Stock Price Forecast Report

1. Introduction

Since the stock market changes in real time according to each trading day, and new data is generated every day, scholars make predictions about the future stock market by studying historical data. However, the price of stocks is affected by many factors. The future price of a stock is not so easy. However, with the development of machine learning, it has become possible to extract effective information from massive data, so it is very important for investors to be able to predict stock prices more accurately. Stock historical data is a time series with noise, how to extract effective information from it is very important. This paper extracts features from YFinace data set, extracts effective information as much as possible, and then predicts stock prices through the model , compared with the traditional time series model and the BP neural network widely used in financial research, it is a good development for stock price forecast analysis. Similarly, in practical applications, the model is trained through historical data to predict the stock price. If the closing price of the second day predicted by the model is higher than that of the day, the stock price will rise in the future, and investors can continue to hold positions to obtain income. Similarly, the forecast of stock prices can also enable the country to have a deeper understanding of the current economic situation and industry development, so as to make timely macro-control.

Building a financial model to achieve accurate stock price prediction has become a work of both theoretical and practical significance. In theory, whether it is a traditional financial time series model or an emerging artificial intelligence model, the prediction of stock prices can be achieved. However, the prediction effect is quite different. Finding a model with better prediction effect through the combination and comparison of models is something that many scholars are always happy to do, and it also has great theoretical significance. In real life, known data can be injected into financial models to realize prediction of future data. If the predicted value of the stock price is greater than the closing price of the day, it means that the model predicts that the stock price may rise in the future, and investors can choose to continue to hold the stock to obtain higher investment returns. If the predicted value of the stock price is less than the closing price of the day, it means that the stock price may fall in the future. In order to avoid risks and reduce losses, investors can choose a trading strategy of moderate selling. Therefore, it has important practical value to realize the prediction of stock price by constructing a financial model. If you can find a way to accurately predict the trend and volatility of stock prices, it will be extremely valuable for countries, listed companies, and individual investors. Furthermore, it is of high practical value to conduct investment strategy research on stock price judgment. A well-effective stock investment strategy can reduce the phenomenon of blind investment in the stock market to a certain extent, can help investors make better investment decisions and asset allocation, reduce investors' losses, and achieve rational trading while also increasing ROI.

2. Machine Learning

Machine learning belongs to the category of artificial intelligence and is the most core and important research direction in the field of contemporary artificial intelligence. The purpose

of machine learning research is to make the machine simulate the human thinking model, and continuously learn and improve, so as to acquire new skills and new knowledge. Research. This learning and optimization process enables the model to have higher efficiency, accuracy, stability, and general applicability.

The principle of machine learning is to form a corresponding relationship between input and output in the system model through sample data, that is, the learning function, which can be realized by optimizing the learning function. By injecting input data into the system, the predicted unknown data can be obtained at the output end. . Now machine learning has penetrated into many disciplines and has been widely used, such as data mining, financial market analysis, medical diagnosis and natural language processing.

The empirical risk minimization principle is to minimize the test error by seeking to minimize the training error. However, many researchers fall into a misunderstanding, thinking that as long as the empirical error is minimized, a better prediction effect can be achieved. In fact, when the empirical error is small enough, the model may overfit, thereby weakening the generalization ability and increasing the real risk. Therefore, it is not wise to blindly seek to minimize the empirical error. Similarly, if the empirical risk is not well controlled, the model will suffer from underfitting, both of which can greatly reduce the predictive power of the model. In order to avoid overfitting and underfitting, it is necessary to not only ensure the model optimization of the learning machine, but also to grasp the amount of model test sample data.

When the data is limited, there are three main conflicts between the complexity of the model and the general level of the model:

(1) The effect of the model is affected by empirical risk, but the degree of influence is not absolute. Adhering to the ERM principle does not necessarily increase the generalization of the model.

(2) The higher the complexity of the model, the lower the empirical risk, but blindly adhering to the ERM principle will directly lead to the complexity of the model.

