Prediction of Apple Stock Price - Comparing the Linear and Nonlinear Regression

Jin Guo*

Shenzhen Academy of International Education, Shenzhen, 518000, China

*Corresponding author: guojin@mails.x-institute.edu.cn

Abstract. Predicting future movements of a stock could be usually better performed within multiple models of analysis. However, different methods of prediction will show completely varied results even in case of a same stock. Due to varied logics and progression process, each model will return different results. These results could have implied certain aspects or properties that the stock owns, and they would reflect keys to acknowledge analyzers and decision makers to in practical. It is imperative for analyzers or decision makers to compare results from different models and algorithms. This paper performs a comparative statement of how would different prediction models or algorithms differentiate in results of predicting a stock price. Furthermore, the researcher will include a comparation of applying separately linear regression and Nonlinear Regression, to analysis on the stock Apple (NASDAQ: AAPL) and compare their prediction outcomes. In addition, the researcher will evaluate the performance according to RMSE (Root-Mean-Square-Deviation) results.

Keywords: Prediction; stock price; linear regression; nonlinear regression.

1. Introduction

In a subfield of artificial intelligence known as machine learning, algorithms are used to simulate how human analysts act and make choices. The progress of machine learning took a turn in the 1990s, with the emergence of the internet and increased availability of useable data, the field began to change from a knowledge-driven to a data-driven approach, paving the way for the machine learning models we see today [1]. Machine learning has evolved into a key component of artificial intelligence after decades of study, and some early applications have demonstrated its maturity in handling a range of human jobs that demand a large amount of computation. Machine learning software is fundamentally different from traditional computer programs, which rely on rules in the form of code created by human specialists to tell computers how to carry out tasks. In other words, traditional programs are designed to execute specified tasks solely [2].

This technology was adopted into human affairs right away. The volume of financial activity has increased, and their shifting tendency has also grown more complicated, as a result of the recent fast economic boom. Consequently, the model formulation of the current study adds to a full knowledge of the variables impacting financial sectors, utilizes the advantages of machine learning techniques [3]. Asset price trend forecasting is a well-known and intriguing topic that has drawn the attention of many scholars from several disciplines. themes for academic and financial study to comprehend stock market dynamics and forecast development and change for many years, intelligent stock selection has been a key component of portfolio development for investors and fund managers [4].

However, the outcomes of the data analysis process do vary and are controversy due to the employment of various methods. Researchers have proposed many approaches for technical behavior of stock prices in order to discover trends in price movement that would advantage investors [5]. These variations are inevitable in actual situations or tests because of variations in the fundamental logics of algorithms. Regression analysis is used in statistics to investigate the connection between an independent and dependent variable [6]. Additionally, basic logical inconsistencies and algorithmic mathematical terminology lead to variances. Nonlinear Regression concentrated on identifying more precise connections between data time phases. The non-linear regression model uses previous stock data to forecast future stock values [7].

One of the most established analytical techniques is technical analysis, which makes use of previous stock prices and trade volumes to forecast the trend of future stock prices. Because this

method merely considers past pricing data and volume, this study, which is built on demand and availability in financial system, may potentially be used to enterprises in financial distress [8]. It is crucial for analysts and investors to undertake stock research that evaluates how various prediction techniques will provide distinctive outcomes.

Although stock price fluctuations are fixed, forecasts of future prices may still change owing to variations in forecasting techniques. Time impacts can have a lot of variance and unpredictability. Any value of a stock can hardly be referred to be noise, because it is the outcome of logical decisions made by agents with variable degrees of information, cautiousness, and functionalities. This means that quantitative studies cannot average out interruptions, and an exceptional case cannot be repaid by an overall average, as has been completed in industrial operation and management, since these data, even if unprecedented, are derived from actual actions and may be relevant to a stock's price evolution [9]. Additionally, it is feasible for the researcher to track the operation of each algorithm and determine the cause of this divergence.

For analysts and decision-makers, identifying distinctions between various algorithms is a necessary and inevitable effort. In order to maximize efficacy, analyzers in the future could be able to use several algorithms in a variety of situations. Multiple regression analysis enables researchers to analyze the strength of the link between a dependent variable and numerous predictor variables, as well as the contribution of each predictor to the relationship, frequently with the impact of other predictors statistically excluded [10].

