Stock price prediction study based on LSTM and random forest model

*Yunkai He[1, *], Bo Yang[2], Yifan Yao[3], Zikai Lu[4], Liqun Yu[5]*

(1: Nanchang Hangkong University, Jiangxi, China

2: Xian Innovation College of Yanan University, Shanxi, China

3: Hubei University of Technology, Hubei, China

4: Jiangsu Ocean University, Jiangsu, China

5: Huaiyin Normal University, Jiangsu, China)

ABSTRACT

The growing impact of global climate change on the economy and financial markets highlights the importance of environmental factors in the financial sector. The purpose of this study is to explore the correlation between environmental factors and the overall performance of the stock market, and the short-term impact of extreme weather events on the stock prices of specific industries (e. g., energy industries), and to build a commodity price prediction model to provide investors with more comprehensive and accurate financial market analysis and decision support. Through multi-source data collection and analysis, we found the correlation between environmental factors and the overall stock performance index, and revealed the impact of extreme weather events on energy stock prices. Meanwhile, the LSTM and random forest regression model were used to predict commodity prices, revealing the importance of environmental factors in financial markets, and providing a useful reference for the study of the correlation between environmental factors and financial markets.

Keywords: Climate Change; Financial Markets; Environmental Factors; Emergencies; Stock Forecast

1 INTRODUCTION

The growing impact of global climate change on the economy and financial markets highlights the importance of environmental factors in the financial sector. Extreme weather events and air quality changes triggered by climate change not only affect crop yields and public health, but also cause food price fluctuations and increased demand for health care, At the same time, it leads to the impact of policy adjustment on the industrial structure and market pattern. In this context, interdisciplinary knowledge, such as climate science, environmental policy, and how environmental factors influence financial markets through economic mechanisms, becomes crucial [1]. Therefore, the purpose of this study is to deeply analyze the correlation between environmental factors and the overall performance of the stock market, study the short-term impact of extreme weather events on the stock prices of specific industries (such as the energy industry), and build a commodity price prediction model to provide a more comprehensive and accurate financial market analysis and provide more reliable decision support for investors [2].

2 MODLE PREPARATION

During data collection, we obtained data at different dates from multiple sources:

First, we collected data on extreme weather events, including data from the National Weather Service, international organizations, scientific research institutes, and professional meteorological websites and applications. Second, we obtained the air quality index data through government websites and third-party data service platforms. In addition, we also collected stock index data from the stock market data download website and the official website of the stock exchange [3]. For the tourism stock data, we reviewed the financial data platform and industry research reports. Energy inventory data comes from data crawl and official agencies such as the National Bureau of Statistics. Oil price data come from the benchmark oil prices in the international market, while natural gas price data are obtained through the international energy market report and data released by national statistical agencies. Finally, we queried the NBS for strategy change data and consulted climate datasets published by international organizations and climate change reports and datasets regularly published by government agencies and research centers [4]. These data will provide important support for our model preparation and further analysis.

The variables counted in this paper include the following Table 1:

Table 1：Study variables

Extreme_Weather_Event	AQI	Stock_Index	Tourism_Stock	Energy_Stock	Oil_Price	Gas_Price	Policy_Change	Climate_Change_News

3 SEF-SMP CORRELATION STUDY

3.1 Spearman correlation analysis model

Spearman Correlation analysis is a non-parametric method used to measure the monotonic relationship between two variables. It uses the ranking information, transforms the raw data into rankings, and calculates the correlation between the rankings. Here is the rationale for the Spearman correlation analysis:

Conversion to rank: For each index, the value is converted to the corresponding rank, that is, the data is ranked in the order from small to large. If there is the same value, the average rank can be taken [5].

Calculate the rank difference: for each pair of data points, calculate the rank difference between the two variables, i. e. If there are multiple identical data points, the average rank was used.

Calculate Spearman correlation coefficient: Spearman correlation coefficient (usually expressed) is calculated by the following formula: ρ

$$\rho = 1 - \frac{6\sum d^2}{n \times (n^2 - 1)} \tag{1}$$

Where, n is the number of data points.

Judge correlation: The value of Spearman correlation coefficient ranges between-1 and 1. Positive values indicate positive correlation, negative values indicate negative correlation, and 0 indicates no correlation. The closer the correlation coefficient is to 1 or-1, the stronger the correlation is.