(3) The complexity of the learning machine has a great influence on the comprehensive performance of the model. A learning machine with better generalization ability cannot only pursue the complexity of the model, but has a level of complexity that matches the reality.

It can be seen that when the sample data is limited, adhering to the ERM principle is not equivalent to the expected risk minimization, because ERM can only ensure that the learning error of the model is small, but it will lead to the problem of poor generalization level and large error of the learning machine.

3. Factors Affecting Stock Prices

The market price of a stock is the market price of a stock. The stock price can reflect the market situation of the stock market. Investors will also decide whether to invest or not

through the stock price. Analyze the factors that affect the rise and fall of stock prices from both macro and micro aspects. At the macro level, it is mainly determined by national policies and the general economic environment; at the micro level, it is mainly determined by the market, including the internal changes of the company, the industry situation, investor psychology, etc. The factors that ultimately affect the stock price mainly include the following aspects:

(1) National policy situation

my country's stock market prices are obviously affected by national policies, and stock prices often fluctuate under the influence of positive and negative information from national policies. Policy factors include macro-control monetary policy and fiscal policy. Monetary policy is mainly the adjustment of interest rates by the central bank. If inflation is severe and economic development is overheated, the central bank will raise interest rates to recover funds; in times of economic crisis, it will increase the money supply by cutting interest rates. There will be more capital into the stock market, which will help the stock price to rise. For example, since 2015, the central bank has issued several announcements of interest rate cuts and RRR cuts, which have had a greater impact on stock price fluctuations. Fiscal policy is mainly about taxation and fiscal expenditure. Tax increases or decreases affect the profits of individuals, companies and investment funds. When the profit of a listed company increases and the intrinsic value of the stock increases, investors will consider investing in the company, and then the stock price will rise, and companies and individuals will have more funds to engage in stock investment. The state's financial expenditure on the industry will also drive the stock price of this industry to rise accordingly, which is regarded as good news of fiscal policy, reflecting the country's optimism about the future of this industry. Investors will inevitably pay attention to the development and changes of the industry, and the stock price will also rise.

(2) Industry factors

The development of the industry will also have an impact on stock prices, which is an influencing factor between macro and micro. When the government encourages the development of certain industries, the stock prices of these industries will rise. For example, due to the impact of the new crown epidemic in recent years, everyone has paid great attention to the issue of vaccine research and development, and the stocks of the pharmaceutical industry have also become the focus of investors' attention.

(3) Internal status of listed companies

To judge whether a listed company is worthy of investors' investment, in the final analysis, it depends on whether its own operation makes investors optimistic about its future development. Therefore, investors can determine profitability, operating capacity, capital, cash flow, etc. through the financial statements disclosed by the company in the official website. The stronger the profitability of a listed company, the higher the profit; the stronger the management ability of the enterprise, it means that it can reasonably use the resources of the enterprise and control the cost reasonably; the faster the enterprise goes, the future of the

enterprise The growth value, development scale, development trend and speed of the enterprise can promote the improvement of the comprehensive strength of the enterprise; and the cash flow of the enterprise is one of the important indicators of the operation and management of the enterprise. more secure. When a listed company's financial performance is good, it will attract more investors to invest and increase the company's stock price.

4. Commonly used analytical methods for stock forecasting

At present, most individual investors use traditional analysis methods, namely fundamental analysis method or technical analysis method, to predict the trend of stock prices, while professional institutions will use statistical and artificial intelligence related theoretical knowledge to assist them. The commonly used methods are as follows:

(1)Technical Analysis

Under this method, we mainly observe the changing trend of the stock price on the relevant chart. Charts include time-sharing charts, daily charts, weekly charts, and related technical indicators include MACD, RSI, MR, etc.

(2)Time series analysis

This method has strict presuppositions and higher requirements on the data itself. Predictive work by leveraging historical information. The main models are ARMA, exponential smoothing, GARCH, etc.