2. Methodology

2.1 Methods

In order to anticipate the stock price of Apple (NASDAQ: APPL) by applying independently the linear regression and Nonlinear Regression approaches, and to examine differences between expected outcomes, and a comparison of particular findings done by presenting actual examples will be done. In order to identify distinctions that have produced distinct outcomes from the standpoint of algorithm performances and structures, it is also required to trace the logics of differentiated algorithms.

The researcher will use a dataset, containing historic prices of Apple stock from NASDAQ (https://www.nasdaq.com/market-activity/stocks/aapl/historical), to separately set programs with linear regression and Nonlinear Regression in the complier of Python. Predicted results will be presented as the form of graph, and will be further evaluated by RMSE (Root-Mean-Square-Deviation) results, and this process will also be granted from a Python algorithm.

2.2 Dataset Applied

The researcher has used a dataset of historical NASDAQ pricing for Apple stock. The dataset contains seven columns of information, separately date, open price, highest price, lowest price, close price, adjusted closing price, and volume of Apple stock. Date was ranged all the way from December 31th, 2021, to December 30, 2022.

2.3 Demonstration Section

This essay presents a comparison of specific results done from presenting actual instances, to firstly predict stock price of Apple (NASDAQ: APPL) by using separately linear regression and Nonlinear Regression methods, and compare differences among predicted outcomes. Furthermore, the researcher will trace the logics of differentiated algorithms, in order to find out differences that has led to distinct results from the perspective of algorithm performances and structures.

2.3.1 Graph Usages

Graphs are important in terms of illustrating and giving clear images of datasets such as stock prices. In this essay, the researcher will adapt scatter plots to illustrate both the dataset of Apple stock

price, and the trend analyzed and predicted by either (Linear Regression and Nonlinear Regression) algorithms.

The researcher has used the python library matplotlib. pyplot() to conduct graphs. Data extracted from a dataset will be plotted on a diagram with X and Y axis separately labeled with certain data types and corresponding values. Next page provides an example of table and graph formats, as table 1 shows.

Table 1. Sample Dataset

Date	Close Price	Data	Close Price
20220104	197.70	20220113	172.19
20220105	174.92	20220114	173.07
20220106	172.00	20220118	169.80
20220107	172.70	20220119	166.230
20220111	172.19	20220120	134.501
29229112	175.53		

The data will be plotted as blue dots while a red line represents the trend, as figure 1 shows:

