JSDP: Journal of Spatial Development and Policy, 2025. 3(6): 39-52.

PANYAPAT
ปัญญาพัฒน์

JSDP
Journal of Spatial Development and Policy
https://so16.tci-thaijo.org/index.php/JSDP/index

Research Article

An Ensemble Multi-Model Approach for Stock Return Forecasting Integrating Emotion-Driven Feature Extraction and Time Series Prediction Based on Variational Autoencoders (VAE) and Temporal Fusion Transformers (TFT)

Minghui Zhong[1*] and Liang Shu Hui[2]

[1] *Student, International College for Interdisciplinary Studies, Payap University, Thailand*
[2] *International College for Interdisciplinary Studies, Payap University, Thailand*
[*] *Corresponding author. E-mail: mhzhong2022@126.com*

ABSTRACT

This article aimed to study: 1) the impact of investor sentiment on stock returns in the automotive industry, and 2) the development and validation of the Adaptive Sentiment-Enhanced Temporal Regression Model (ASTRM), an ensemble model integrating Variational Autoencoders (VAE), Temporal Fusion Transformers (TFT), and Ordinary Least Squares (OLS) regression. This quantitative research utilizes historical secondary data from 269 companies in the Chinese automotive industry. Data was collected from the China Stock Market & Accounting Research (CSMAR) database, comprising historical stock prices and a sentiment index derived from social media posts from May 23, 2019, to September 21, 2022. The statistical analysis involved: 1) using VAE for feature extraction and data denoising, 2) employing TFT with LSTM and multi-head attention mechanisms for time-series forecasting, and 3) applying OLS regression to optimize the final predictions. Model performance was evaluated using metrics such as accuracy and adjusted R-squared. The findings indicate that the ASTRM model achieved a prediction accuracy of 91.12%, significantly outperforming the traditional LSTM model (70.63%) and a model without denoising (77.48%). Furthermore, investor sentiment was found to have a statistically significant influence on stock returns (p-values < 0.001). The ASTRM also demonstrated faster convergence and a lower final loss value, enhancing overall predictive performance.

Keywords: Investor Sentiment, Stock Returns, Automotive Industry, Multi Model

Introduction

In financial markets, investor sentiment acts like an ocean tide, profoundly affecting market volatility and investment decisions (Shleifer, 2000). With the popularity of social media, sentiment analysis has become a vital tool for studying investor behavior. The massive amount of data from these platforms provides an unprecedented opportunity to construct and analyze sentiment indices, enabling a deeper understanding of the psychological shifts of market participants and their impact on market performance (McGurk et al., 2020).

Received: 2 July 2025;
Revised: 12 December 2025;
Accepted: 18 December 2025
© 2025 JSDP: Journal of Spatial Development and Policy
ISSN: 2985-220X (Online)

Several academic studies have shown that during periods of high market volatility, changes in sentiment indices can significantly predict short-term returns (Andleeb & Hassan, 2023). Similarly, there is supporting evidence that as the sentiment index rises, market trading volume also tends to increase significantly, reflecting that market liquidity can serve as a sentiment indicator (Baker & Stein, 2004).

In this context, this article uses trading data from the China Stock Market & Accounting Research (CSMAR) database, a comprehensive and well-regarded academic resource. The sample consists of stocks from the Chinese automotive industry for the period of May 23, 2019, to September 21, 2022, to study how sentiment indices constructed from social media data affect stock returns. This approach not only improves the accuracy and reliability of the prediction models but also offers a new perspective for understanding investor behavior in the financial market. Based on the empirical results that demonstrate a significant correlation between the sentiment index and stock returns in the automotive industry (CSMAR, 2025). we aim to provide a more accurate forecasting tool for market participants and valuable insights for policymakers, enabling them to better seize investment opportunities and mitigate risks in a complex and ever-changing market environment.

Research Objectives

1. To investigate the impact of investor sentiment on stock market returns in the automotive industry by constructing a comprehensive sentiment index from social media data and analyzing its influence on market volatility and investment decisions.

2. To develop and validate the Adaptive Sentiment-Enhanced Temporal Regression Model (ASTRM) that integrates Variational Autoencoders (VAE), Temporal Fusion Transformers (TFT), and Ordinary Least Squares (OLS) regression to achieve superior stock return prediction accuracy compared to traditional methods.

