# Research on the Netflix Stock Price based on Linear regression Models

*Xiangming* Xie

Education department, SanDiego State University, 92182 SanDiego, United States

**Abstract.** This article uses a linear regression model to predict Netflix's stock from 2018 to 2022, and divides the data set into pre- and post-epidemic analysis. The before-and-after analysis aims to explore the impact of COVID-19 on Netflix stock price. By collecting and collating Netflix's historical stock price data, financial indicators, and other relevant information, a linear regression model was constructed. Through model training and verification, the prediction results of future stock prices are obtained, and a comparative analysis before and after the epidemic is carried out. The research results have important reference value for investors, analysts and decision makers in formulating investment strategies and risk management.

## 1 Introduction

Netflix is a world-leading online streaming service provider, founded in 1997. Through the Internet, Netflix provides its subscribers with a rich content library of various types of movies, TV series, documentaries, variety shows and cartoons. Due to relatively successful policies, Netflix has achieved very good success in recent years, including but not limited to investing in original content, expanding globally, bringing better experience to users through technological innovation, and the impact of the global epidemic. These successful strategies and changes to traditional industries have had a crucial impact on Netflix's success. In addition, the epidemic has also played a certain role in promoting its development. These factors have combined to make Netflix one of the world's leading providers of online streaming services. Studying the reasons for the success of Nextflix is also an important basis for studying the development of Internet companies.

Since the invention of stocks, people's interest in stocks has never diminished. People buy stocks with the goal of capital appreciation through stock price appreciation or return on investment through dividends. Therefore, paying attention to stock prices can help people judge their investment performance and wealth growth. Moreover, the stock market is generally regarded as a barometer of the economy, and the rise and fall of stock prices can reflect the trend of the economy and industry. People's attention to stock prices can help them understand the macroeconomic conditions and industry developments, so as to make relevant decisions. And stock price fluctuations provide investment opportunities and trading strategies. Investors and traders make investment decisions and trading strategies based on changes in stock prices in order to obtain investment returns or realize short-term trading profits. Therefore, attention to stock prices can help people make more informed investment decisions.

Investors have never stopped predicting stocks, and stock price predictions can help investors make more informed investment decisions. By predicting stock prices, investors can assess the potential returns and risks of a stock and decide whether to buy, hold or sell a stock. Accurate stock price forecasts help investors choose stocks that suit their investment goals and risk tolerance. Prediction of stock prices helps investors in risk management. By predicting the volatility and potential risks of stock prices, investors can adopt appropriate risk management strategies, such as setting stop loss points, diversifying investment portfolios, using derivative instruments, etc., to reduce potential losses. Predictions on stock prices can help traders make trading decisions. Short-term traders often rely on the volatility and trends of stock prices for short-term trading profits. Accurate stock price forecasts can guide traders in determining when to buy and sell, develop trading strategies, and manage trading risk. The prediction of stock price plays an important role in investment decision-making, risk management, trading decision-making, asset allocation and economic analysis, and can guide investors and decision makers to make more informed decisions and manage risks.

Using machine learning to predict stock prices is a method that everyone is using. The advantage of this method is that machine learning algorithms can process large-scale historical stock market data, including prices, trading volume, financial indicators, etc. By analyzing large volumes of data, machine learning can uncover underlying patterns and trends, thereby providing more accurate predictions. Machine learning models can automatically process data and make predictions, reducing the intervention of manual processing and subjective judgments. This increases the efficiency of forecasting and reduces human bias in the forecasting process. In addition, machine learning models can handle the computational demands of large-scale data, resulting in faster predictions. More and more excellent methods and architectures have been invented. Like

Corresponding author: xxie6645@sdsu.edu
© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

