ARIMA Model for Accurate Time Series Stocks Forecasting

Shakir Khan1, Hela Alghulaikha2
College of Computer and Information Sciences
Imam Mohammad Ibn Saud Islamic University
Riyadh, Saudi Arabia

Abstract—With the increasing of historical data availability and the need to produce forecasting which includes making decisions regarding investments, in addition to the needs of developing plans and strategies for the future endeavors as well as the difficulty to predict the stock market due to its complicated features, This paper applied and compared auto ARIMA (Auto Regressive Integrated Moving Average model), Two customize ARIMA(p,D,q) to get an accurate stock forecasting model by using Netflix stock historical data for five years. Between the three models, ARIMA (1,1,33) showed accurate results in calculating the MAPE and holdout testing, which shows the potential of using the ARIMA model for accurate stock forecasting.

Keywords—ARIMA; forecasting; prediction analysis; time series; stocks forecasting; data mining; big data

I. INTRODUCTION

The increasing availability of historical data with the need for production forecasting has attracted the attention of Time Series Forecasting (TSF), which gives a sequence of predicting future values, especially with the limitations of traditional forecasting, such as complexity and time-consuming [1]. The future prediction of system behavior by TSF based on current and past information. The role of TSF is part of several real-world problems, such as network traffic, petroleum, weather forecasting, and financial markets [2]. The empowered institutions and individuals to make decisions to invest and the need to develop plans and strategy of future endeavors made the prediction exciting area for the domain researchers to work and improve the predictive models [3, 4, 5]. Especially when the decision-making process, in general, considered not the role of forecasting stock to setting a trading strategy, determining the actions for appropriate timing to buy or sell stocks and to study future investment opportunities as well as the importance of developing and improving time series forecasting models and study its effectiveness and success, this paper aim to get the accurate stocks forecasting model by comparing the results of accuracy of auto ARIMA model and two customize ARIMA (p,D,q) models which will be applied on Netflix stocks historical data for last five year. By applying this model to forecast for Netflix's future, especially since it showed an essential role in people’s life today with what the world is facing from COVID-19. Therefore, it is quite essential having a clear understanding of the present as well as forecasting the future when aiming to have a safe investment. It also contributes to understanding the role of the time series forecasting ARIMA model and the accuracy of its techniques.

The rest of the paper is organized as follows. Section II literature review. Section III methodology, Section IV result and discusses and the conclusion in Section V.

II. LITERATURE REVIEW

Improving the accuracy of stock forecasting was part of cao et al. [18], study, where they combined Empirical Mode Decomposition (EMD) with the Long Short-Term Memory (LSTM) in their proposed model, where the result showed better performance. It has also been studied the efficiency to improve the predictability of stock returns by proposing simple way which was based on existing predictors with low correlations instead of new powerful predictors [19]. As well as the complexity of stock data has been discussed and its need for efficient prediction system and proposed model which shows better accuracies forecasting [11]. Using hybrid models was part of the studies to face the complexity of linear and nonlinear components where the ANN-ARIMA hybrid model has been evaluated and showed more accurate results than the conventional ARIMA-ANN model [12]. The majority of prior research on stock time series forecasting focused on proposing...
III. METHODOLOGY

Prediction the future of stocks values using ARIMA model it will be by testing the auto ARIMA values as well as build customize ARIMA (p,D,q) models to get better forecasting model. The ARIMA model applied on real Netflix stock data which is available for public on Yahoo! Finance [20]. The dataset contains Netflix daily stock price data for five years, starting from 7 April 2015 to 7 April 2020. The data in Fig. 1, describes the date, open which is the price at the beginning of the day, high which is the highest price during the day, low which is the lowest price during the day, close which is the price at the end of the day, adjusted closing which is the price of stock's closing price amended to accurately reflect that stock's value after accounting for Netflix actions, and the volume which is the number of stocks of a security traded during that day. The forecasting process adjusted closing values which had only counted, since it is representing the real closing value of the day as well as this value has been scaled for more accurate readings. The model applied using R language in R Studio. Determine the model accuracy and the comparing between the several experiments in the model will be based on calculate Autocorrelation Functions (ACFs), Partial Autocorrelation Function (PACF) as well as Mean Absolute Percentage Error (MAPE).

