International Conference on Digital Intelligence Management and Innovation (DIMI 2025)

ESG Investment Analysis and Stock Price Prediction Using FinBERT and LSTM Models

Nan Zheng
International Business School, Xi'an Jiaotong-Liverpool University, Suzhou, China

Abstract: This study explores the integration of Environmental, Social, and Governance (ESG) data with deep learning models for stock price prediction. We propose a hybrid framework combining FinBERT, a pre-trained financial language model, for sentiment analysis of ESG-related texts, and Long Short-Term Memory (LSTM) networks for time-series forecasting. Experimental results on Ping An Bank (stock code: 000001.SZ) demonstrate that the model achieves a Mean Absolute Error (MAE) of 0.107 and Root Mean Squared Error (RMSE) of 0.163, indicating robust convergence. The analysis highlights the potential of ESG sentiment as a predictive factor and underscores challenges such as data scarcity and inconsistent ESG ratings. Future research directions include multi-enterprise validation and model comparison.

Keywords: ESG Investment; FinBERT; LSTM; Stock Prediction; Natural Language Processing.

1. Introduction: ESG Investment: Opportunities and Hurdles

ESG is a concept and practical framework that focuses on the sustainable development of enterprises. It is an abbreviation for "Environment", "Social", and "Governance". With the continuous development of economic globalization, countries around the world are paying increasing attention to environment, society, and governance. Issues such as global warming, improper corporate management policies, frequent occurrences of social inequality, and the widening gap between the rich and the poor have led countries to place greater emphasis on ESG.At the same time, many investors and asset management companies are increasingly emphasizing the role of ESG in investment. The investors' emphasis on ESG has led to an improvement in companies' ESG practices. This is a mutually complementary and mutually reinforcing relationship.Research has indicated that there is a positive connection between a corporation's financial performance and the degree of sustainability in its business model[1]. Moreover, the ESG investment approach can assist in decreasing portfolio risk. Meanwhile, it can produce returns that are frequently on par with those of traditional portfolios (Friede, Busch, and Bassen 2015)[1].Simultaneously, when confronted with extreme risks like financial crises and the global COVID-19 pandemic, companies with excellent ESG performance tend to have more stable market values. Their stock prices also demonstrate greater resilience (Broadstock et al., 2021; Lins et al., 2017)[2][3]. Although the importance of ESG in investment is constantly increasing, there are many difficulties in the process of studying ESG data and corporate performance. Henriksson et al. (2019) further explore the drawbacks of existing ESG rating methods[4]. They point out that the input data for these ratings usually relies on voluntary disclosure, resulting in a scarcity of data. Moreover, the scoring methods vary among different ESG data providers.Most of the data collected for ESG is voluntarily provided by companies, and its fairness and completeness are subject to verification. Some companies may deliberately conceal information that has a negative impact on themselves. At the same time, the amount of data that can be currently collected is limited. Moreover, different data providers use different scoring methods, making it impossible to compare. This prevents investors from making better decisions.

Zhou, Liu and Luo (2022) introduced the company's financial performance as a mediating variable into the analysis framework of the impact of ESG performance on the market value of listed companies, and constructed linear regression models and mediating effect models to examine it from different aspects such as profitability, operational capacity, and growth capacity[5].They also assign scores to ESG

information to obtain different ratings. This is the approach used by some researchers to solve the problem. Later in this paper, we will present a different method.

2. ESG Investment : Technologies, Models, and Cutting-Edge Tools

As the global economy grows, businesses and investors are increasingly prioritizing Environmental, Social, and Governance (ESG) factors, which are important influence on investment decisions. Despite the growing attention paid to ESG, there are still deficiencies in research. The key factors are the lack of a defined absolute index and the inconsistent evaluation criteria among various institutions.Researches show a positive link between firms' financial performance and business model sustainability. ESG investment can lower portfolio risk and yield competitive returns (Friede, Busch & Bassen, 2015)[1]. So we can use the machine learning to analyze the scores of ESG whether they influence the prices of stock.

