Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/15/20230840

Chinese stock price prediction under COVID-19 period based on linear Regression Model

Yijia Zhang[1, †], Yiran Yang[2, 3, †]

[1]The department of artificial intelligence, Beijing university of chemical technology, Beijing, 102202, China
[2]The department of computer science, Xiamen university of technology, Xiamen, 361000, China

[3]2010716438@s.xmut.edu.cn
[†]These authors contributed equally.

Abstract. To anticipate and assess the price trend of scientific and technical stocks in a time of high volatility following the breakdown of China's COVID-19 Epidemic Economic Policy at the end of 2022, we plan to utilize a training linear regression model in this work. assisting people and businesses in making better analyses and decisions during this time of high risk, thereby lowering investment risks. This paper uses the stock price data of Alibaba, Tencent, and Xiaomi in Quandl during the new crown epidemic to represent the general trend of Chinese technology stock prices. We preprocessed the data to retain features that better reflect the characteristics of the data and remove features that have less impact on the data analysis. This study chose a linear regression model to model various relationships through the classifier used by Scikit-Learn for regression and estimated the unknown parameters in the linear regression model from the data. The data from the training and test sets are used to train the linear regression model, and the results are visualized as graphs, which can more intuitively convey the overall trend and local volatility of stock prices. In this study, the accuracy of the model can reach more than 85%. The visualized figure shows that although stock values recovered quickly when the regulation was lifted, they are still lower than they were before the outbreak due to the impact of China's past epidemic policies on their long-term fall. The experimental findings demonstrate the great accuracy with which the advanced regression model in this work can forecast the price trend of technology stocks throughout the period of high volatility following the policy's unsealing and the extent to which it may represent price volatility.

Keywords: linear regression, COVID-19, stock prediction, machine learning.

1. Introduction

Stocks represent a unit of ownership in a company and can be acquired through the sale of shares in exchange for monetary compensation. Investors purchase stock in anticipation of a gain in price that will allow them to profit when they sell it in the future [1]. In light of COVID-19, China has implemented a series of restrictive policies which have had significant impacts on the country's trade activities. At the beginning of the COVID-19 epidemic, the growth rate of China's foreign trade

© 2024 The Authors. 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/).

DOI: 10.54254/2755-2721/15/20230840

dropped sharply over the same period, and both the supply side and the demand side of China's trade have been seriously impacted [2]. As a result, the prices of numerous stocks have been adversely affected.

In 2022, China's COVID-19lift restrictions, and the prices of many stocks will fluctuate significantly. For example, removing restrictions on entry and exit and foreign trade. As such, it is imperative to accurately predict high-volatility stocks, which will aid individuals, companies, and governments in making informed investment decisions and mitigating risks. The development of the science and technology industry is changing with each passing day, and the competition is fierce. It is a high-growth industry, and its growth rate is far greater than that of general industries. Therefore, this paper selects technology stocks as the data source of the model to predict the price trend of technology stocks after the COVID-19 epidemic in China in 2022. Especially, Artificial Intelligence (AI) was chosen in this study due to its robustness and satisfactory performance [3].

Stocks are initially predicted and analyzed using economic models. Early in the 1960s, Sharpe et al. created the Capital Asset Pricing Model (CAPM) to predict the stock, which offered the first logical framework for connecting the needed return on investment to the investment's risk [4]. Chen et al. put forward a market model based on multiple factors in Economic Forces and the Stock Market, which forecasts the changes in stock prices by considering macroeconomic variables and the characteristics of stocks [5]. However, these economic-based models are not very accurate.

With the development of AI technology, AI has become the main technology for predicting stock price trends due to its high accuracy and robustness. Yoon et al. proposed a neural network-based method that uses historical stock prices, transaction volume, and other factors as input to build a stock price prediction model to predict the rise and fall of stock prices [6]. Lee proposed to use the mixed feature selection method to process the data of the stock market and input the selected feature into the Support Vector Machine (SVM) model as an input variable to predict the rising and falling trend of the stock [7]. Although these models work well, many of them focus on stable or low-volatility stock forecasts. For the prediction of Chinese stocks in 2022, the model that can better predict high-volatility stocks is more critical. Therefore, this paper intends to use the linear regression model in the field of machine learning to analyze and predict the price trend of stocks in the period of high volatility after the lifting of restrictions on China's COVID-19 epidemic policy in 2022.

The data used in this paper are from the data set in Quandl, this dataset collects information on the stocks of representative technology companies' financial market in China's epidemic period from the end of 2019 to the beginning of 2023. In the selection of technology stocks, this paper selected the stock data of Alibaba, Tencent, and Xiaomi as the representative of technology stock data for analysis and prediction. These three companies' global market value is at the top of the technology companies with high popularity and high technology level, so their stock data can better represent the overall trend of the China technology industry. We forecast the stock trend by building a linear regression model and training the stock data of three companies in 2022. The experimental findings demonstrate the great accuracy with which the advanced linear regression model in this study can forecast the price trend of technology stocks in a period of high volatility following the policy's unsealing and the extent to which it may represent price volatility.