Research Methodology

This study employs a quantitative research design grounded in the principles of behavioral finance and machine learning-driven financial forecasting. The research design comprises the following key stages.

1. Conceptual Framework and Hypothesis

The conceptual framework posits that investor sentiment, as a behavioral factor, is a significant predictor of stock market returns, an area where traditional financial theories may be insufficient. The central hypothesis (H_1) of this research is formulated as:

H_1: Investor sentiment, derived from social media data, has a statistically significant impact on stock return predictions in the Chinese automotive industry.

2. Data Source and Preprocessing

This study utilizes secondary data from the China Stock Market & Accounting Research (CSMAR) database, a well-established and comprehensive source for academic research on China's financial markets.

2.1 Sample: The research focuses on companies within the Chinese automotive industry.

2.2 Period: The dataset covers daily trading data from May 23, 2019, to September 21, 2022.

2.3 Sample Selection Process: To ensure the quality of the dataset and mitigate potential biases, a systematic screening process was implemented:

- Inclusion Criteria: The initial sample included all companies classified under the automotive industry in the CSMAR database as of April 30, 2019.
- Exclusion Criteria: The following were excluded: (1) firms in the bottom 30% of total market value, to reduce the influence of illiquid stocks; and (2) firms with significant missing data, a short listing history, or those that were delisted during the study period, in order to control for survivorship bias.

2.4 Final Sample: After the screening process, the final sample consisted of 269 companies for analysis.

3. Variable Index Design

The primary variables for this study were constructed as follows:

Investor Sentiment Index: This index was formulated using data on the daily volume of bullish posts, bearish posts, and total posts for each stock, as sourced from the CSMAR database. These raw counts were transformed into firm-level and industry-level sentiment indices using the logarithmic formulas specified in Equations (4) through (7).

Stock Return: Stock returns were calculated based on the daily change in the closing price at both the individual firm level (Equation 8) and the aggregate industry level (Equation 9).

4. The ASTRM Research Model

This study proposes and validates an ensemble model named the Adaptive Sentiment-Enhanced Temporal Regression Model (ASTRM). The model's architecture consists of three sequential stages: Feature Extraction and Denoising: A Variational Autoencoder (VAE) is employed to process the high-dimensional and noisy input data. The VAE maps the data into a low-dimensional latent space, effectively extracting key features while removing noise, thereby providing a cleaner input for the subsequent forecasting stage.

Time-Series Forecasting: The denoised, low-dimensional features from the VAE are integrated with other time-series variables and fed into a Temporal Fusion Transformers (TFT) model. Leveraging its LSTM and multi-head attention mechanisms, the TFT model captures complex short-term and long-term temporal dependencies to generate stock return predictions.

Prediction Optimization: The predictions generated by the TFT model are then utilized as a new independent variable in an Ordinary Least Squares (OLS) regression model. This final stage further refines and optimizes the forecast, producing the final, high-accuracy stock return prediction.

The Conceptual Framework

The conceptual framework of this study is constructed based on the principles of behavioral finance and machine learning-driven financial forecasting. The central hypothesis posits that investor sentiment, extracted from social media, significantly influences stock returns in the automotive industry, here are three core components of the framework:

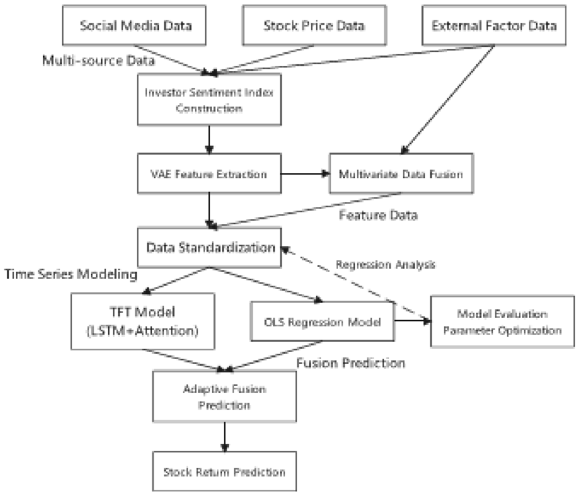

Figure 1 Conceptual Framework

Research Results

Research has found that:

1. The Impact of Investor Sentiment on Stock Returns The analysis of the relationship between the constructed sentiment indices and stock returns confirmed that both bullish and bearish sentiment are statistically significant predictors, supporting the study's main hypothesis.