Given machine learning's potential for ESG score analysis, exploring its powerful sub - field, Deep Learning, is crucial. Deep Learning helps machines better handle complex ESG - related data, opening new doors for more accurate analysis. Deep Learning is a crucial subfield within the realm of Machine Learning.It operates by constructing neural network models with multiple layers, enabling computers to automatically learn features and patterns from vast amounts of data. Deep Learning (DL) represents a significant advancement in the domain of Machine Learning (ML). Over the past few decades, ML has seen substantial progress in complex learning algorithms and effective pre - processing techniques. One of these key developments is the transformation of Artificial Neural Networks (ANNs) into increasingly deep neural network architectures. This transformation has enhanced learning capabilities, and the overall advancement is what we now know as DL (Goodfellow et al., 2016; LeCun et al., 2015)[6][7].

NLP is an interdisciplinary field at the intersection of computer science and linguistics, which is used to understand and process human language. With the rapid development of deep learning, NLP can leverage machine learning models to conduct in - depth understanding and analysis of texts such as news risk assessment and credit rating and customer service. With the continuous development of the digital economy, the volume of various financial texts, such as news and information, listed company financial reports, and financial research reports, has increased significantly, leading to a growing demand for Natural Language Processing (NLP). The importance of NLP has also become more prominent day by day.

Some scholars, in an attempt to explore the impact of ESG - related information on the stock market, employed the following methods. By collecting news articles of 30 constituent companies and calculating the sentiment of the news articles using the dictionary method of Loughran and McDonald, ESG sentiment data is impromptu constructed, and the VADER - LM method is used for the robustness test. They applied the Autoregressive Distributed Lag (ARDL) model, taking the ESG sentiment index as the independent variable and the idiosyncratic returns as the dependent variable, and set multiple control variables. Through different model specifications, they distinguished between positive and negative sentiment and considered the impact of the current financial performance of stocks on investors' reactions.

Another research model is the Random Forest. It is a machine learning method based on the principle of ensemble learning. It constructs multiple distinct decision trees and then integrates their prediction results to perform classification or regression tasks.The main advantages of the Random Forest lie in its ability to reduce the risk of overfitting, accurately evaluate and quantitatively analyze data, and remain unaffected by abnormal data. However, it is difficult to interpret the process of ESG analysis. Moreover, its cost will increase due to the high demand for resources and the relatively long training process.In the finance and economics domain, financial text mining is a crucial component of financial technology. Diverse financial text data, sourced from various materials like analyst reports and official company announcements, act as invaluable resources. They are used to analyze and forecast future market trends.Financial text mining serves as an essential tool, enabling the extraction of meaningful insights. These insights empower market participants to make well - informed decisions, giving them an edge in the ever - changing financial landscape(Liu et al., 2020)[8]. Therefore, to address these issues, FinBERT, a

pre - trained language model based on BERT (Bidirectional Encoder Representations from transformers), was developed. It can perform tasks such as classifying financial texts and conducting sentiment analysis through text. Moreover, it is capable of bidirectionally understanding and processing financial text data.The working method of the FinBERT model is based on two stages. The first stage is pre - training. It is similar to the BERT model and conducts large - scale analysis on a large amount of financial and economic data, such as data from FinancialWeb, Yahoo, and Reddit. During the pre - training process, this data is fed into the model. Subsequently, it will be fine - tuned according to specific training tasks. Meanwhile, the attention mechanism of FinBERT is the core of its Transformer architecture. It can understand by integrating the context of the text, unlike other machine learning algorithms that can only analyze text as a collection of independent words without considering grammar and word order.Meanwhile, the attention mechanism of FinBERT is the core of its Transformer architecture. It can understand by integrating the context of the text, unlike other machine learning algorithms that can only analyze text as a collection of independent words without considering grammar and word order (Devlin et al. 2019)[9].Finally, we can input the collected ESG data into the FinBERT model. Let the model analyze and interpret the data, and then conduct a quantitative assessment to intuitively understand the enterprise's performance in ESG, providing data support for relevant activities.