2. Method

2.1. Dataset description and preprocessing

In this project, we selected the stocks of three companies for analysis, which are Tencent, Xiaomi, and Alibaba. The stock data for these companies was obtained from the data.nasdaq.com platform, which offers a wealth of financial and economic data, encompassing a range of asset classes such as stocks, commodities, derivatives, fixed income, and mutual funds. The platform provides access to real-time, historical, and end-of-day data, sourced from various providers such as Quandl, Sharada, and Brave new coin, among others. In addition, Python also provided the package of quandl to help the researchers get the dataset easily. We chose the period from the end of 2019 to the beginning of 2023,

DOI: 10.54254/2755-2721/15/20230840

including the period of COVID-19 in China. The datasets for each company comprised 12 volumes and approximately 797 rows, with each row indicating the date within the study period, and each volume representing a specific feature of the company's stock.

For a stock, the main features include high price, low price, close price, and turnover. And combined with the dataset obtained, we selected 'Nominal Price', 'High', 'Low', 'Previous Close', and 'Share Volume (000)' as the features that we used. This study's main purpose is not to study the complex relationship between the various eigenvalues, but only to predict a certain eigenvalue, so we chose the 'Nominal Price' as the label we want to forecast. To improve the utilization of data, we assign an outlier to the missing value, therefore, when a machine learning classifier processes the data, this will just be recognized and treated as an outlier feature. The data is nearly 800 days, and we want to predict the stock price for the next half month, which means the goal is to forecast out 2% of the entire length of the dataset. Based on this we defined the forecasted days variable' forecast_out' and then added a new column 'label' in the dataset to record the predicted value. The newly added column is expressed by moving the data in the 'close' column forward by 2% rows. Next, we defined X (all columns except label) as the feature and y (just 'label') as the label to predict. To speed up processing and improve accuracy, we used preprocessing module in sklearn to preprocess the X. 2% of the data was left when the label column was generated above. These rows do not have label data, so we can use them as input data for prediction. Then we created a new variable 'X_lately' to keep these data.

2.2. Proposed approach - Linear regression

In the context of regression problems in stock forecasting, Python provided an effective tool called Scikit-Learn. Scikit-Learn offers several classifiers for regression analysis, and after evaluating the accuracy of each model, the linear regression model was selected for this study. The linear regression algorithm is a frequently used method in solving estimation problems [8]. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Usually, we have a dataset D, x as a feature vector of dimension d, t as the target value (sometimes denoted by y), and the objective is to get a good mapping y through the linear regression model.

$$D = \{(x_i, t_i)\}_{i=1}^{N} \tag{1}$$

Parameterize y with w:

$$y = y(x, w) \tag{2}$$

For input vector, the prediction is:

$$y_n = y(x_n, w) \tag{3}$$

The model is defined as:

$$y(X, W) = \omega_0 + \omega_1 x_1 + \cdots + \omega_D x_D \tag{4}$$

Then, extending the formula by using the matrix to represent a set of fixed non-linear functions of the input vector:

$$y(X, W) = \sum_{j=0}^{M-1} \omega_j \phi_j(X) = W^T \phi(X) \tag{5}$$

The above formula is aimed to find the appropriate ($1 \leqslant i \leqslant D$) parameter for the linear regression model. The actual class values ($y_1, y_2, \ldots y_n$) will be approximately equal to the predicted y_n values. Linear regression algorithm has strong data processing speed, it doesn't require complex calculations, so it is easy to understand and could run fast even with large amounts of data. What's more, when some new data was added, it was easy to update the model.

After importing the sklearn, we used the Linear Regression classifier which provided training and testing data. The dataset was divided into two parts, one part is used as the training data and accounts for 80% of the entire dataset, and the other is used as testing data and accounts for 20%. We trained the machine learning classifier by training data and got the accuracy of the model by testing data.

3. Result and discussion

This section examines the forecasting ability of Regression Models for stock data. Then, we will illustrate the prediction results of the selected machine learning model based on the Hong Kong Stock

DOI: 10.54254/2755-2721/15/20230840

Exchange. The following Table 1, Table 2 and Table 3 demonstrate the accuracy and the forecast for the stock's Nominal price of these three companies for the next sixteen days. In addition, the stock price trend of these companies is also shown in Figure 1, Figure 2, Figure 3 and Figure 4. From the table, the accuracy of the model can reach more than 85%. The folding line chart shows the stock price trend from the beginning of 2020 to 2023. It is a period that included the outbreak of COVID-19, the implementation of epidemic prevention and control policies, and the lifting of the control policy. In this chart, the red line represents the past stock price fluctuations, and the blue one represents the future stock price forecast.