1.1 To provide a more intuitive view of the characteristics of the dataset, this paper conducted a statistical analysis of the fundamental features of stocks, as shown in Table 3. This analysis includes several key characteristics of stocks, including opening price (Mopnprc), closing price (Mclsprc), trading amount (Mnvaltrd), total market value (Msmvosd), trading days (Ndaytrd), net return (Mretnd), market type (Markettype), cumulative return (Cmretmdeq), and turnover rate (Turn). Among these, the average opening price and closing price are 11.83 and 11.94, respectively, indicating that the market price is generally stable. The large standard deviation between trading amount and total market value indicates significant differences in company size and trading volume in the market. The means of net return and cumulative return are close to 0, suggesting that returns and risks in the market are generally low, but there is volatility. The high mean and large standard deviation of turnover rate indicate that some stocks exhibit high trading activity and speculation.

Table 1 Statistical analysis of stock basic characteristics

指标	Mopnprc	Mclsprc	Mnvaltrd	Msmvosd	Ndaytrd	Mretnd	Markettype	Cmretmdeq	Turn
count	2029.000000	2029.000000	2.029000e+03	2.029000e+03	2029.000000	2029.000000	2029.000000	2029.000000	2029.000000
mean	11.832225	11.942586	1.860269e+09	7.678786e+06	20.292262	0.016188	4.258748	0.021899	139.464577
std	9.686918	9.874311	3.203237e+09	2.439384e+07	2.167048	0.124735	4.820708	0.074787	220.117353
min	0.444000	0.449000	1.376408e+06	1.077600e+05	5.000000	-0.617333	1.000000	-0.049598	0.555696
25%	5.470000	5.480000	4.103375e+08	1.406339e+06	19.000000	-0.055266	1.000000	-0.022471	33.787970
50%	9.340000	9.470000	8.685270e+08	2.840820e+06	21.000000	-0.003249	4.000000	-0.006254	72.619812
75%	15.180000	15.250000	1.939247e+09	5.164113e+06	22.000000	0.068655	4.000000	0.015512	152.704214
max	85.980000	94.980000	4.466420e+10	3.257772e+08	23.000000	0.963201	16.000000	0.222753	2811.197135

This study also analyzed a dataset of investor sentiment characteristics, as shown in Table 1. Analysis shows that the number of positive posts (BullishPosts) and negative posts (BearishPosts) related to investor sentiment fluctuates significantly, with mean values of 182.46 and 216.11, respectively, and maximum values of 4710 and 7258, indicating that market sentiment may have a significant impact on fund returns during certain periods. The number of neutral posts is relatively small, with an average of 41.61, indicating that the emotions of most market participants show a clear tendency.The average number of IPOs is 16.95, with a maximum of 37, indicating that the frequency of new stock issuance activities during a specific period may also have an impact on market sentiment and fund returns. The mean of the Bull_SI index is -0.11, indicating a relatively cautious overall market sentiment, while the mean of the Bea_SI index is 0.11, indicating that market concerns about risk may suppress fund returns. The average total number of posts (sum) is 440.18, but the maximum value is as high as 13185, suggesting that the intensity of market discussions during periods of high or low sentiment may be related to fluctuations in fund returns. Although the mean of the difference (bul_SI-diff) of the bullish sentiment index 2 is close to 0, its standard deviation and extreme values indicate frequent fluctuations in market sentiment, which may directly affect the returns of equity funds.

Table 2 Statistical analysis of emotion characteristics of Stockholders

指标	BullishPosts	NeutralPosts	BearishPosts	ipo_num	bul_SI2	Bea_SI2	bul_SI2_diff	sum
count	2029.000000	2029.000000	2029.000000	2029.000000	2029.000000	2029.000000	2029.000000	2029.000000
mean	182.461311	41.610646	216.107935	16.952193	-0.112871	0.112871	0.001025	440.179892
std	355.569367	77.110382	408.398237	9.179945	0.448405	0.448405	0.475330	812.752121
min	0.000000	0.000000	0.000000	6.000000	-1.832581	-2.484907	-2.426883	0.000000
25%	45.000000	9.000000	48.000000	11.000000	-0.387381	-0.124298	-0.273569	104.000000
50%	83.000000	19.000000	98.000000	15.000000	-0.136576	0.136576	0.003214	201.000000
75%	168.000000	41.000000	208.000000	27.000000	0.124298	0.387381	0.275446	411.000000
max	4710.000000	1217.000000	7258.000000	37.000000	2.484907	1.832581	2.890372	13185.000000