3. The Principles and Architecture of Transformers

Since they were first introduced in 2017, Transformers (Vaswani et al., 2017) have revolutionized the field of Natural Language Processing (NLP)[10]. Their ability to enable enhanced parallelization and effectively model long-range dependencies has made them a game-changer. Among the various models based on the Transformer architecture, BERT (Devlin et al., 2019) stands out as the most well - known[9]. BERT has achieved top - tier results in numerous benchmark tests and remains an essential baseline model for NLP tasks.BERT is based on the Transformer architecture, a deep - learning model. Its core component is the self - attention mechanism. By integrating Tokenization, various types of Embeddings, and output layers tailored to specific tasks, BERT can capture bidirectional context information within text. Unlike Recurrent Neural Networks (RNNs) that preceded it, which could only analyze text sequentially, either strictly from left - to - right or right - to - left, BERT can understand the meaning of a text by analyzing both its preceding and succeeding context. Essentially, BERT is constructed by piling up Transformer encoder layers (Vaswani et al., 2017)[10]. Each of these layers contains multiple self - attention “heads” which play a crucial role in processing information.The Transformer architecture is primarily composed of two components: the Encoder and the Decoder. The Encoder can be utilized for language representation learning tasks in Natural Language Processing (NLP), while the Decoder can be applied to tasks such as translation.The Encoder layer mainly consists of two parts: the multi - head self - attention mechanism and the feed - forward neural network.

The multi - head self - attention mechanism involves, after operations such as word embedding, taking the sequence of word vectors. These vectors are then mapped into three different spatial vectors: query, key, and value. It queries the degree of dependency and correlation among various vectors, and then captures the dependency relationships between different positions.

Multi-Head Attention Mechanism：

$$\text{Attention}(Q,K,V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Among them, Q, K, and V represent the query, key, and value vectors respectively. softmax is the softmax function, and d_k is the dimension of the vector.

The feed - forward neural network receives the results processed by the multi - head self - attention mechanism. Then it transforms the input data into a higher - dimensional space. After that, a function is used to introduce non - linear changes to the data, enabling the model to learn more complex patterns. The second step is to transform the data back to its original dimension and output the processed data. Feed - Forward Neural Network:

$$FFN(x)=\max(0,xW_1+b_1)w_2+b_2$$

Among them, W_1, b_1, $W_2$and b_2 represent the weights and biases of the two fully - connected layers respectively.

BERT learns language representations through

bidirectional training and task-specific fine-tuning, and applies them to various natural language processing tasks.Pre-training refers to the process of self-learning on a large amount of unlabeled data. Through these tasks, the model can learn the internal laws and patterns of language, thus providing strong support for its performance in subsequent specific tasks, such as text classification and question answering.Pre-training mainly consists of two components, one is the Masked Language Model (MLM), and the other is the Next Sentence Prediction (NSP).BERT eases the previously mentioned unidirectionality limitation by utilizing the “Masked Language Model” (MLM) for pre - training. The idea for this MLM pre - training objective is sourced from the Cloze task (Taylor, 1953)[11].MLM involves randomly masking or replacing some words in the input text and requires the model to predict the original content of these masked or replaced words. - This enables the model to utilize bidirectional context information to predict the masked words, thereby learning deeper semantic representations.The Next Sentence Prediction (NSP) task is to present one sentence followed by another, and the model needs to determine whether the two sentences are adjacent. This can help the model understand the coherence and causal relationships between sentences.After completing pre - training, the BERT model can be fine - tuned on various specific tasks, such as text classification, question - answering systems, named entity recognition, etc.In our multilingual experiments, we selected the paraphrase - multilingual - MiniLM - L12 - v2 model [20], a distilled variant. As noted by Afzal et al. (2025), it is smaller and more lightweight compared to BERT and similar models. This characteristic enables it to provide efficient performance in large - scale applications. Moreover, it is optimized for high - quality sentence embeddings and is applied in multilingual tasks such as semantic similarity assessment, clustering, and paraphrase detection[12].

4. Analysis of Experimental Results

Figure 1. Experimental Results

We based on the CSMAR database and used Python analysis tools to obtain the relevant word clouds and the rankings of the top - ranked themes.From the three aspects of ESG, the word cloud is analyzed as follows. In terms of environmental protection, there are keywords such as green plan, sustainability, and environmental protection. These reflect the enterprise's measures in environmental protection. The enterprise may reduce carbon emissions, save energy and resources by implementing green plans. It pays attention to environmental protection measures in the operation process and may adopt green technologies to reduce pollution. In terms of society, there are keywords such as corporate social responsibility, safety, and talent. Corporate social responsibility is a key element of the social factor in ESG. By fulfilling social responsibility, the enterprise can have a positive impact on employees' salaries, benefits, and public social activities. In terms of safety, many enterprises, such as those related to food safety and pharmaceutical safety, etc., are very important for consumers' physical health. In terms of corporate governance, there are keywords such as system, report, and management. These involve the internal management and planning of the company, which are crucial for the enterprise itself and its future development.