Table 1. Stock price prediction for Alibaba.

Company	Predicted results				Accuracy
Alibaba	83.113	80.948	87.004	82.987	0.912
	87.829	83.407	83.725	84.368	
	99.642	105.993	102.931	115.844	
	110.544	118.839	113.713	112.415	

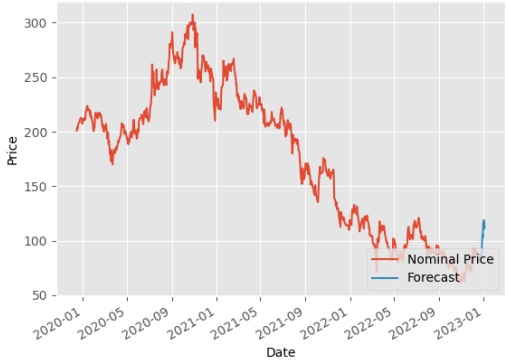

Figure 1. 2020.01-2023.01 Alibaba stock price trend.

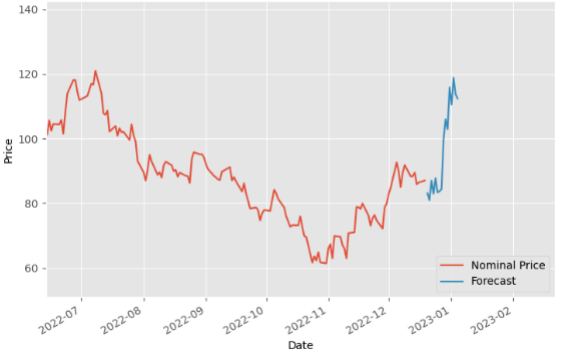

Figure 2. 2022.07-2023.01 Alibaba stock price trend.

Table 2. Stock price prediction for Tencent.

Company	Predicted results				Accuracy
Tencent	317.570	317.158	331.954	324.483	0.893
	335.034	340.295	343.000	346.762	
	369.789	361.420	358.326	367.628	
	365.678	383.966	370.260	368.247	

DOI: 10.54254/2755-2721/15/20230840

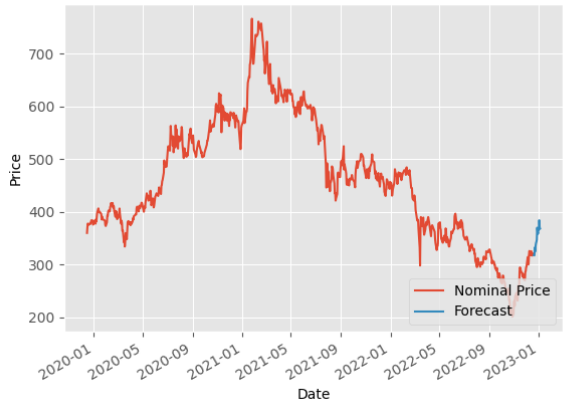

Figure 3. 2020.01-2023.01 Tencent stock price trend.

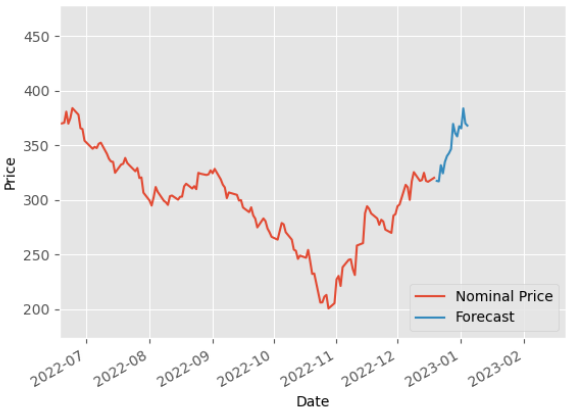

Figure 4. 2022.07-2023.01 Tencent stock price trend.

Table 3. Stock price prediction for Xiaomi.