(3)Neural network prediction method

A neural network is a set of input/output units, each of which is associated with a weight. It turns out that neural networks can approximate any function, and the stock market is a nonlinear system. There are many laws and characteristics among the variables that affect stock price changes. By studying historical transaction data, neural networks can find some rules in complex data.

5. Datasets and Preprocessing

The choice of yfinance for this article aims to solve this problem by providing a reliable, threaded, Pythonic way to download historical stock exchange market data. yfinance is a popular open source library developed by Ran Aroussi for accessing financial data available on Yahoo Finance.

Reasons to choose this dataset are: free, quick and easy setup, simplicity, high data granularity, return data directly in Pandas dataframes/series. Also, yfinance is completely open source and free. There are other ways to access Yahoo Finance data, some free, some paid, and some paid options have certain benefits, such as ensuring a certain level of maintenance for the solution. In particular, installation is easier. yfinance has only 4 dependencies, all of which ship with Anaconda and are fully installed in one line of code. There is no need to create an

account, and no need to register and use an API key. The design of yfinance is very Pythonic and very lean. It's as simple as creating a ticker object for a specific ticker, then just calling all the methods on this object.

Because the study of stock prices is of great significance to investors, those stocks with good performance and large market capitalization are usually selected as prediction objects. This paper randomly selects a stock in the U.S. stock market as the research object, in order to eliminate the interference of subjective selection and judgment. Finally, MU was randomly selected as the target stock studied in this paper. The selection of data scale is also different according to the change of the stock market fluctuation law and the length of the movement cycle. If the period of the stock is short, the data on the small scale should be selected, otherwise, the data on the large scale should be selected. When the amount of selected data is large, the generalization performance of the model is enhanced, but the accuracy of local prediction may be correspondingly reduced. In addition, if the data scale is too large, the parameters of the model will be over-fit due to the complicated calculation, which will lead to the difficulty of solving. Similarly, if the data size is too small, the learning machine cannot learn well, resulting in poor model fitting effect. Therefore, the selection of data scale is very important for the model of machine learning. This article is mainly aimed at the prediction of medium and long-term stock prices, so the data interval is selected from January 1, 2014 to September 17, 2022.

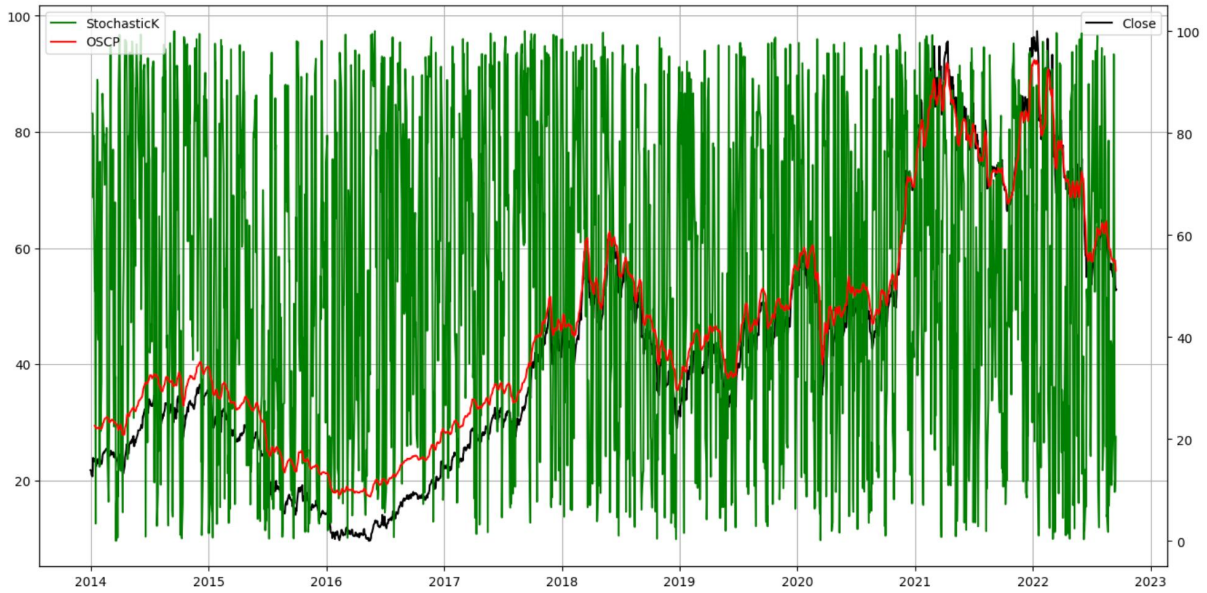
Micron Technology, Inc., incorporated on April 6, 1984, is a global leader in semiconductor devices (primarily NAND flash, DRAM and NOR flash, as well as other memory technologies, packaging solutions and semiconductor systems for computing, user, networking, Manufacturers and sellers of automotive, industrial, embedded and mobile products). In addition, the company manufactures semiconductor components for CMOS image sensors and other semiconductor products. The company has four divisions: NAND Solutions Group (NSG), DRAM Solutions Group (DSG), Wireless Solutions Group (WSG) and Embedded Solutions Group (ESG). The company's products include NAND flash memory, dynamic random access memory (DRAM), and NOR flash memory. The company's production facilities are located in the United States, China, Israel, Italy, Malaysia, Puerto Rico and Singapore.