A. ARIMA Model

Auto Regressive Integrated Moving Average (ARIMA) is a model describes time series given based on observed value which can be used to forecast future values. Applying ARIMA models on Any time series show patterns with no random white noise and non-seasonal [21]. The model introduced by Box and Jenkins in 1970. To generate short-term forecasts, ARIMA models showed efficient capability outperformed complex structural models [3]. The future value of a variable in ARIMA model is a combination of linear to the past values and errors, expressed as follows:

\[ Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \] (1)

Where,

\( Y_t \) is the actual value and \( \varepsilon_t \) is the random error at time \( t \), \( \phi \) and \( \theta \) are the coefficients, \( p \) and \( q \) are integers that are often referred to as autoregressive and moving average, respectively. [22].

B. Auto ARIMA and ARIMA (p, D, q) Models Implementation

Exploring Netflix stocks data from 7th April 2015 to 7th April 2020, showed the non-stationary characteristics of time series as shown in Fig. 2. To reduce the noise and uncover patterns in the data as well as smoothing the data, moving average calculates, where measured as weekly, monthly and yearly value was is shown in Fig. 3. The weekly moving average (K=7) is more looking like as the data itself, and to not lose much of the data pattern the weekly moving average is showing as the most appropriate option.

Fig. 4 shows the decomposing of the time series data which identifying seasonality, trends, data and reminder, where the seasonality was removed to improve the stationary of data.
ACF and PACF shown in Fig. 5 represent the data non-stationary form. The data converted to stationary by changing the differencing as shown in Fig. 6.

Dicky fuller test used to confirm that the data became stationary. Fig. 7 shows the values before changing the differences and Fig. 8 shows the data became stationary with first-difference.

Applying auto ARIMA which showed (4,1,4) as its value, where showing some significant spikes ACF and PACF and it's over the limits as shown in Fig. 9.

Customize ARIMA (1,1,33) showed no significant spikes above the limits of ACFs and PACFs as shown in Fig. 10, which means more accurate model of Auto ARIMA.

Customize ARIMA (1,2,33) tested as well in order to get the best accuracy, as it showed no significant spikes above the limits of ACFs and PACFs as shown in Fig. 11, and not much different from (1,1,33) results.

Forecasting the three models of ARIMA, Auto ARIMA (4,1,4) and ARIMA (1,2,33) showed the same prediction which predicted that stocks will go up while in ARIMA (1,1,33) has different prediction where predicted that the stocks will remain the same as shown in Fig. 12.

Comparing the accuracy results by calculating Mean Absolute Percentage Error MPE showed no much difference between the three models see Fig. 13, where the accuracy of auto ARIMA is 98.88 %, ARIMA (1,1,33) is 99.74% and ARIMA (1,2,33) is 99.75%.

The MAPE calculating showed that the accuracy of forecasting the three models have almost similar value, since Auto ARIMA(4,1,4) showed significant spikes of its ACF and PACF it will not take the results of this model, while the ARIMA (1,1,33) and ARIMA (1,2,33) it has almost the same accuracy, yet it's forecasting to give different values, to reach
the best accuracy further tests have been done by holdout 50% of data and forecasting rest 50% of data and then comparing the result with actual data, where ARIMA (1,1,33) showed better result from ARIMA (1,2,33) as shown in Fig. 14 and Fig. 15.

![Fig. 12. Forecasting Auto ARIMA (4,1,4), ARIMA (1,1,33) and ARIMA (1,2,33).](image)

IV. RESULTS AND DISCUSSION

ARIMA (1,1,33) model showed better accuracy. Although within the measurement of MAPE, the accuracy was 99.74% and ARIMA (1,2,33) was 99.75% which is almost the same. However, owing to its result from holdout test it is considered the best accuracy among the three models. The prediction of Netflix stocks on ARIMA (1,1,33) showed continuity in value, where this prediction to 100 days which mean for next three months, there will be no significant increase in the value of stocks as shown in Fig. 16.

![Fig. 16. The Prediction of Netflix Stocks on ARIMA (1,1,33).](image)

V. CONCLUSION

The research used Netflix stocks historical data for the past five from 7 April 2015 to 7 April 2020 to compare the results of auto ARIMA model and two customize ARIMA (p,D,q) models. After several tests ARIMA (1,1,33) showed accurate results in its calculating values which showed the potential of using ARIMA model on time series data to get accurate prediction on stocks data which will help investors in stocks in their investment decisions. The forecasting of Netflix stocks on ARIMA (1,1,33) showed continuity in value.

This research compared the results and calculated the accuracy based on one model which is ARIMA model the future work will compare more than one model and calculate the accuracy to reach the most accurate one.

REFERENCES