1.2 The Influence of Stock Investor Sentiment on Stock Return Forecast

1.2.1 Predictions Using the Traditional OLS Model

As shown in Table 2, the experimental results indicate that when using only the stock features of the OLS model for prediction, although the coefficients of each feature in the model are different, their statistical significance performance is not ideal. For example, the p-values of transaction amount (Trdmnt) and other key variables are relatively large, indicating that these features have a weak influence in the model. In addition, although the R-squared value shows that the model can explain some changes in returns, it is evident that the contributions between features are not significant, and there is still significant room for improvement in the overall explanatory power of the model. This reflects that there may be some issues with the fitting effect of the model in terms of variable selection.

Table 3 shows the importance analysis of model features when only stock characteristics are used for forecasting

variable	Coefficients	Std. Error	t value	Pr(>|t|)	R-squared	Adjusted R-squared
Trdmnt	0.009522.	0.017068	0.558	0.576990	0.6912	0.6499
Mopnprc	0.011219	0.013115	0.855	0.392420	0.6912	0.6499
Mclsprc	0.007280	0.011114	0.655	0.512551	0.6912	0.6499
Mnvaltrd	0.512551	0.009426	1.660	0.097136	0.6912	0.6499
Msmvosd	0.011237	0.010948	1.026	0.304812	0.6912	0.6499
Ndaytrd	0.011174	0.013902	0.804	0.421602	0.6912	0.6499
Mretnd	0.011752	0.011069	1.062	0.288521	0.6912	0.6499
Markettype	0.011492.	0.016346	0.703	0.482111	0.6912	0.6499
Cmretmdeq	0.012977	0.016193	0.801	0.422979	0.6912	0.6499
Turn	0.016655.	0.012769	1.304	0.192291	0.6912	0.6499

When using the emotional characteristics of stock investors for prediction, the situation is similar. Although some coefficients in the emotional variables show some significance, overall, the fitting effect of the model is still not satisfactory. The interaction between emotional features and the limited explanatory power of the model have resulted in no significant improvement in the predictive performance of the model.

Table 4 Analysis of Feature Importance in Model Prediction Using Investor Sentiment Characteristics

variable	Coefficients	Std. Error	t value	Pr(>|t|)	R-squared	Adjusted R-squared
BullishPosts	0.001256	0.022794	0.055	0.956151	0.5203	0.5026
NeutralPosts	0.014597	0.013212	1.105	0.269356	0.5203	0.5026
BearishPosts	0.005568	0.027055	0.206	0.836976	0.5203	0.5026
ipo_num	0.010407	0.019960	0.521	0.602140	0.5203	0.5026
bul_SI	0.004580	0.018759	0.244	0.807155	0.5203	0.5026
bea_SI	0.005309	0.011003	0.482	0.629517	0.5203	0.5026
bul_diff	0.005566	0.029992	0.186	0.852797	0.5203	0.5026

1.2.2 Predictions Using the Astrm Model

The experimental results indicate that the performance of the first two sets of features in the OLS model is not ideal, with weak feature correlation and model interpretability not meeting expectations. This indicates the need for further introduction of more complex models to capture potential feature interaction relationships. When using only stock features for yield prediction in the ASTRM model, as shown in Table 1, variables such as trading amount (Trdmnt), opening price (Mopnprc), closing price (Mclsprc), market value (Mnvaltrd), and residual market value (Msmvosd) all contribute significantly to the model's predictive ability, and their coefficients are statistically significant (p-values are extremely low, indicating that the impact of features on the model is very significant). Especially, the coefficients of the opening price and closing price are 0.039280 and 0.033946, respectively, indicating that these characteristics have a significant impact on stock returns. In addition, market type and turnover also have a significant impact on the prediction results of the model, and their coefficients are significant, indicating that these variables cannot be ignored when predicting stock returns. From the overall performance of the model, the Adjusted R-squared value is 0.7755, indicating that the model can explain most of the changes in returns. These results indicate that the model already has strong predictive ability when using only these stock features, but in practical applications, combining more emotional factors and other non stock features may further enhance the performance of the model.