In this paper, the label is set to the stock price of the second day, and the data set is shown as follows:

Date	Close	Label
2013-12-31	21.750000	21.660000
2014-01-02	21.660000	20.969999
2014-01-03	20.969999	20.670000
2014-01-06	20.670000	21.730000
2014-01-07	21.730000	23.870001
...	...	...
2022-09-12	57.939999	53.619999
2022-09-13	53.619999	53.110001
2022-09-14	53.110001	52.689999
2022-09-15	52.689999	52.849998
2022-09-16	52.849998	NaN

6. OSCP Regression

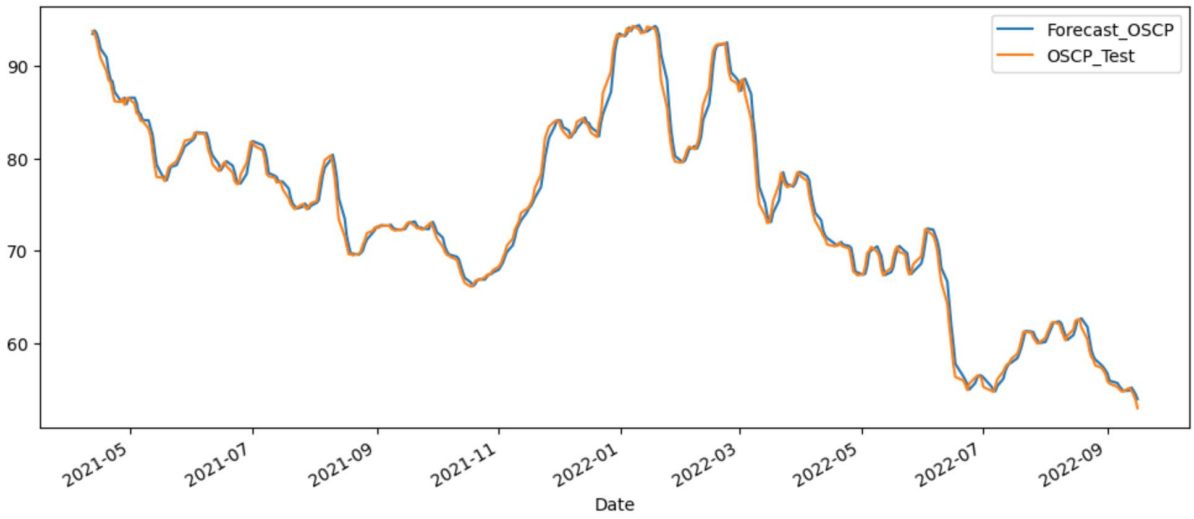
The Price Oscillator uses two moving averages, one short term and one long term, and then calculates the difference between the two moving averages. Price oscillators can give overbought and oversold conditions, as well as try to confirm bullish and bearish price moves. The length of the moving average is user-determined. Price swings can be used to detect when the trend is slowing down and possibly reversing. This is released when the price oscillator returns to the 0 line. Conversely, when the price oscillator moves away from the 0 line, the price trend is accelerating. Also, price oscillators may reveal overbought and oversold areas where traders may look for buying opportunities when the price oscillator bottoms. Of course, other technical indicators need to be used to assist judgment. Similarly, in an overbought zone, when a price oscillator tops out, traders may look to sell. Other technical indicators need to be used as an aid, however, the price makes a biased recommendation on whether to execute a buy or sell signal: usually, a trader may want to sell when he sees an area of overbought; In the area, traders may want to buy. By performing feature extraction on the data, the OSCP features are as follows:



Through the extraction of various features of the data (StochasticK, OSCP, Close_MA5, Close_MA10), some recent data are shown in the following table:

Date	Close	Label	Ln	Hn	StochasticK	Close_MA5	Close_MA10	OSCP
2022-09-12	57.939999	53.619999	53.740002	58.139999	9.599991	56.199999	56.448000	54.873865
2022-09-13	53.619999	53.110001	53.139999	58.139999	11.443688	55.877999	56.109000	54.494774
2022-09-14	53.110001	52.689999	52.459999	58.139999	11.525959	55.499999	55.790000	53.951884
2022-09-15	52.689999	52.849998	51.980000	58.139999	20.451087	54.959999	55.406000	53.025012
2022-09-16	52.849998	NaN	51.490002	58.139999	NaN	54.041999	54.959999	NaN

The stock price is predicted by the OSCP feature, and the effect is as follows:



7. Random Forest Regression and Linear Regression

Random forest regression algorithm is an important application branch of random forest. The random forest regression model builds multiple unrelated decision trees by randomly extracting samples and features, and obtains prediction results in a parallel manner. Each decision tree can obtain a prediction result through the extracted samples and features, and obtain the regression prediction result of the entire forest by taking the average of the results of all trees.

The random forest regression algorithm is used in scenarios that require relatively low data dimensions (tens of dimensions) and high accuracy requirements. For example, predicting the popularity of a topic on Twitter can be handled using an SVR regression model. The input of the model can be the characteristics of the topic, such as the number of discussion groups generated by the topic at a certain time, the number of people participating in the discussion of the topic, and the topic attention. The output of the model is the average number of active discussion groups per hour, which is a positive floating point number, and the larger the value, the higher the popularity.

In statistics, linear regression is a regression analysis that models the relationship between one or more independent and dependent variables using a least squares function called a linear regression equation. Such a function is a linear combination of one or more model parameters called regression coefficients. The case with only one independent variable is called simple regression, and the case with more than one independent variable is called multiple linear regression.