CNN-BiLSTM-AM. 'This method is composed of convolutional neural networks (CNN), bi-directional long short-term Memory (BiLSTM), and attention mechanism (AM).' [1]. In another article, the ARIMA method is used, and the ARIMA model is relatively simple compared to other complex time series models. It is based on the parameters of the autoregressive and moving average parts of the time series to model and does not require much domain knowledge and complex calculation processes. This makes the ARIMA model highly operable in practice. The ARIMA model performs well and has high forecasting accuracy when dealing with some time series data with obvious trends and seasonal patterns. [2]. This article compares ARIMA and artificial neural networks models and shows which model is more suitable for the current data based on the results. 'The findings further resolve and clarify contradictory opinions reported in literature over the superiority of neural networks and ARIMA model and vice versa.' [3]. This article uses a lot of different structures and methods to make predictions and give optimal solutions. The article points out that some existing linear and nonlinear algorithms focus on using daily closing prices to predict the likely direction of a company's stock price. [4]. The authors in this article use an innovative structure proposed by themselves. The article points out that they use structural support vector machines to classify more complex inputs differently, and use graph structures to connect companies in the information technology field. They used structural support vector machines to predict any movement, positive or negative, that these companies might have. [5]. Artificial networks and random forest techniques are also compared. In this study, the researchers used random forests and artificial neural networks for comparison. They used both techniques to predict stock prices and trends in different business units in five companies. [6]. Linear regression is widely used as a relatively simple and efficient method. These researchers used linear regression methods to build and predict models. They experimented with the model and used previous stock prices and other stock characteristics and made predictions from them. The resulting coefficients are 0.9989 and 0.9983, respectively. [7]. As a relatively simple and easy-to-use method, linear regression is widely used as an experimental method to predict various aspects. 'In this study, the researchers used linear regression to predict the behavior of the TCS dataset. They believe that the linear regression method is the most suitable method, it can give better results and shareholders can use this method to invest confidentially. [8]. As a more excellent improvement method, IMLR is also one of the widely cited methods. In this study, multiple linear regression (IMLR) based on a modified linear regression method was used. This method is embedded in the mobile application. And it is applied to forecasting stock results. Multiple linear regression is a hybrid multiple linear regression that uses a moving average technique. [9]. The clustering and regression are also two models that are often compared. 'The clustering and regression are the two techniques of data mining used here, Validation index is used for analyzing the performance of different clustering methods such as partitioning technique, hierarchical technique, model based technique and density based technique.' [10].

This article mainly uses Linear Regression to predict the stock price of Netflix from 2018 to 2022. And divide the data into two sub-datasets before the epidemic and after the epidemic to study the impact of the epidemic on Internet companies such as Netflix, and draw conclusions.

Researching Netflix's stock price can help investors make more informed investment decisions. Understanding Netflix stock price trends, volatility, and valuation can help investors evaluate whether to buy, hold, or sell Netflix stock. This is very important for individual investors, institutional investors and fund managers. As a leading online streaming media company, Netflix's stock price fluctuations can often reflect the development trend and market sentiment of the entire streaming media industry. By studying the stock price of Netflix, we can understand the competition situation of the company's industry, changes in market share and market expectations for the company's future development. This has important reference value for industry analysts, research institutions and investment banks. As a company operating on a global scale, Netflix's stock price changes can also reflect the macroeconomic and market conditions. Researching Netflix's stock price can provide information on economic trends, consumer behavior and market demand, which is of reference value for macroeconomic analysis, industry research and policy formulation.

## 2 Data and Method

### 2.1 Data

#### *2.1.1 Sample and Descriptive Statistics*

The data set for this study is the stock price of Netflix from February 5, 2018 to February 5, 2022, collected from the Kaggle website. The data set contains 1009 rows and 7 columns. The data in the data set is fully available and there is no invalid data. And the dataset contains Date, Open Price, Close Price, Highest Price, Lowest Price, Adj Price and Volume. The data set contains 1009 rows and 7 columns. The data in the data set is fully available and there is no invalid data. After statistical analysis of the results, it is found that every data is available, and the averages of open, high, low and close are not much different. The overall distribution of the data is relatively average. The volume of the data is not very big. There is no big difference in other parts, and the overall distribution is relatively average. This dataset is very useful and contains all the required data. Therefore, this study was carried out based on this data set. In fact, investors are very interested in the field of stock price forecasting research. To invest well and successfully, many investors want to know what the future holds for them in the stock market. A good and effective stock market forecasting system helps traders,

investors and analysts by providing additional information. This time, Linear Regression is used to predict the stock price. The data set contains 1009 rows and 7 columns. The data in the data set is fully available and there is no invalid data.

In order to study the impact of the new coronavirus on stock price prediction, this experiment divides this data set into two sub-data sets, one data set from February 5, 2018 to January 1, 2020, and one data set in 2020 From January 2 to February 5, 2022, it is divided into another data set, which is predicted separately to obtain comparative data and further analyze the impact of the new coronavirus on the experimental results (table 1).