Table 5 Analysis of the Importance of Model Features When Predicting Only with Stock Features

variable	Coefficients	Std. Error	t value	Pr(>|t|)	R-squared	Adjusted R-squared
Trdmnt	0.048029	0.008492	5.656	1.77e-08	0.7688	0.7755
Mopnprc	0.039280	0.002911	13.493	8.63e-40	0.7688	0.7755
Mclsprc	0.033946	0.002636	12.876	1.62e-36	0.7688	0.7755
Mnvaltrd	0.016241	0.002651	6.127	1.07e-09	0.7688	0.7755
Msmvosd	0.016240	0.003738	4.344	1.47e-05	0.7688	0.7755
Ndaytrd	0.012323	0.005723	2.153	3.14e-02	0.7688	0.7755
Mretnd	0.044647	0.005723	9.135	1.55e-19	0.7688	0.7755
Markettype	0.034045	0.003621	9.402	1.40e-20	0.7688	0.7755
Cmretmdeq	0.038323	0.006507	5.890	4.52e-09	0.7688	0.7755
Turn	0.010823	0.002255	4.799	1.71e-06	0.7688	0.7755

According to the analysis results in Table 5, the importance of investor sentiment in predicting stock returns has been significantly reflected. The analysis includes multiple emotion related variables, such as BullishPosts, Neutral Posts, BearishPosts, and their corresponding emotion index variables (P, N, P_diff), all of which show significant statistical significance in the model, indicating their important impact on the prediction results. Especially, the coefficients of NeutralPosts and BearishPosts are 0.039280 and 0.033946, respectively, with extremely high significance, indicating that these emotional variables have strong explanatory power for predicting stock returns. In addition, the Multiple R-squared value of the model reached 0.9541, and the Adjusted R-squared value was 0.8653, indicating that the model can explain most of the changes in returns. These results fully demonstrate that incorporating investor sentiment into the model can significantly improve the accuracy of stock return prediction, and sentiment data is one of the key factors affecting the prediction results.

Table 6 Analysis of Feature Importance in Model Prediction Using Investor Sentiment Characteristics

variable	Coefficients	Std. Error	t value	Pr(>|t|)	R-squared	Adjusted R-squared
BullishPosts	0.048029	0.008492	5.656	1.77e-08	0.9541	0.8653
NeutralPosts	0.039280	0.002911	13.493	8.63e-40	0.9541	0.8653
BearishPosts	0.033946	0.002636	12.876	1.62e-36	0.9541	0.8653
ipo_num	0.016241	0.002651	6.127	1.07e-09	0.9541	0.8653
bul_SI	0.016240	0.003738	4.344	1.47e-05	0.9541	0.8653
bea_SI	0.044647	0.004888	9.135	1.55e-19	0.9541	0.8653
bul_diff	0.010823	0.002255	4.799	1.71e-06	0.9541	0.8653

2. The experimental results and model validation confirm that the developed Adaptive Sentiment-Enhanced Temporal Regression Model (ASTRM) demonstrates significantly superior performance and accuracy compared to the benchmark models.

2.1 Accuracy Comparison

The research results indicate that the addition of emotional factors further amplifies the performance differences between models, as shown in Figure 2. The performance of the LSTM model is relatively stable, but its accuracy improvement is slow in the early stages of training, ultimately reaching 70.63%. This is because the LSTM model can gradually learn the basic patterns in the data when processing time-series data after adding emotional factors, but its structure is relatively simple and cannot effectively deal with noise and complex features in the data, which limits its predictive performance. This further indicates that the LSTM model has insufficient performance in integrating complex emotional information and has not fully explored the potential impact of emotional factors on stock return prediction. In contrast, models without data denoising performed better than LSTM in the early stages of training, with an accuracy rate rapidly increasing to nearly 77.48%. However, in the later stages, the accuracy tended to stabilize, showing sensitivity to data noise and limitations in model learning. The ASTRM model performed the best throughout the entire training process, with an accuracy rate rapidly increasing from 14.84% in the initial stage to 91.12% and stabilizing at around 90.15%. This significant improvement indicates that the ASTRM model can not only handle complex time-series data, but also effectively integrate user emotional information, significantly improving prediction accuracy. By combining multi head attention mechanism and deep learning network, the ASTRM model can effectively capture key patterns in data and suppress the interference of noise on prediction.