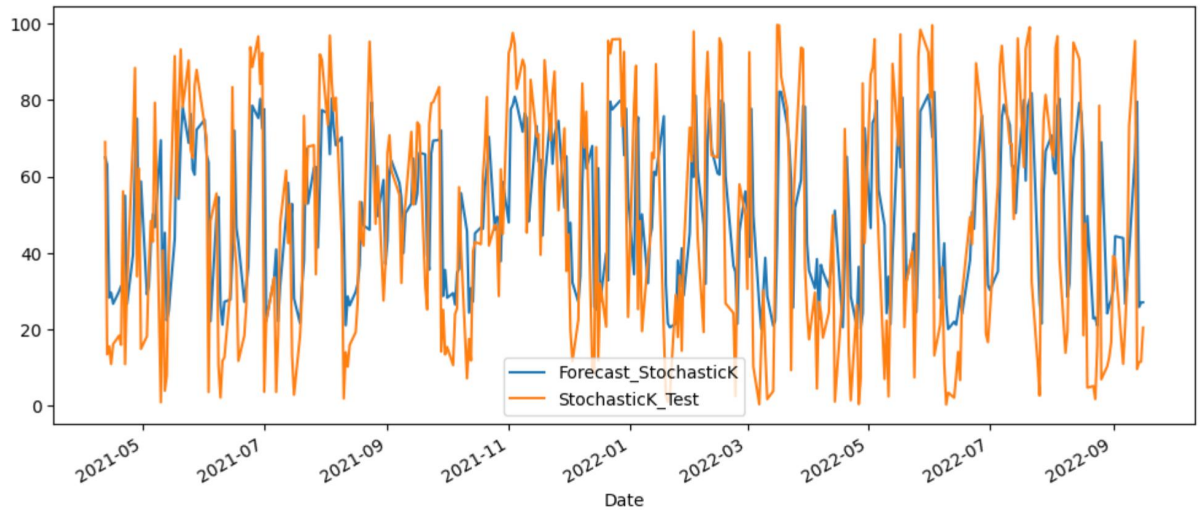
In linear regression, the data is modeled using a linear prediction function, and unknown model parameters are also estimated from the data. These models are called linear models. The most commonly used linear regression modeling is that the conditional mean of y for a given value of X is an affine function of X. Less generally, a linear regression model can be a median or some other quantile of the conditional distribution of y given X, expressed as a linear function of X. Like all forms of regression analysis, linear regression also focuses on the conditional probability distribution of y for a given value of X, rather than the joint probability distribution of X and y (the field of multivariate analysis).

Linear regression was the first type of regression analysis to be rigorously studied and widely used in practical applications. This is because models that depend linearly on their unknown parameters are easier to fit than models that depend nonlinearly on their unknown parameters, and the statistical properties of the resulting estimates are easier to determine.

If the goal is prediction or mapping, linear regression can be used to fit a prediction model to the observed data set and the values of X. After completing such a model, for a new value of X, the fitted model can be used to predict a value of y without a given y paired with it.

Linear regression models are often fitted using least squares approximation, but they may also be fitted in other ways, such as by minimizing "fit imperfections" in some other specification (such as least absolute error regression), or in bridge Penalty for minimizing the least squares loss function in regression. Instead, least squares approximation can be used to fit those nonlinear models. Therefore, although "least squares" and "linear models" are closely related, they are not equivalent.

