Netflix Stock Price Prediction Based on Machine Learning

Jiahao Shi *

Tandon School of Engineering, New York University, Brooklyn, New York, 11201, United States

* Corresponding Author Email: js12848@nyu.edu

Abstract. As big data and Artificial Intelligence (AI) are more and more getting into society, people are beginning to ask for predictions in all fields to make prevention. With the technological improvements, the stock market has become more and more predictable as compared to before. The finance field, especially the stock market field, was being put on for making predictions. This paper evaluates the pros and cons, accuracy, and precision when making predictions using the ARIMA model and the Linear Regression model on Netflix stock price trends. The result shows that both models perform moderately. From that, this paper also found that the Linear Regression Model seems to be performing better in the short term and the ARIMA model predicted that it might perform better in the term of long term. This paper proved that these models could play a role in the field of finance but are still restricted to the dataset's size and the application of the model. Hopefully, this paper will find well to finance related members, give and inspire some advice to them.

Keywords: Machine Learning; Stock Market; Stock Price; Linear Regression Model; ARIMA Model.

1. Introduction

The stock market has been a big topic in the field of finance and one company's stock price was largely correlated to their business success and recent achievements. With influence from companies and investors, the stock market price prediction is complicated and presents an intrinsic large volatility market trend.

Market trends were influenced by multiple different aspects, financially and non-financially, which, in respect, company & investors and media, events, etc. It's hard to put everything into one and make a good prediction in such noisy environments. It's also hard to tell whether one variable will be weighted more or less than another variable. Therefore, to be precisely predicted, the stock price can reduce a lot of risk for a company, which is important [1,2,3].

Machine learning is a branch of Artificial Intelligence (AI), that has been widely put in use in the data prediction field. Many researchers have tried to use different machine learning models to predict stock price trend changes. Just as D.P. Gandhmal mentioned Neural Networks, Genetic Algorithm were commonly used for investment decisions. Another technique mentioned was ANN, which is an Artificial Neural Network, that is found to predict the stock market with 29% of works employing ANN [4]. Machine learning relied largely on the dataset provided. This means that the dataset obtained needs to be accurate, large, and detailed. Netflix is one of the largest entertainments in the U.S. Its reputation provided a large non-financial variable that the research wanted to cover so that the prediction method that this paper use could potentially applied to other company with high social influences [5,6].

Therefore, this paper will be researching Netflix's stock price trend. This paper will review the use of machine learning to train models that could predict Netflix stock price trends. The model will be reviewing Linear Regression and the ARIMA model to make predictions on the Netflix Stock Price trend. These two models are the most used and relatively mature methods nowadays people use to predict. By using these two models it can tell the possibility of Netflix stock price trend which could be used by stock market buyers as advice on determining to engage in certain stock markets.

2. Method

2.1. Dataset

This paper obtains data from the Kaggle website, using Netflix's past stock price data from 2018 to 2022. The data included a series of variables including data time and its respective open, close, high, and low price. Open refers to the company's stock opening price in a day, close is the opposite which is the price when the stock market ends. High is the highest stock price in a day and low is the lowest price in a day [7].

Fig. 1 is the basic figure of the Netflix Stock's close price from 2018 to 2022. This is the basis for all future predictions. Date and close price are the main variables this paper's linear regression model covered. This shows a constant like increase across time allowed a linear regression model to take place.