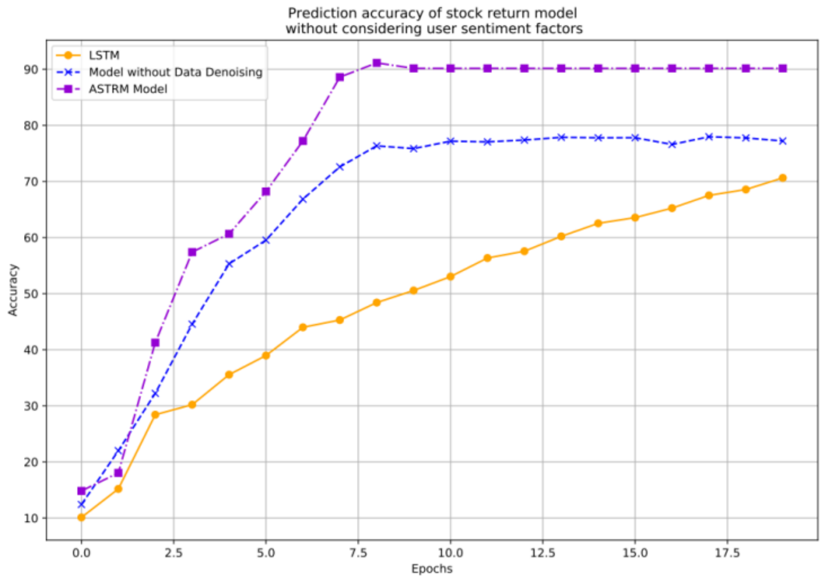

Figure 2 Comparison of Yield Prediction Accuracy with User Emotion Index Added

When considering user emotional factors, the loss values of each model are shown in Figure 2. The ASTRM model (dark purple curve, dotted line) once again demonstrated excellent performance, with its loss value rapidly decreasing from 0.77 to 0.02 and remaining stable. Compared with the first figure, the loss value of the ASTRM model decreases more rapidly and stably after adding emotional factors, indicating that emotional factors provide additional useful information for the model, significantly improving its convergence speed and performance. The model without data denoising (blue curve, dashed line) also showed a decrease in loss value after adding emotional factors, but ultimately remained around 0.36, indicating that the influence of noise was not completely removed. The LSTM model (orange curve, solid line) showed a slight decrease in loss value after considering emotional factors, but the rate of decrease remained slow, ultimately remaining at around 0.29, demonstrating its limited ability in integrating emotional factors.

2.2 Learning Efficiency (Convergence and Loss)

In addition to accuracy, the ASTRM model demonstrated superior learning efficiency (as shown in Figure 3 The ASTRM model (dark purple curve, dotted line) once again demonstrated excellent performance, with its loss value rapidly decreasing from 0.77 to 0.02 and remaining stable. Compared with the first figure, the loss value of the ASTRM model decreases more rapidly and stably after adding emotional factors, indicating that emotional factors provide additional useful information for the model, significantly improving its convergence speed and performance. The model without data denoising (blue curve, dashed line) also showed a decrease in loss value after adding emotional factors, but ultimately remained around 0.36, indicating that the influence of noise was not completely removed. The LSTM model (orange curve, solid line) showed a slight decrease in loss value after considering emotional factors, but the rate of decrease remained slow, ultimately remaining at around 0.29, demonstrating its limited ability in integrating emotional factors.