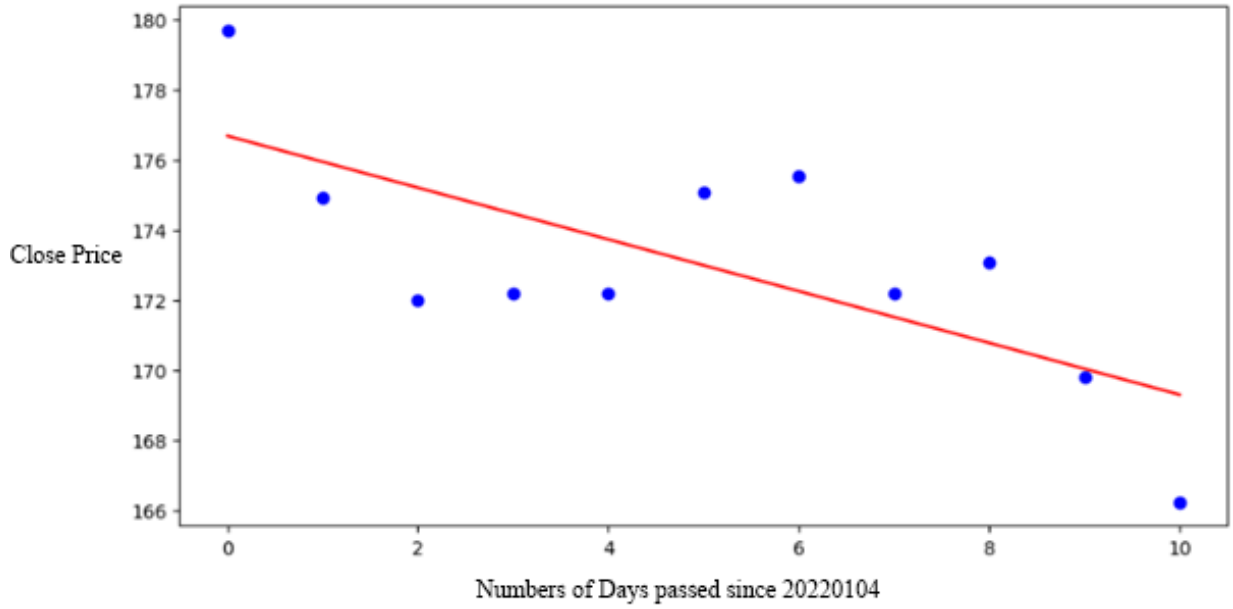


Figure 1. Sample Graph

3. Comparasion of Linear and Nonlinear Regression

3.1 Linking Machine Learning with the Field of Finance

The field of finance have been explored during the recent centuries, with the latest technologies being adapted and utilized to make profit during any previous time. Since 1966, the NYSE has used computers to automate trading. After decade, the 2 to 6 billion in trading volume that flows through the NYSE each day is now handled entirely by computers. High-frequency electronic deals, which may be completed in as little as 10 milliseconds, account for more than 60% of all trading that takes place on the exchange today.

Adaptation of computer technology of the modern stock market is undoubtably noteworthy, and as computer science developed to the point of 2000's, machine learning has evolved and developed to be more mature. It was to be foreseeable that machine learning is going to be largely applied into the field of finance, moreover, to stock trading. As the aspects of finance is largely suitable for machine learning programs. The core of machine learning is statistics, photos, or text.

Bank transactions, photographs of persons or even particular bakery items, service information, time series data, or revenue numbers are all examples of data [11]. Machine learning has brought changes to the industry as well. Ability to handle large amount of data at almost immediate time and to identify information or signals was inherited well from Artificial Intelligence and the field of big data. Especially in Banking industry, machine learning technologies help banks to transform their data sources into meaningful insights across the board, from operations to corporate development and marketing. Firms often apply deep learning utilize applications in banking to deliver faster service, better enhanced security, and cleaner, more efficient operations.

3.2 Importance of Identifying Differences Among Algorithms

If we perform our first data analysis, we may choose the best model to use in each specific situation. Data Scientists or AI researchers have never come up with a singular solution or a universally dominating model. In addition, linear regression, to be specific, does not appear to be like multiple regressions in case that linear regression takes only one input variable as dependent, while the latter has two or more independent variables. Differences between linear and non-linear methods can also lead to greatly varied outcomes, as figure 2 shows.

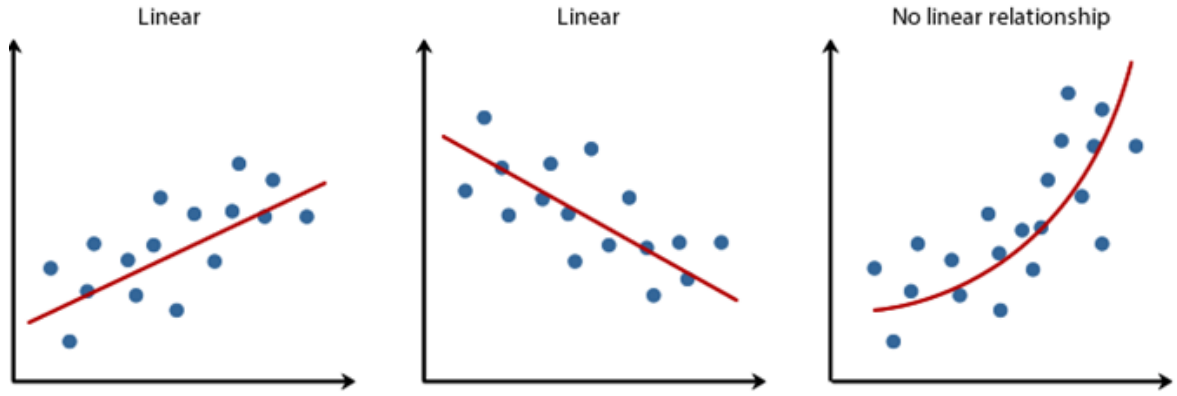


Figure 2. Graphing Trends Illustrated by Different Algorithms [11]