**Table 1** Descriptive Statistics

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| open | 1009 | 419.06 | 108.54 | 233.92 | 331.49 | 377.77 | 509.13 | 692.35 |
| high | 1009 | 425.32 | 109.26 | 250.65 | 336.3 | 383.01 | 515.63 | 700.99 |
| low | 1009 | 412.37 | 107.55 | 231.23 | 326 | 370.88 | 502.53 | 686.09 |
| close | 1009 | 419 | 108.29 | 233.88 | 331.62 | 378.67 | 509.08 | 691.69 |
| Adj close | 1009 | 419 | 108.29 | 233.88 | 331.62 | 378.67 | 509.08 | 691.69 |
| volume | 1009 | 757.07 | 546.55 | 114.4 | 409.19 | 593.45 | 932.24 | 589.04 |

### *2.1.2 Data Visualization*

Visualizing close stock prices is also an effective way to help analyze data better. Use the style() method to draw the close stock price to get the result as show in figure1.

As can be seen from Figure 1, the stock price of Netflix has fluctuated slightly from February 2018 to mid-May 2020, but the overall price is relatively stable. However, there was a substantial increase from May 2020 to October 2021, and a downturn began in October 2021.

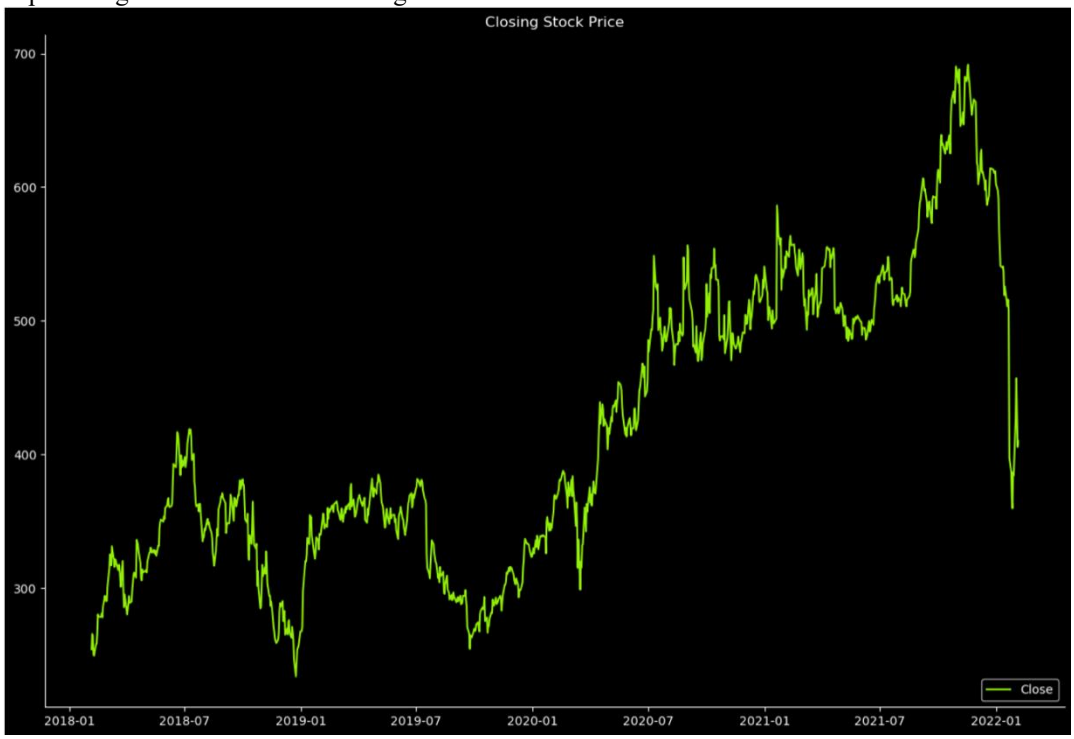


**Figure 1.** Closing Stock Price of Netflix from 2018 to 2022

## 2.2 Method

After the data set is fully analyzed, the data set can be used for data preparation. In order to save the original data, all forecast data is based on the replicated data. The advantage of this is to save the original data so that the original data will not be lost when the experiment is repeated or the experiment goes wrong. Three different sets of data, that is, all data, data before the epidemic, and data after the epidemic were copied and brought into linear regression. Linear regression is a statistical analysis method used to establish and predict linear relationships between variables. Linear regression is a method of finding the best parameter value by minimizing the difference between the actual value and the predicted value, and the method of least squares is usually used to determine the best predicted result of the parameter value. The goal of the least squares method is to minimize the sum of the squares of the prediction

errors of all training samples, that is, to minimize the loss function: $L(\beta_0, \beta_1, ..., \beta_n) = \Sigma(y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_n x_{ni}))^2$. Among them, $\Sigma$ represents the summation operation, and i represents the index of the training sample.