3.2 Model results

Programming with MATLAB, the correlation coefficient table is obtained as follows Figure 1:

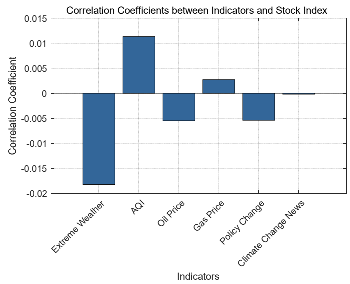

Fig.1：Correlation coefficient diagram of each index on the stock overall performance index

According to the correlation coefficient diagram, we draw the following conclusions: First, in the relationship between specific variables, there is a weak positive correlation between extreme weather events and AQI, while there is a weak positive correlation between policy changes and natural gas prices [6]. None of the other variables were significant. This suggests that the linear relationship is relatively weak and more thorough analysis or other factors may be needed to explain the association between these variables.

4 EWE-SMP CORRELATION STUDY

The research ideas adopted in this part are as follows:

(1) Time series drawing: By drawing the time series drawing, take the time column as the independent variable and the energy stock price column as the dependent variable to show the relationship between them.

(2) Mark extreme weather events: mark the time point of extreme weather events on the time series map. This is achieved by looking at each time point and judging whether the corresponding number of extreme weather events is greater than or equal to 1.

(3) Observe the stock price trend: observe the standard red time point, and check whether the energy stock price rises or falls for a length of time after the time point. If falling, it is marked as descending point, then as rising point.

(4) Statistical falling points and rising points: the number of falling points and rising points in all the standard red time points.

(5) Determine the impact of extreme weather events on stock prices: By comparing the number of falling points and rising events, determine whether extreme weather events will promote the decline or rise of energy stock prices [7].

This process can be expressed by a mathematical formula as follows:

Make T is the time series, P is the energy stock price series, and E is the number of extreme weather events.

For each time point ti, if Eti 1.

For the marked red time point ti, observe the stock price change after n days: if Pti + n <Pti, mark as falling point; if Pti + n> Pti, mark as rising point.

Number of statistical falling points D and number of rising points U.

If D> U, extreme weather events will promote energy stock price decline; if D <U, extreme weather events will drive energy stock price rise.

This model explores possible associations through time series analysis and statistical principles, and observations of extreme weather events.

Based on the above ideas, the solution results are obtained as follows Figure 2:

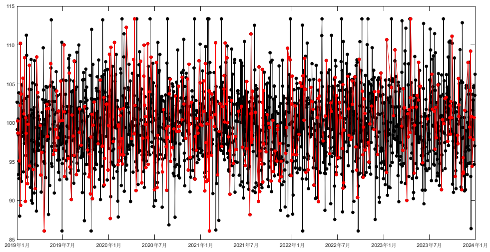

Fig.2: Model Results

By observing the curve, it can be seen that after the occurrence of the extreme events, the change trend of the energy stock price will have relatively obvious changes, which proves that the extreme events have a direct impact on the change of the energy stock price.

5 EF-DATA COMMODITY PRICE PREDICTION

4.1 Independent variable prediction based on the LSTM

Long-and short-term memory network (LSTM) is a recurrent neural network model for processing time-series data. Compared to the standard recurrent neural networks, LSTM has a more complex structure, including the forget gate, input gate and output gate, to control the cell state update and output [8]. Through the structure of the gate, LSTM can selectively decide to retain or discard information in the cell state and control the state characteristics of the output cells according to the input and gate. The working principle is to transfer and calculate the information between layers and layers within the time step, and finally output the results of each time step.

The final predicted values of the respective variables are Table 2:

Table 2: Predictive values of the independent variables

Date	Extreme weather events	Air quality index	Share index	Travel stocks	Energy inventory	Policy changes	Climate change news
2024/1/1	1.0000	49.5663	994.5337	99.1479	100.4051	0.0000	2.0000
2024/1/2	0.0000	48.5175	1019.0396	100.5818	98.9493	0.0000	3.0000
2024/1/3	0.0000	49.1536	1017.0290	100.4535	100.1623	0.0000	2.0000
2024/1/4	1.0000	56.7704	986.4772	100.3275	99.5406	0.0000	3.0000
2024/1/5	0.0000	53.4141	987.6823	100.6417	101.5366	0.0000	1.0000
2024/1/6	0.0000	55.0242	998.1662	100.6647	98.8678	0.0000	2.0000
2024/1/7	0.0000	54.8871	1009.5639	100.6347	99.6461	0.0000	2.0000
2024/1/8	0.0000	46.2179	985.1569	103.2123	100.8814	0.0000	3.0000
2024/1/9	1.0000	56.0414	987.5442	103.2277	99.8754	0.0000	3.0000
2024/1/10	0.0000	54.1737	1000.3491	99.3133	98.2882	0.0000	2.0000