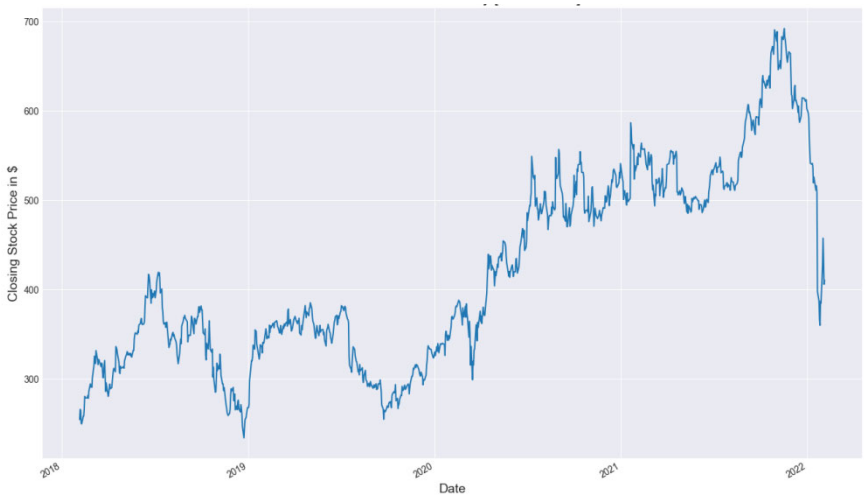

Fig. 1 The history data of close stock price from 2018 to 2022

2.2. Model

2.2.1 Auto Regressive Integrated Moving Average Model

The first model this paper will be talking about is the Auto Regressive Integrated Moving Average model. This model uses time series and observes value to forecast trends. This model outperformed in the aspect of short-term forecasts ARIMA contains three parts, which are AR, I, and MA. AR or Auto Regressive allows the user to find the connection between current value and past value. It used variable past time series values to make predictions. To reach a good AR model first find the difference between past values and current values. This encounters the p, which is the order of the autoregressive component that represents the connection between current values and past values difference. Following MA is the Moving Average, it appears to remove random variants in the prediction. It encounters the q components, which is the order of the moving average component. Last, there is d is the degree of difference (number of times the series is differenced to achieve stationarity) [8,9].

From Fig.2, it can be seen that in the second order the function drops to the shaded, which is known as the confidence level, so set q to 2. show found.

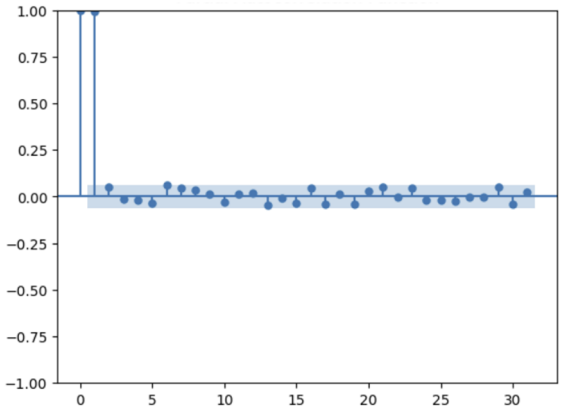

Fig. 2 Partial Autocorrelation Function

Fig.3 will be used to set the p value. Same as how q value was set the order will be counted to the point when points drop to the confidence level. However, in this situation, no points were in the confidence level the largest value was decided to use, so p value is 30.

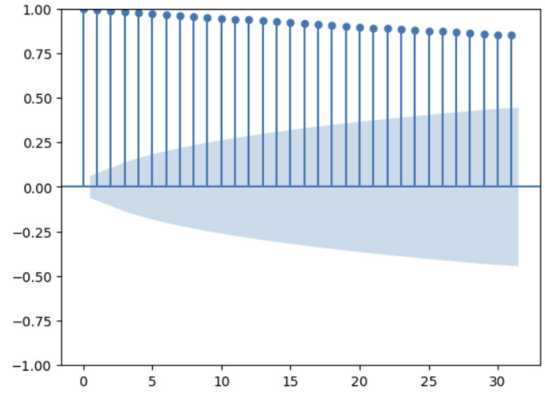

Fig. 3 Autocorrelation Function

2.2.2 Linear Regression Model

The Linear Regression Model was another way to predict the trend of Netflix stock price. The Linear Regression model assumes that relationships will always be linear. However, stock prices were debated as to whether they should be assumed as linear. But the linear regression model contains a simple and multiple linear regression model, which comes with one single independent variable the other allows the model to have multiple independent variables [10,11].

2.3. Evaluation Method

2.3.1 ARIMA Model Evaluation Method

Evaluating an ARIMA Model's fit residual is one way. By checking the residual plot and histogram the basic statistic of the ARIMA Model can be told. A normal Q-Q graph can also be used to evaluate the ARIMA Model. It will indicate residual behavior too.