3.3 Results of Linear Regression

Outcome of Linear Regression prediction appears to be inaccurate. RMSE (Root Mean Square Error) in the case of Linear Regression reached 9.846. And the predicted trend is graphed below, as figure 3 shows:

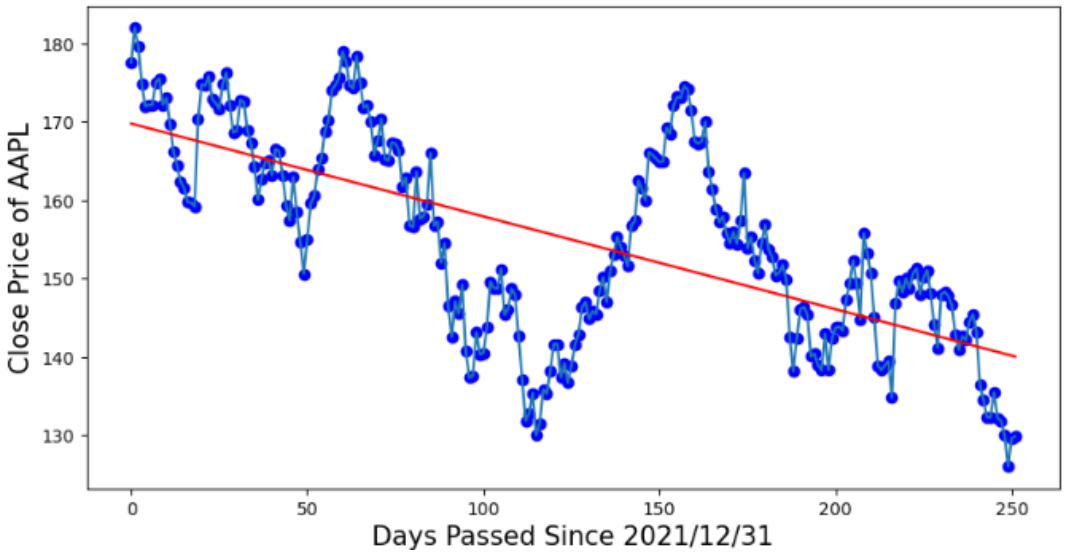


Figure 3. Results of linear regression

Linear regression is basically an extension of simple regression. From the results conducted previously, the researcher notices that linear regression basically conduct its predictions based on following a linear data model of one independent variable. After building an equation, with slope and a parameter determined, the dependent variable will be calculated by repeating adding up the constant with parameter times the independent variable squared. Furthermore, there will be a linear relationship between inputted variable (the X axis) and the dependent variable (the Y axis). Also, since linear regression constructs its models only through learning data input before, it is considered a supervised machine learning algorithm. According to this process, smaller curves that might have significant impacts on the results might be ignored. Linear regressions are apparently useful for predicting continuous variables, and not able to perform classifying functions.

3.4 Results of Nonlinear Regression

Prediction using linear regression appears to be inaccurate. In the example of Nonlinear Regression, the RMSE value has performed better comparingly. The RMSE value calculated was 1.873. And the graph below shows the anticipated trend:

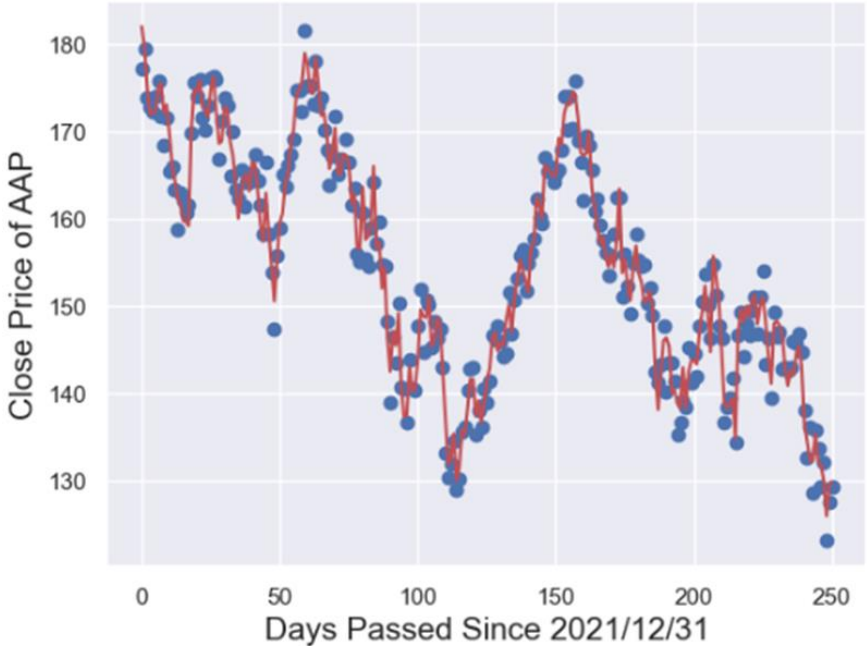


Figure 4. Results of Performing Nonlinear Regression

Nonlinear regression relates X and Y values by a curved nonlinear model. As shown in the graph, nonlinear regression is more conservative in applying data to build its model. Curves frequently appears when Y data varied from one to another, brings flexibility to the nonlinear regression. The predicted and model trend appeared to be close to the plotted original dataset. Furthermore, as the dependent and independent variables' changes are included by the nonlinear regression as its slope, nonlinear regression will illustrate an accurate model for nonlinear regression. In predicting stock prices, nonlinear regression can be utilized since change in time and data do not share a linear relationship.

3.5 Comparison

After comparing to the graphs, the researcher concludes that linear regression does not appear to be comparable to multiple regressions in the sense that linear regression accepts just one input variable as dependent, whereas the latter involves two or more independent variables. Differences between linear and non-linear approaches can also provide widely disparate results. Nonlinear regression inclined to be more conservative and closer to the original data, despite possible cases of

noises were considered. Whereas Linear Regression is more confined to the original data, and has only given an approximate response of its prediction. Linear regression is relatively limited by the shape of original data and appear to be not fitting well with long time-based data. Considering the RMSE results, nonlinear regression has shown to be more accurate in terms of predicting stock price of a long-time span.

4. Conclusion

This essay makes a comparison remark about how different prediction models or algorithms perform when projecting a stock price. The researcher investigates how each method operates and determine how this differentiation happened. Finding discrepancies amongst algorithms is an unavoidable and necessary effort for analysts and decision makers. The researcher compares the results of using linear regression and nonlinear regression independently to analyze the stock Apple and compare their predicted outcomes. The researcher also assesses performance using RMSE (Root-Mean-Square-Deviation) findings. To conduct the results, the researcher applied both algorithms into separate programs written in programming language Python, and provided graphs along with RMSE results being carried by programs.

References

[1] George Lawton. Deep Learning's Role in the Evolution of Machine Learning. TechTarget, 2020.

[2] Onder Ozgur, Erdal Tanas Karagol, Fatih Cemil Ozbugday. Machine Learning Approach to Drivers of Bank Lending Evidence from an Emerging Economy. Springer Open, 2021.

[3] Kodelja, Zdenko. Is Machine Learning Real Learning? Eric.ed.gov, 2019.

[4] Padhi D K, Padhy N, Bhoi A K, Shafi J, Yesuf S H. An Intelligent Fusion Model with Portfolio Selection and Machine Learning for Stock Market Prediction. Research Gate, 2022.

[5] Sidra Mehtab, Jaydip Sen, Abhishek Dutta. Stock Price prediction Using Machine Learning and LSTM – Based Deep Learning Models. OpenAIRE, 2020.

[6] Alvin Nicolas. Linear Regression in R. Researchprospect, 2021.

[7] Nazar M. Al-shaiby. Using the Nonlinear Regression to Predict the Prices of Stocks and its Impact on the Investor Decisions in the Financial Market. Research Gate, 2020.

[8] Pyo S, Lee J, Cha M, Jang H. Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. National Library of Medicine, 2017.

[9] Nils Karlsson. Comparison of Linear Regression and Neural Networks for Stock Price Prediction. Uppsala Universitet, 2020.

[10] Kazuhisa Takemura. Decision Strategies and Bad Group Decision-Making: a Group Meeting Experiment. Escaping from Bad Decisions, 2021.

[11] Wayne Duggan. This Day in Market History: NYSE Gets Computerized. Benzinga Website, 2022.