Linear regression models can be used to explore the influence and bias caused by independent variables on dependent variables and can also be used to predict the expected results of target variables. The precondition for using linear regression is that there must be a reasonable linear relationship between the dependent variable and the independent variable, and the error term between them must satisfy some basic assumptions. If the error obeys the normal distribution of independent and identical distribution, the error has the same variance, etc. In addition, linear regression can also be extended, such as multiple linear regression, ridge regression, lasso regression, etc., to deal with more complex problems and data. Linear regression has some key ideas and assumptions, namely the linear relationship assumption: Linear regression assumes a linear relationship between the independent and dependent variables. This means that there is a fixed proportional relationship between the change of the independent variable and the change of the dependent variable, which can be described by a linear equation. Minimize error assumptions: Linear regression determines the best-fit line by minimizing the sum of squares of residuals between the predicted and observed values. The residuals are the difference between the predicted and observed values, and by minimizing the sum of squares of the residuals, we can find the regression coefficients that minimize the forecast error. Model Fitting and Evaluation: Linear Regression uses the training dataset for model fitting, i.e. calculating the values of the regression coefficients according to the method of least squares. After the fitting is completed, the evaluation index (is used to evaluate the fitting degree and performance of the model. In general, the logic of linear regression is to find the best Fitting lines for prediction, inference and analysis.

## 3 Results

### 3.1 full sample

After processing the dataset using linear regression analysis, I got the following result: Mean Square Error:15.204. Round Mean Square Error:3.899. Mean Absolute Error:2.796. Mean Absolute Percentage Error:0.007. R2 Score:0.999. Overall, the predicted score is high.

**Table 2** The results of predicting Close Stock Price using Linear Regression

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| close | 202.0 | 423.027326 | 102.426464 | 253.669998 | 340.849999 | 379.509995 | 508.972504 | 688.289978 |
| close_prediction | 202.0 | 422.967075 | 102.404707 | 246.863561 | 340.680293 | 379.127532 | 510.353683 | 687.226145 |

It can be seen that the predicted results (table 2) are relatively close to the predicted results of the data set itself, and are generally at a relatively excellent level. In general, because the data set is very rich and highly available, and the data content is very comprehensive. And the algorithm of linear regression is excellent and very suitable for this experiment, so the overall result is ideal.

And COVID-19 has had a broad and far-reaching impact on stock prices. This impact can manifest in different patterns and degrees in different aspects and over time periods. Due to the particularity of the new crown epidemic, stock prices have generally experienced sharp declines and fluctuations due to heightened uncertainty and panic. Stock prices are affected by investor sentiment, and market panic often leads to sharp falls and violent fluctuations in the stock market. However, different industries and companies are affected to different degrees by the epidemic. For example, the tourism industry or airline companies or hotel chains have received a severe impact, and stocks have fallen sharply. However, stocks in businesses such as healthcare and online commerce have seen big gains. With the development and release of the new crown vaccine, the market generally has hope for the control of the epidemic and economic recovery. Good or bad vaccine progress and improved vaccination coverage could have a positive impact on the stock market.

### 3.2 Pre-covid19 sample

The price of a stock is also affected by many factors. Fundamental factors such as a company's profitability, growth potential, and financial health are important drivers of changes in stock prices. The macroeconomic environment also has an important impact on the stock market. For example, the country's economic growth rate, inflation level, interest rate changes, monetary policy and other macroeconomic factors will directly or indirectly affect the profitability of enterprises and the risk preference of market investors, thereby affecting stock prices. And external events and natural disasters: Sudden external events and natural disasters may have a major impact on the market and stock prices. Therefore, this experiment separates the data sets and conducts different experiments to find out how the epidemic since the beginning of 2020 has affected Netflix's stock price.

For this experiment, in order for the experiment to accurately study the impact of COVID-19 on stock prices, the only variable must be set to time. Therefore, this experiment uses exactly the same method and test

model, and divides the data set into pre-epidemic and post-epidemic, and treats the two sub-data sets in exactly the same way. Get the experimental results and analyze them (table 3).