2.3.2 Linear Regression Model Evaluation Method

To evaluate whether a linear regression model is a good prediction by checking its residual, r squared, is a good method. R-squared will tell you how many precents of y predicted value are accounted for by x values. The larger the R-squared value the better this linear regression model is.

3. Result

3.1. ARIMA Model Result

From the result, it can be shown that the p value was rather large which means that these data are statistically insignificant. This said to be that the model was either not stable or it's just not a very good prediction.

This residual plot shows how the model's difference between prediction and actual data will look (Fig.4). From the graph we can tell that residuals are evenly distributed between (-2,2). This said the ARIMA model performs with some acceptable accuracy.

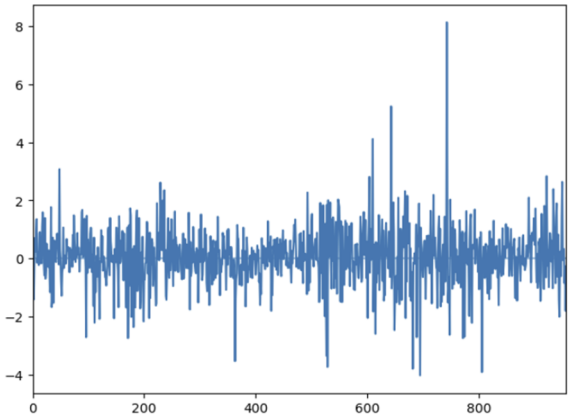

Fig. 4 Standardized residual plot of ARIMA

Fig. 5 is a histogram of residue to show whether the prediction's results will show a normal distribution or not. From the histogram of residue, a normal distribution pattern was shown which proves that the prediction should be performed an acceptable accuracy.

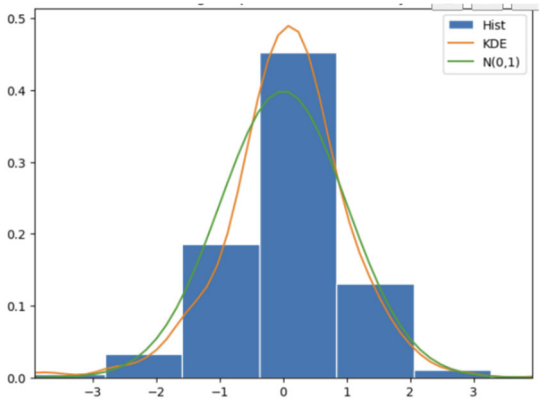

Fig. 5 The histogram of residue

The normal Q-Q graph as shown in Fig.6, was to represent whether the graph performs a linear like pattern or not. From this graph, it is easily shown that there seems to be an exponential increase at the end while having linear in the middle. This makes the prediction sort of unsatisfied that it is not fully normally distributed instead. Its pattern showed that there might be a potential pattern the ARIMA model couldn't include.

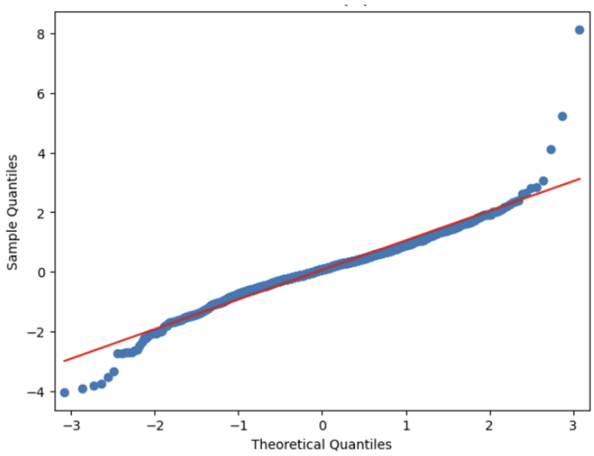

Fig. 6 Normal Q-Q graph

3.2. Linear Regression Model Result

Fig.7 shows the final simple linear regression model for predicting the Netflix stock price. With the x-axis (integer Data) representing each day from 2018-2022, this was to make the equation and prediction clearer and easier while not ruining the data. This line in this graph has an approximate slope of around 0.32 and an intercept of around 259.28. This means that this linear regression model predicted that every day after Feb 5, 2018, the stock price for Netflix will increase by $0.32 starting from $259.28.