Table 1. Distribution of Keywords

Topic One	Bank	Shares	Awards	Interest	company	Risk Management	Economy	Finance	Consumer	Number
Topic Two	Peace	Development	Report	Speech	Key	Performance	Honor	Management	Strategy	Topic
Topic Three	Environmental Impact	Environmental Benefits	Technology	Manufacturing	Low-carbon	Environmental Protection	Carbon Peak	Carbon Neutrality	Photovoltaic	Energy Conservation
Topic Four	True Micro	Reduce	Small and micro	Decline	Unchanged	Unswerving	Acceptance	Accessibility	Consumer Protection	Boycott
Topic Five	Catalogue,	Architecture	Bottom line	Promote	Environment	Third Party	Complex and Intricate	Strength	Model	One-Stop
Topic	Limited	Longevity	Shenzhen	Shenzhen	China	Southern	Asia-Pacific	Shanghai	Asia-Pacifi	Asia

Six	Company			Development Bank		Daily			c Region	
Topic Seven	Inclusive	Green	ShenZhen City	Domestic and international	Culture	Special New	Science and Technology Innovation	Beginning of the Year	Life	Theme
Topic Eight	Sustainable	Bank Manager	Substantial	Steady	Improve	Adhere to	Compliance	Enthusiastic	Disclosure	Suggestion
Topic Nine	Greater Bay	Living Water	Upstream	Port	Downstream	Pool Item	Liquidity	Hanoi	Aquatic Products	Canal
Topic Ten	Organic	Product	Chemical Department	Health	Medical	Random Inspection	Biomedicine	Injection	Medicine	New Drug

Meanwhile, from the environmental perspective, keywords such as environmental impact, environmental benefits, technology, low - carbon, and environmental protection are highly relevant to the environmental aspect of ESG. From the social responsibility perspective, themes like "real small and micro - enterprises", "reduction", and "small and micro - enterprises" are involved. From the corporate governance perspective, themes related to "safety", "development", and "reports" are relevant.These are the analyses of the experiment.

5. The Principles and Architecture of LSTM (Long Short-Term Memory)

After analyzing ESG data using the BERT model, we will use the LSTM model to predict the future value of stocks.Before introducing the LSTM model, we can first introduce the RNN model. Essentially, the LSTM model is an improvement of the RNN model. According to Zargar (2021), convolutional neural networks are specifically designed to process grid - structured data like images, while recurrent neural networks are intended for processing sequence data[13].Recurrent neural networks (RNNs) fall under the category of deep - learning neural networks (DNNs). Unlike traditional feed - forward neural networks, RNNs maintain a hidden state when processing each input. This hidden state is then passed on to the next time step, enabling the model to remember previous information.As Zargar (2021) pointed out, a basic single - hidden - layer RNN is composed of one set of input, hidden, and output units, and its hidden units have feedback loops[13].”

The input layer is responsible for receiving external data and transforming it into a format suitable for network processing.

The hidden layer is the core part of an RNN. It is used to extract features and sequential information from the input data. It updates its own state by combining the current input with the hidden state from the previous time step, enabling it to capture long - term dependencies in the sequence.

Calculation formula:

$$S_t=f（U* x_t+W* s_{t-1}+B1)$$

The output layer generates the final prediction or output based on the output results of the hidden layer.

Calculation formula:

$$O_t=g(V* s_t+B2)$$

Lipton, Berkowitz, and Elkan (2015) indicated that the problems of vanishing and exploding gradients are encountered during the process of backpropagating errors over numerous time steps.When training neural networks, backpropagation calculates gradients for parameter updates. During backpropagation across time steps, two issues might arise. Vanishing gradients occur when gradients shrink, making it hard for models to learn long - range dependencies. Exploding gradients happen when gradients grow unbounded, causing model instability and potential training failure.Therefore, the LSTM model was developed by improving the RNN to address the problems of gradient explosion and gradient vanishing.The core improvement of LSTM lies in the introduction of the cell state and a series of gating mechanisms.As stated by Fang et al. (2021), LSTM is the first proposed RNN gating algorithm, and its corresponding loop unit, the LSTM unit, is composed of three gates: the input gate, the forget gate, and the output gate[15].