4.2 Stock price forecast based on random Forest

Random forest regression (RFR) is a regression model based on decision trees. Multiple decision trees are constructed by bootstrap, and then their prediction results are combined by average to predict continuous dependent variables. Compared to other methods, RFR is insensitive to multivariate collinearity and is robust to modeling missing and non-balanced data, which is suitable for large-scale data and multivariate cases [9]. In a given training sample, 80%, 300 decision trees, and the number of node variables is 5. The model results obtained from modeling are as follows: The results obtained by Matlab modeling show that the mean square error of the training sample is 0.0395, the average error is 0.0439, and the average relative error is 2.31%. Through this model, the price of oil and natural gas can be predicted [10]. The specific prediction results are shown in Table 3.

Table 3：Oil and gas price forecast results

Date	Natural gas prices	Oil price
2024/1/1	5.0113	49.8815
2024/1/2	5.0024	49.7230
2024/1/3	4.9987	50.1620
2024/1/4	4.9736	49.7952
2024/1/5	4.9746	50.0353
2024/1/6	5.0011	49.6323
2024/1/7	4.9896	49.8440
2024/1/8	5.0331	50.3689
2024/1/9	4.9801	50.8332
2024/1/10	5.0187	49.4736

6 CONCLUSION

This study provides an in-depth analysis of environmental factors with financial markets by integrating interdisciplinary knowledge. First, we collected multiple data sources, including extreme weather events, air quality indexes, stock indices, travel stocks, energy inventories, strategy changes, and climate change news. Then, we studied the relationship between these factors and the overall stock performance index by using the Spearman correlation analysis method, and found that the correlation was low. On this basis, we further explore the short-term impact of extreme weather events on energy stock prices using time series analysis and statistical methods, and the results show that extreme weather events can have a significant impact on energy stock prices. In addition, we also used LSTM and random forest regression models to predict commodity prices, providing investors with more comprehensive and accurate financial market analysis and decision support. In conclusion, this study reveals the importance of environmental factors in the financial field and provides a useful reference for studying the correlation between environmental factors and financial markets.

REFERENCES

[1] Park, H. J., Kim, Y., & Kim, H. Y. (2022). Stock market forecasting using a multi-t ask approach integrating long short-term memory and the random forest framewo rk. *Applied Soft Computing*, *114*, 108106.

[2] Park, H. J., Kim, Y., & Kim, H. Y. (2022). Stock market forecasting using a multi-t ask approach integrating long short-term memory and the random forest framewo rk. *Applied Soft Computing*, *114*, 108106.

[3] Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements o f stock prices for intraday trading using LSTM and random forests. *Finance Researc h Letters*, *46*, 102280.

[4] Ray, S., Lama, A., Mishra, P., Biswas, T., Das, S. S., & Gurung, B. (2023). An ARI MA-LSTM model for predicting volatile agricultural price series with random fores t technique. *Applied Soft Computing*, *149*, 110939.

[5] Chen, J. (2023). Analysis of bitcoin price prediction using machine learning. *Journal of Risk and Financial Management*, *16*(1), 51.

[6] Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., & Fareed, M. (2022). Stock market forecasting using the random forest and deep neural network models befor e and during the COVID-19 period. *Frontiers in Environmental Science*, *10*, 917047.
[7] Darapaneni, N., Paduri, A. R., Sharma, H., Manjrekar, M., Hindlekar, N., Bhagat, P., ... & Agarwal, Y. (2022). Stock price prediction using sentiment analysis and d eep learning for Indian markets. *arXiv preprint arXiv:2204.05783*.
[8] Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using CNN-BiLSTM-Atte ntion model. *Mathematics*, *11*(9), 1985.
[9] Eghtesad, A., & Mohammadi, E. (2023). Portfolio optimization with return predicti on using LSTM, Random forest, and ARIMA. *Journal of Financial Management Perspe ctive*, *13*(43), 9-28.
[10] Vuong, P. H., Dat, T. T., Mai, T. K., & Uyen, P. H. (2022). Stock-price forecasting based on XGBoost and LSTM. *Computer Systems Science & Engineering*, *40*(1).