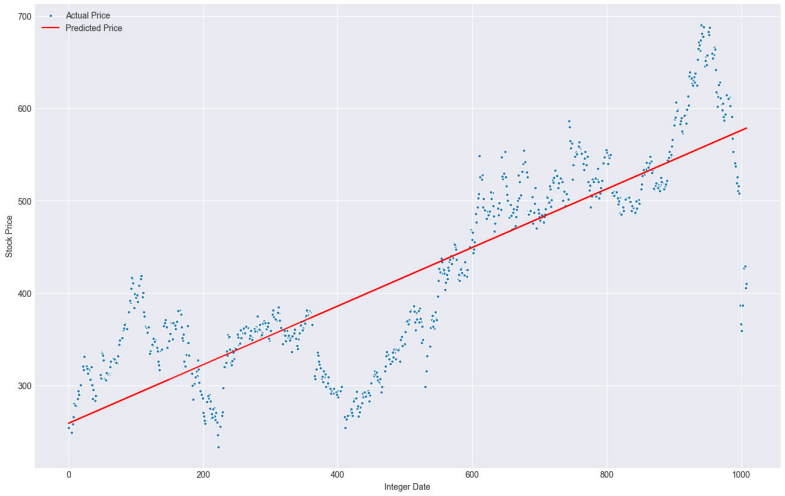

Fig. 7 Linear regression stock price prediction

Fig.8 is about the residual histogram of the linear regression model. It shows a Normal Distribution but slightly skewed to the left which means that the model tends to overvalue stock price. From the histogram, it presented that fifty percent of data varied around 35 dollars. The final R-squared value for this linear regression model is 0.69. This represents that 69 percent of variables can be explained by the independent variables in the regression model or it can be said to be 69 percent accurate. This suggests a moderate level of fit for this linear regression model.

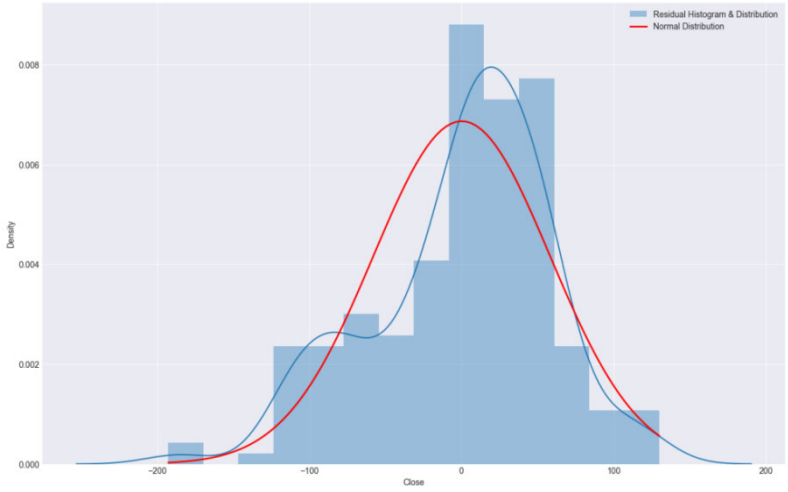

Fig. 8 Residual Histogram for Linear Regression Model

4. Discussion

The normal Q-Q graph showing the ARIMA model indicated there should be some potential problems within. Some reasonable explanations could be that outliers and sample size is too small. Outlier is one of the reasons because the stock price is often influenced by society and some events happening in society might abruptly influence Netflix's stock price. For example, one of the movies or TV show that years goes hot and brings Netflix's social media attention to another level. Another reason could be in this paper the dataset used was from 2018 to 2022 which is a rather small sample compared to that Netflix has participated in NASDAQ which is since 2002. Since stock markets usually show instability during the short term, if the dataset's size used was larger the prediction could be more accurate. Moreover, in statistics larger dataset brings more precision and accuracy.