The input gate specifies the degree to which the current input information is written into the memory unit. The forget gate decides whether to "forget" the historical information stored in the unit, allowing the model to selectively remove irrelevant information.When processing sequential data, not all data is useful. The forget gate acts like a "filter", determining through calculation which old information can be discarded, enabling the model to focus on more relevant information and effectively handle long - term dependency issues in long sequential data. The output gate, in turn, determines the output of the information in the memory unit at the current

time step.The forget gate contains a sigmoid function, which acts like a valve controlling the proportion of information flow. It can generate values ranging from 0 to 1, thereby determining how much of the information in the memory cell from the previous time step will be retained at the current time step. The input gate, on one hand, uses a sigmoid function to decide which parts of the input information should be updated into the memory cell. On the other hand, it employs the TanH function to control the direction of information increase or decrease. The TanH function restricts the values within the range of -1 to 1 and generates candidate information for updating the memory cell.

Forget gate:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f$$

Input gate：

$$i_t=\sigma(W_i\cdot[h_{t-1}x_t]+b_i)$$

$$\widetilde{C}_t=\tanh(W_C\cdot[h_{t-1},x_t]+b_i)$$

Memory update:

$$C_t=f_t\cdot C_{t-1}+i_t\cdot\widetilde{C}_t$$

Output gate:

$$o_t=\sigma(W_o\cdot[h_{t-1},x_t])+b_0)$$

$$h_t=o_t\cdot\tanh(C_t)$$

However, LSTM still has some limitations. For example,In the research of Ismoilov and Jang (2018), overfitting is described as a situation where the model's training error decreases while the validation error increases because it memorizes training data rather than learning its distribution. Conversely, underfitting occurs when a model can't learn enough due to insufficient data[16].For this phenomenon, it can be addressed by increasing the amount of data and using regularization methods.A frequently used regularization technique is the autoencoder approach (Ismoilov & Jang, 2018)[16]. It's a neural network where the input and output layers have the same number of neurons, using unsupervised learning without needing labels during training.

The experimental procedure is described as follows: First, we loaded the 2023 ESG (Environmental, Social, and Governance) textual data of Ping An Bank (stock code: 000001.SZ) and performed natural language processing (NLP). The raw text data were tokenized using the Jieba library for Chinese word segmentation. A predefined stopwords list was applied to filter out irrelevant terms and tokens with a length of one character. Subsequently, sentiment scores were computed using the BERT model with the parameter configuration paraphrase-multilingual-MiniLM-L12-v2 to generate semantic embeddings. Wang et al. (2020) demonstrated that MiniLM reduces model size by 50% while maintaining competitive performance, making it ideal for multilingual applications with limited resources[17].This is the reason why we chose to use this parameter.We will normalize the "return" and "sentiment" features. This helps to eliminate the differences in dimensions and value ranges between these two features, making them comparable. At the same time, we will use the Min - Max Normalization method for normalization. Patro and Sahu (2015) indicate that Min - Max Normalization is a method utilized for linearly transforming the original data range. It is capable of preserving the correlations among the original data points. This normalization approach is relatively straightforward and can efficiently align the data to a predefined boundary. At the same time, normalization makes the data distribution more reasonable, reduces the occurrence of gradient anomalies, and ensures stability. Moreover, the more concentrated and regular data allows the model to learn the features in the data more effectively[18]. When performing normalization, we default to scaling the data between 0 and 1. The window size is a core parameter in time series forecasting, determining the length of historical data on which the model relies for each prediction. In this experiment, we set the window size to 20 as our time steps. That is, the model will analyze the data of the past 20 days to predict the results of the 21st day.We set the window size to 20. This is because the stock market is typically open for 20 to 22 trading days in a month. By setting it to 20, we can better capture the market volatility cycle.Next, we will define the LSTM model. Its input is set to 2, meaning there are two features in each time step of the input. The two features are return and sentiment score, where the return reflects temporal patterns in price movements, and the sentiment score captures the potential impact of market sentiment on stock prices.The hidden layer is set to 64, which determines the complexity of the features learned by the LSTM. The number of stacked layers is 2, enabling the model to learn more complex temporal patterns. The output is set to 1, as we are predicting the stock price for the next day. The dropout is set to 0.2, which means 20% of the data will be

