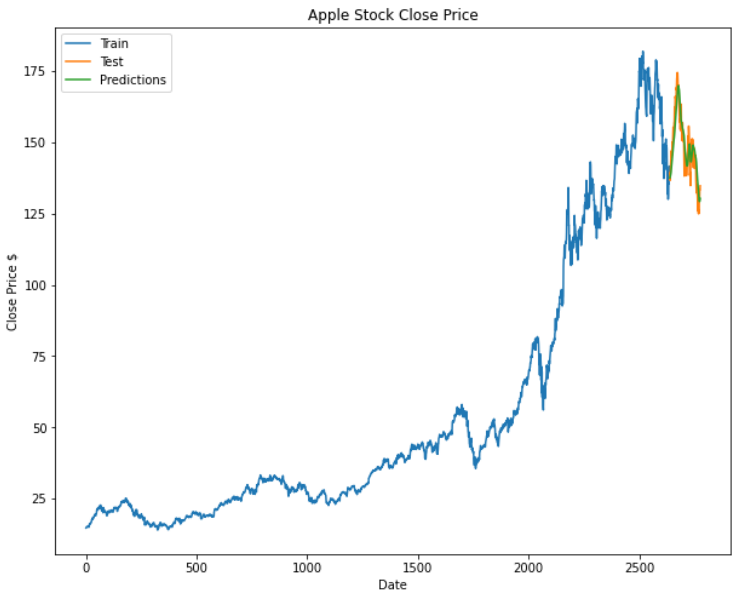


Figure 5. AAPL Inc. Closing price predicted by ARIMA with p, d, values by auto ARIMA.

2.2 Modelling of Apple Inc. Closing Price by SVM model

Explanatory or independent variables are used to predict the value response variable. The dataset with the variables for prediction consists of variables such as 'Open – Close' and 'High – Low', which are indicators based on which the algorithm will predict tomorrow's trend. The target variable is the outcome which the ML model will predict based on the explanatory variables. If tomorrow's price is greater than today's price, the signal will be buyed and stored as 1. Otherwise, the signal will be stored as 0 in the target dataset y. The accuracy of SVM model is computed by comparing the actual values of the signal with the predicted values of the signal. The accuracy in the test data is greater than 50%, indicating that the classifier model is effective. The modelling of AAPL Inc. stock price by SVM is shown in Figure 6. The strategy by SVM totally outperforms the performance of the reliance stock after year 2019. The strategy (Blue Line)

provided the return of 32% in the last 1 year whereas the stock of reliance (Red Line) only provides the return of just 27% in the last 1 year.

Figure 6. Prediction and actual values by SVM model for AAPL Inc. stock price.

F1 score is an alternative ML evaluation metric that assesses the predictive skill of a model by elaborating on its class-wise performance. An F1 score reaches its best value at 1 and worst score at 0. The F1 score of our model is 0.49.

2.3 Model Comparison

To compare the effects of the two models, the Mean Square Error (MSE) is used as the criterion for evaluation. The smaller the MSE is, the better the model is. Equation to calculate MSE are given is below.

$$MSE = \frac{1}{n}\sum(Y - \hat{Y})^2 \tag{2}$$

The MSE of ARIMA model is 0.0134, which is much lower than 30.7685 of the SVM model. If only MSE is considered, ARIMA is a better fit than SVM model to predict AAPL Inc. stock price. As suggested by SVM model, the prediction after 2019 highly outperforms the actual AAPL price. This might due the fact that SVM cannot predict the unexpected external factors. Obviously, since 2019, the widespread of Covid-19 worldwide had a great effect on world market. The SVM model and most of current ML algorithms cannot include predict unexpected external factors. In the future, new model will be developed to include external factors.

3. Conclusion

It is difficult to predict the stock prices due their dynamic, chaotic and volatile nature. The analysis of AAPL Inc. stock prices in this paper suggests that forecasting ARIMA and SVM models, especially the ARIMA model can be used effectively to predict stock prices with a reasonably high accuracy. The drawback of the prediction models is that they are unable to include unexpected external factors, which could be a direction for future research.

References

[1] G. AF. Seber, A. J. Lee. (2012). Linear regression analysis. John Wiley & Sons: 329.

[2] E.F. Fama. (1965). Random Walks in Stock Market Prices. Financial Analysts Journal, 21(5): 55-59.

[3] H. Hamzah. (2022). Effective Stock Prediction Model Using MACD Method. International Journal of Informatics and Computation (IJICOM) 2022. 4(2): ISSN: 2685-8711, E-ISSN: 2714-5263. DOI: 10.35842/ijicom

[4] A. A. Ariyo, A. O. Adewumi and C. K. Ayo. (2014). Stock Price Prediction Using the ARIMA Model. 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, 106-112.

[5] A. Ganesan, A. Kannan.(2021). Stcok Price Prediction using ARIMA model. International Research Journal of Engineering and Technology. 2021. 8(8): 226-234.

[6] K. Liagkouras, K. Metaxiotis. (2020). Stock Market Forecasting by Using Support Vector Machines. Machine Learning Paradigms. 2020. 259–271.