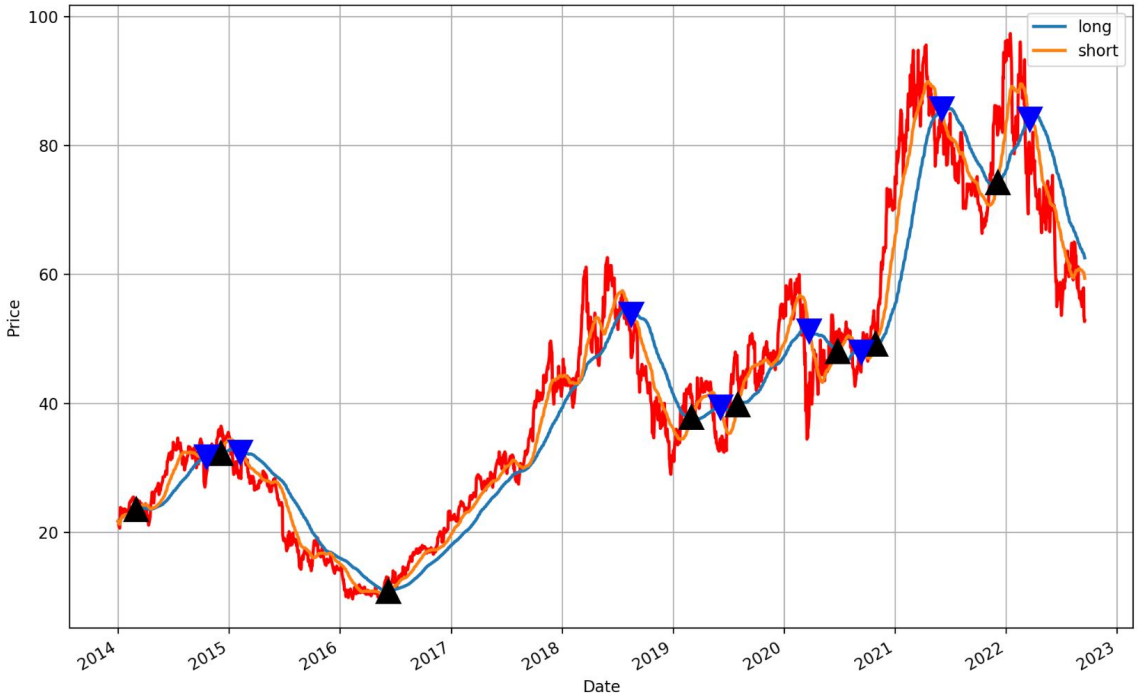
By performing linear regression on the data, the stock price prediction effect is as follows:



8. Stock trading decisions

The ultimate goal of stock price prediction is to help people make decisions about stock trading. With the development of artificial intelligence technology, artificial intelligence has gradually been widely used in more and more fields. In the field of economy and finance, the role of artificial intelligence is becoming more and more prominent. The high returns in the stock market make many investors enter one after another, and the accompanying high risks make investors suffer heavy losses. Therefore, the study of the inherent laws of the stock market and the prediction of stock prices have great application value. The change of a stock price has a certain relationship with its historical data, and we can make predictions by analyzing the historical information. The high noise, nonlinearity and other factors of the stock market make forecasting difficult. Therefore, the traditional prediction method is not feasible, and it is difficult to establish an effective mathematical model. Machine learning can predict time series by virtue of its high tolerance to noise data, nonlinear mapping ability, generalization ability, self-learning and adaptive ability, etc. It does not need to make any analysis assumptions in solving practical problems.

By comprehensively processing the above extracted features, the model's prediction of stock price trading decisions is as follows:



9. Conclusion

In this paper, the machine learning feature extraction and analysis of stocks are carried out. It can be seen that the model can be better applied to individual stocks in the financial market and has a good promotion value. In order to further test the stability and sensitivity of the model, the model was processed for stock replacement, position holding period and time period respectively, and it was found that the prediction effect of the model was always relatively stable. In addition, the training set data in this paper is large, the model fitting effect is good, the prediction accuracy rate is high, and the prediction error is low. Therefore, when building a model, high-frequency data should be selected, and the training set data should be as large as possible to ensure that the model fits well and the prediction accuracy rate is high.

10. References

A.A.Ariyo, A.O. Adewumi and C.K.Ayo. 2014. Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation: 106-112.

Cao L, F.E.H.Tay. 2001. Financial forecasting using support vector machines. Neural Computing & Applications 10(2): 184-192

H.Ince,T.B.Trafalis. 2008. Short Term Forecasting with Support Vector Machines and Application to Stock Price Prediction. International Journal of General Systems, 37(6): 677-687.

B.M.Henrique, V.A.Sobreiro, H.Kimura. 2018. Stock Price Prediction Using Support Vector Regression on Daily and up to the Minute Prices. The Journal of Finance and Data Science, 4(3):

183-201.

Xiao C,Xia W,Jiang J. 2020. Stock price forecast based on combined model of ARIMA - LS -SVM. Neural Computing and Applications, 32(6): 5379 - 5388.