randomly dropped out to prevent overfitting.Then, we will train the LSTM model. We select the AdamW optimizer and the MSE loss function to achieve effective model parameter updates and accurate evaluation of the training effect. The learning rate is set to 0.001, and the weight decay coefficient is set to 1e-4. These settings control the step size of parameter updates and constrain the model weights to prevent overfitting to noise and details. Setting the number of training epochs to 100 helps the model fully learn the complex features and patterns in the data. Meanwhile, an empty list named `train_losses` is initialized to record the loss values, which is used to determine whether the model converges and whether the training process is stable.Then, set the model to evaluation mode, use the test data for prediction, and convert the prediction results into a NumPy array.Then, evaluate the performance of the model.

```
[**********100%**********]  1 of 1 completed
Epoch 0, Loss: 0.130901
Epoch 10, Loss: 0.025671
Epoch 20, Loss: 0.022769
Epoch 30, Loss: 0.024299
Epoch 40, Loss: 0.023070
Epoch 50, Loss: 0.023103
Epoch 60, Loss: 0.023165
Epoch 70, Loss: 0.022933
Epoch 80, Loss: 0.022870
Epoch 90, Loss: 0.022733

📊 Prediction Evaluation Metrics:
📌 MAE (Mean Absolute Error): 0.107344
📌 RMSE (Root Mean Squared Error): 0.163182
```

Figure 2. The Prediction Evaluation Metrics

Calculate the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE), where RMSE is obtained by taking the square root of the Mean Squared Error (MSE). Keep six decimal places for the results. After calculation, the value of MAE is 0.107344 and the value of RMSE is 0.163182.

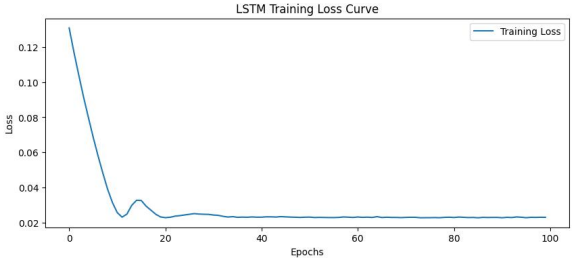

Figure 3. Training Loss Curve

Finally, plot the training loss curve.Create a figure with a width of 10 and a height of 4, with the number of training epochs on the x-axis and the loss values on the y-axis. Use `plt.plot` for plotting. Finally, use `plt.show` to display the figures, and observe the changing trend of the loss values to determine whether the model converges.In the initial 0 - 20 epochs, the loss value initially decreased from 0.12 to 0.02. There was a rebound in the middle, yet it still ended up decreasing. From 20 to 40 epochs, it was a process of overall decline with fluctuations. After 40 epochs, it remained consistently stable at around 0.02.Therefore, this model is convergent.

6. Conclusion

This experiment mainly focuses on predicting the stock price using the ESG textual data of Ping An Bank (stock code: 000001.SZ) in 2023. The purpose of the experiment is to analyze the textual data and construct a model to predict the stock price. The experimental process is mainly divided into five steps.

In the first step, the original text is first segmented using the Jieba library, and irrelevant words are filtered out using stop words. Then, a BERT model with the parameter paraphrase-multilingual-MiniLM-L12-v2 is used to calculate the sentiment scores, which are then normalized. The normalization method used is the Min-Max normalization method.

In the second step, an LSTM model is set up. The window size is set to 20, and the data of the previous 20 days are used to predict the result of the 21st day. The input is set to 2, which means there are two features: the return and the sentiment score. The hidden layer has 64 units, the number of stacked layers is 2, the output is 1, and the dropout rate is 0.2.

In the third step, regarding model training, the AdamW optimizer and the MSE loss function are selected. The learning rate is set to 0.001, the weight decay coefficient is set to 1e-4, the number of training epochs is set to 100, and a list named "train_losses" is initialized to record the loss values.

In the fourth step, model evaluation is carried out. The test data are used for prediction, and the results are converted into a NumPy array to calculate the MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error).