**Table 3.** The results of predicting Pre-covid19 Close Stock Price using Linear Regression

| | count | mean | std | min | 25% | 50% | 75% | 100% |
|---|---|---|---|---|---|---|---|---|
| close | 96.0 | 326.431771 | 44.726287 | 233.880005 | 288.645005 | 326.645005 | 362.895005 | 418.970001 |
| close_prediction | 96.0 | 326.608374 | 44.280695 | 242.725420 | 289.319349 | 325.591091 | 364.793975 | 415.031291 |

### 3.3 Post-covid19 sample

Splitting the rest of the sample data and processing it likewise yielded little change for the forecast. In this way, we can conclude that the sample data has not been divided according to before and after the epidemic and processed separately to produce any changes in the prediction results (table 4).

**Table 4.** The results of predicting Post-covid19 Close Stock Price using Linear Regression

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| close | 106.0 | 490.996227 | 82.178713 | 326.000000 | 434.992508 | 499.565002 | 538.929977 | 677.719971 |
| close_prediction | 106.0 | 490.482932 | 83.040337 | 317.096183 | 433.924674 | 500.238247 | 540.112348 | 679.333658 |

### 3.4 Comparsion

It can be seen from the results that there is no significant difference in the accuracy of Netflix's use of linear regression to predict the stock price from 2018 to 2022 before and after the epidemic. According to the analysis, the main reason for the results is diversity, and the stock price is affected by many factors, including company fundamentals, industry trends, market sentiment, competitive environment, etc. The epidemic is only one of the factors, and other factors may play a more important role in the forecast, thus masking the impact of the epidemic on the stock price forecast. The factors affecting the epidemic are very complex, including the time of outbreak and control, the progress of vaccine research and development, and government intervention measures. The changes and interactions of these factors may be difficult to predict accurately, so the impact of epidemic factors may be concealed or not obvious when predicting stock prices. Financial markets typically react quickly to and digest the impact of major events, including outbreaks. Therefore, the market may have made some adjustments and responses to the epidemic, so that the impact of the epidemic on stock prices has been digested by the market to a certain extent or reflected in the current stock price. The most important reason for analyzing the market is that Netflix is an online streaming media company whose main business is to provide online video services. During the epidemic, due to home isolation and other policies, everyone had nothing to do. The online video services provided are gradually valued, and due to more users, the market is more optimistic about online streaming services. This in turn makes Netflix's stock price more predictable and higher.

## 4 Conclusion

In general, using Linear Regression to predict Netflix's stock price from 2018 to 2022, and dividing this data set into two sub-data sets for prediction, the results are very ideal. Facts have proved that the impact of the epidemic on stock prices will not cause a large deviation in the model that uses Linear Regression for prediction. The overall results of the experiment are ideal. And Linear Regression, as a simple and easy-to-use old-fashioned model, has shown its excellent performance in predicting stock results. It is a very useful mathematical model. This article proves that Linear Regression is very suitable for predicting the stock market. And this study reveals the stock price of Netflix in recent years, and thus has a better evaluation and prediction direction for the overall development level of large Internet companies in recent years. However, this study uses a single method to predict the stock market, and it may be because of this that the academic results of this study cannot be used to make a good comparison with other methods and models.

## References

1. W. Lu , J. Li., J. Wang. et al. Neural Comput & Applic, **33**, (2021)
2. A. A. Ariyo, A. O. Adewumi and C. K. Ayo, *Stock Price Prediction Using the ARIMA Model*, 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, (2014)
3. Adebiyi, A. Ariyo, et al. Journal of Applied Mathematics, **2014**, (2014)

4. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, *Stock price prediction using LSTM, RNN and CNN-sliding window model*, 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, (2017)
5. L. Carson, et al. *A Machine Learning Approach for Stock Price Prediction*, ACM International Conference Proceeding Series, ACM, (2014)
6. V. Mehar, et al. Procedia Computer Science, **167**, (2020)
7. Y. E. Cakra and B. Distiawan Trisedya, *Stock price prediction using linear regression based on sentiment analysis*, 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, (2015)
8. D. Bhuriya, G. Kaushal, A. Sharma and U. Singh, *Stock market predication using a linear regression*, 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, (2017)
9. A. Izzah, Y. A. Sari, R. Widyastuti and T. A. Cinderatama, *Mobile app for stock prediction using Improved Multiple Linear Regression*, 2017 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, (2017)
10. B. S.Bini, and T. Mathew. Procedia Technology, **24**, (2016)