Company	Predicted results				Accuracy
Xiaomi	10.670	10.647	11.178	10.881	0.876
	11.454	10.987	11.217	10.926	
	11.667	11.954	11.480	12.179	
	12.158	11.980	11.797	11.913	

The overall trend for these three companies revealed a fluctuating pattern with initial growth and subsequent decline in the stock prices of these companies. Notably, the peak period for all three companies was observed between the end of 2022 and the beginning of 2021. Both Tencent and Xiaomi stock prices have peaked January in 2021 but the Alibaba's peak occurred November in 2020. It is worth noting that, in comparison to the other two companies, Alibaba's stock price exhibited no significant growth, and after reaching the peak, it demonstrated a downward trend. It drops to the lowest point at the end of 2022, at this point the stock price is just 1/4 of it was at the beginning of 2020. After that, it began to rise again, and the predicted outcome indicates an upward trajectory. Regarding Tencent, it showed a trend with first steady fluctuations and then growth before reaching the highest point of the price. The price at the peak was twice that of the beginning of 2020. After that, it began to fall until the end of 2022. But there is some difference in the period of the second half of 2021, its stock price was relatively stable, it showed some fluctuations, but the average value remains steady. Since November 2022, the price had been increasing including the prediction result. For Xiaomi, in the early period, it revealed dramatic growth from the beginning of 2020 to January 2021, during this period it occurred a little peak in September 2020. And the highest price was up to 2.5 times as the beginning. Subsequently, it began to decline, but the price also recovered in the middle of 2021. For the forecasted prices, it shows an upward trend. In addition, for the lowest point of the price, both Alibaba and Tencent dropped to a level lower than the price at the beginning, but Xiaomi dropped at the same as the beginning.

The coronavirus outbreak in early 2020 was recognized as a "public health emergency" by the WHO and posed a serious challenge to the country's economy, with most industries being strongly affected by the outbreak [9]. Due to the epidemic policy, it can be observed that the epidemic policy has led to an overall downward trend in the stock prices. The transmission mechanism of the COVID-19 epidemic to the stock market was the contagion effect of negative investor sentiment, which led to

DOI: 10.54254/2755-2721/15/20230840

a resonance in the share prices of companies in different regions, resulting in a general decline in the stock market [10]. However, as China lifted the measure imposed to control the epidemic at the end of 2022, there has been a noticeable increase in the stock prices of all three companies. Therefore, it can be concluded that for the coming period, the stock prices of all three companies show a volatile increase, which is a very good thing for the recovery and development of the Chinese economy. Table 4 presents the general trends of the three companies.

Table 4. The general trends of the three companies.

Company / Trend	Alibaba	Tencent	Xiaomi
Past	Small growth and significant decline	Volatility growth volatility decline	First, grow and then fall back to the original level
Future	Volatility rises	Volatility rises	Volatility rises

4. Conclusion

In this work, we intend to use a training linear regression model to predict and analyze the price trend of scientific and technological stocks in a period of high volatility after the unpacking of China's COVID-19 Epidemic Economic Policy at the end of 2022. We trained a linear regression model, trained the dataset after data preprocessing, and performed visualization and accuracy calculations. The experimental results show that the advanced linear regression model in this paper can predict the price trend of technology stocks in a period of high volatility after the policy is unsealed with high accuracy and can reflect the price volatility to a certain extent. From the visualized image, it can be seen that due to the impact of China's previous epidemic policies on the continuous decline of stock prices, the policy lifting has enabled stocks to recover in a short period. In the future, we plan to use more complex and sophisticated machine learning models for forecasting, so that the prediction results can reflect more accurate fluctuations.

References

[1] Ibbotson R G and Sinquefield R A 2010 Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926-2008) (Chicago: Morningstar)
[2] Qiang Z 2020 Impact of the global spread of the new crown pneumonia epidemic on China's trade development and countermeasures vol 15 (Chinese: Business Economics Research) p 4
[3] Al-Blooshi L and Nobanee H and Nobanee H 2020 Applications of Artificial Intelligence in Financial Management Decisions: A Mini-Review
[4] Perold A F 2004 The Capital Asset Pricing Model (Journal of Economic Perspectives) vol18 pp 3-24
[5] Naifu C and Richard R and Stephen A R 1986 Economic Forces and the Stock Market vol 59 (The Journal of Business) chapter 3 pp 383-403
[6] Yoon Y and Swales G 1991 Predicting Stock Price Performance: A Neural Network Approach// Twenty-fourth Hawaii International Conference on System Sciences (IEEE)
[7] Lee M C 2009 Using support vector machine with a hybrid feature selection method to the stock trend prediction vol 36 (Expert Systems with Applications) chapter 8 pp 10896-10904
[8] Şahın D Ö and Akleylek S and Kiliç E 2022 LinRegDroid: Detection of Android Malware Using Multiple Linear Regression Models-Based Classifiers vol 10 (IEEE) pp 14246-14259
[9] DuanYou Y 2020 The impact of the COVID-19 epidemic on China's stock market - an empirical analysis based on the pharmaceutical industry vol 18(China: China Business News) p 3
[10] Hong X and Hongxia P 2021 The Impact of the COVID-19 Epidemic on the Chinese Stock Market - A Study Based on the Event Research Method vol 7 (China: Financial Forum) p 11