In the fifth step, the results are visualized. The abscissa is set as the number of training epochs, and the ordinate is set as the loss value, and a training loss curve is plotted.

Finally, the experimental results are obtained. The MAE value is 0.107344, and the RMSE value is 0.163182. In the initial 0 to 20 epochs of the model, the loss value decreases from 0.12 to 0.02. There is a rebound in the middle, but it still

decreases. From 20 to 40 epochs, it generally decreases with fluctuations. After 40 epochs, it stabilizes at around 0.02. Therefore, the model converges.

Not only can the sentiment of ESG data be utilized for stock price prediction, but also AI models can be employed for stock trading. By inputting a company's financial reports, news content, and market sentiment into the model, the AI can then select stocks and conduct autonomous trading. Furthermore, stock research can be carried out by promptly collecting and analyzing news. The Chinese market is likely to be significantly influenced by policy factors. News content can be continuously fed into the model, after which an analysis of the market stock prices can be conducted to determine what kind of news will be beneficial to which sectors. Once a piece of news is published, the model can automatically collect and analyze it, and then buy specific stocks. This could serve as a new direction for future research. However, there are many aspects that can be improved in this experiment. There is a phenomenon of insufficient data, and the amount of data content is relatively small.The data imported into this model is that of Ping An Bank. The impact degree of ESG data varies for different types of enterprises, so experiments should be conducted on different enterprises. Only a single LSTM model is selected for the model without conducting comparisons. The setting of the window size to 20 is only based on experience and lacks more solid evidence.

References

[1]Friede, G., Busch, T., & Bassen, A. (2015). ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. Journal of Sustainable Finance & Investment, 5(4), 210 - 233. DOI: 10.1080/20430795.2015.1118917

[2]Broadstock, D. C., Chan, K., Cheng, L. T., & Wang, X. (2021). The role of ESG performance during times of financial crisis: Evidence from COVID - 19 in China. Finance Research Letters, 38, 101716. Retrieved from https://doi.org/10.1016/j.frl.2020.101716

[3]Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. The Journal of Finance, 72(4), 1785–1824. https://doi.org/10.1111/jofi.12505

[4]Henriksson, R., Livnat, J., Pfeifer, P., & Stumpp, M. (2019). Integrating ESG in Portfolio Construction. The Journal of Portfolio Management, 45(4), 67–81.

[5]Zhou, G., Liu, L., & Luo, S. (2022). Sustainable development, ESG performance and company market value: Mediating effect of financial performance. Business Strategy and the Environment, 31(7), 3371–3387. https://doi.org/10.1002/bse.3089

[6]Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.

[7]LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436 - 444. Available at: https://doi.org/10.1038/nature14539

[8]Liu, Z., Huang, D., Huang, K., Li, Z., & Zhao, J. (2020). FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20) Special Track on AI in FinTech, 4513 - 4519.

[9]Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

[10]Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. In Advances in Neural Information Processing Systems (pp. 5998 - 6008).

[11]Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. Journalism Bulletin, 30(4), 415–433.

[12]Afzal, T., Abdul Rauf, S., Malik, M. G. A., & Imran, M. (2025). Fine-Tuning QurSim on Monolingual and Multilingual Models for Semantic Search. Information, 16 (2), 84. https://doi.org/10.3390/info16020084

[13]Zargar, S. A. (2021). Introduction to Sequence Learning Models: RNN, LSTM, GRU. Preprint. DOI: 10.13140/RG.2.2.36370.99522

[14]Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. arXiv preprint arXiv:1506.00019v4.

[15]Fang, W., Chen, Y., & Xue, Q. (2021).

Survey on Research of RNN-Based Spatio-Temporal Sequence Prediction Algorithms. Journal on Big Data, 3(3), 98-110. DOI:10.32604/jbd.2021.016993

[16]Ismoilov, N., & Jang, S. - B. (2018). A Comparison of Regularization Techniques in Deep Neural Networks. Symmetry, 10 (11), 648.

[17]Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. arXiv preprint arXiv:2002.10957. Retrieved from https://arxiv.org/abs/2002.10957

[18]Patro, S. G. K., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. arXiv:1503.06462 [cs.OH]. https://doi.org/10.48550/arXiv.1503.06462