The linear regression model has an R-squared value of 69 percent which is acceptable but still can be improved. Some potential improvements for the linear regression model could include a larger dataset sample which has the same reason as the ARIMA model. Another reasonable method to improve the linear regression model could be adding some holidays or events days into the linear regression model so that the model could take into consideration the stock price on these certain days and give different weighted values to these stock prices when training the model.

5. Conclusion

Both models have proven that there could be a way to make models that can predict the stock market. While the ARIMA model seems to need more datasets to a hundred percent utilize its power it still shows a good prediction short term. However, the ARIMA model has the characteristic to reach deeper into the data's pattern this paper predicted that ARIMA will perform better in the long term has it not only finds a patter at the first level which is the dataset original, but it also find a pattern that is current value and previous data difference and integrates to find a new pattern which its pattern will be more complete with more data integrate which is the basic of calculus. On the other hand, the Linear Regression model focuses on the surface of the dataset that is more obvious. For such reason linear regression model is predicted to perform better in the short term because in the long run, the stock price prediction using linear regression is kind of useless as most is increase, which you must increase to stay in the stock market for the long term. The Linear Regression model also was predicted to be less accurate as the time series of data is being tested. However, this paper's dataset is rather short term, so the Linear Regression model is still performing well.

It has shown that Machine Learning is ready to be put into more fields and benefit people. In the future, with more time being put into the model and some expertise in the stock market field that could help train the model into a more fitted one, it is promising that there will be an accurate method

to predict the stock market. Other new models could also be found to be more fitted for stock market predictions.

References

[1] M Umer Ghania, M Awaisa and Muhammad Muzammul. Stock Market Prediction Using Machine Learning(ML)Algorithms. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 2019, 8(4): 97–116.

[2] Obthong, Mehtabhorn, Tantisantiwong, Nongnuch, Jeamwatthanachai, Watthanasak and Wills, Gary. A survey on machine learning for stock price prediction: algorithms and techniques. 2nd International Conference on Finance, Economics, Management and IT Business, Vienna House Diplomat Prague, Prague, Czech Republic. 2020: 63-71.

[3] Jiménez-Preciado AL, Venegas-Martínez F, Ramírez-García A. Stock Portfolio Optimization with Competitive Advantages (MOAT): A Machine Learning Approach. Mathematics., 2022, 10(23):4449.

[4] Dattatray P. Gandhmal, K. Kumar. 16(4): 798-805. Systematic analysis and review of stock market prediction techniques. Computer Science Review, 2019, 34(14): 100190.

[5] Osman Hegazy, Omar S. Soliman, Mustaga Abdul Salam. A Machine Learning Model for Stock Market Prediction. International Journal of Computer Science and Telecommunications, 2013, 4(12):17-23.

[6] Ishita Parmar, Navanshu Agarwal, Sherish Saxena, et al. Stock Market Prediction Using Machine Learning. First International Conference on Secure Cyber Computing and Communication. 2018: 574-576.

[7] Mehak Usmani, Syed Hasan Adil, Maran Raza, Syed Saad Azhar Ali. Stock Market Prediction Using Machine Learning Techniques. 3rd International Conference On Computer And Information Sciences. 2016: 322 - 327

[8] Shakir Khan, Hela Alghulaiakh. ARIMA model for accurate time series stocks forecasting. International Journal of Advanced Computer Science and Applications. 2020, 11(7), 524-528.

[9] Irani Hoeronis. Netflix Stock Price Trend Prediction Using Recurrent Neural Network. Journal CoreIT. 2022, 8(2): 97-102.

[10] Verma, Shikha, Arti, Shiwani. Prediction of Netflix Stock Prices using Machine Learning. Learning 1(2), 9-14.

[11] Xu Xinwen. Research on the Stock Price Forecasting of Netflix Based on Linear Regression, Decision Tree, and Gradient Boosting Models. Proceedings of the 2022 2nd International Conference on Business Administration and Data Science. 2022 :1232-1242