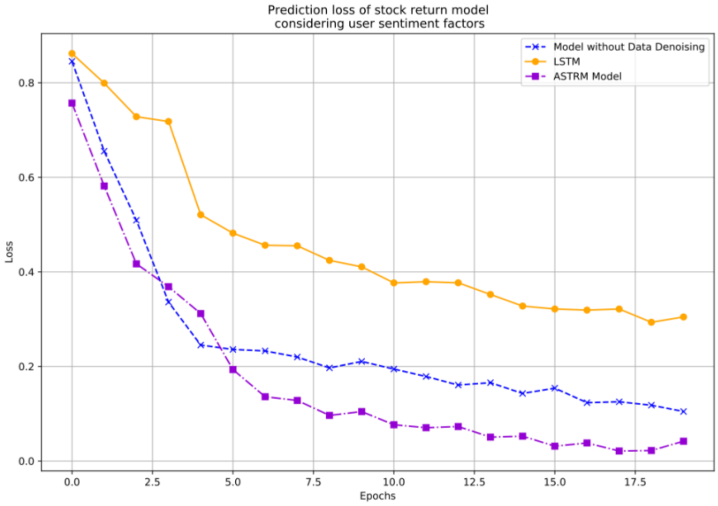

Figure 3 Comparison of Profit Prediction Loss Values with User Emotion Index Added

From the above analysis, it can be seen that the ASTRM model exhibits faster convergence speed and lower loss values in both scenarios, especially after incorporating emotional factors, its advantages are more pronounced. This reflects the powerful ability of the ASTRM model in processing complex data and effectively integrating diverse information. Although the LSTM model and the model without data denoising have improved, their performance in integrating data noise and emotional factors is significantly inferior to the ASTRM model.

Discussions

In today's rapidly developing financial market, investor sentiment plays an increasingly important role in predicting stock market returns (Zhou et al, 2024). With the advancement of machine learning technology, more and more research is focusing on how to more effectively integrate emotional factors into stock return prediction models to improve the accuracy and robustness of predictions (Mahmud et al, 2024). This study delves into the impact of investor sentiment on stock market returns in the automotive industry by introducing the ASTRM model, and compares it with traditional LSTM models and models without data denoising.

1. This study found that incorporating investor sentiment factors into the prediction model significantly improved the accuracy of the model. Similarly, Chen C. (2022) studied investor sentiment and found that the constructed investor sentiment index could actually explain the returns of the Thai stock market. The ASTRM model, through the combination of multi-head attention mechanism and deep learning network, can effectively capture complex patterns in emotional data and suppress noise in the data, ultimately demonstrating excellent performance in stock return prediction. In contrast, although LSTM models perform well in processing time-series data, their prediction accuracy is relatively low due to their failure to fully integrate emotional information. In addition, models that have not undergone data denoising processing have significantly inferior predictive performance compared to the ASTRM model under the influence of noise.

2. This study also found that the addition of emotional factors not only improves the prediction accuracy of its model but also accelerates its convergence speed. The ASTRM model demonstrated a

faster loss reduction rate and lower final loss value in the early stages of training, indicating significant advantages in multivariate information processing and complex data feature capture. These findings emphasize that the integration of emotional information in financial markets can provide more reliable support for investment decisions, thereby improving overall investment returns.

3. Although the ASTRM model performs well in handling complex data, its performance still depends on the quality and accuracy of emotional data. Future research can further explore how to optimize the acquisition and processing of emotional data to maximize the robustness and generalization ability of the model (Yasser et al, 2024). In addition, the differences in emotional impact across different industries can also serve as a direction for future research, providing more targeted recommendations for the development of industry investment strategies.

Originality and Body of Knowledge

The results of the study can be summarized into new knowledge as follows:

1. The study confirmed the impact of investor sentiment on stock market returns and proposed a new investor sentiment index.

2. The study utilized VAE in the model for feature extraction of stock market data, effectively reducing the dimensionality of the data and removing noise, generating key low-dimensional features, and providing high-quality input for subsequent time series predictions.

3. The study introduced the TFT model, combined with the LSTM network and a multi-head attention mechanism, to capture complex dependency relationships in multivariate time series data, thereby enhancing the model's performance in capturing both short-term and long-term market fluctuations and significantly improving prediction accuracy.

4. Through experimental verification, the model framework combining VAE, TFT, and OLS regression achieved an accuracy of 91.12% in predicting stock returns in the actual automotive industry. Compared with traditional LSTM models and models without data denoising, it significantly improved prediction performance.

5. This study not only demonstrated the importance of investor sentiment information in financial market forecasting but also highlighted the significant advantages of the proposed model framework in processing complex financial data, improving prediction accuracy, and enhancing model robustness, thereby providing new research directions for the further development of intelligent financial systems.

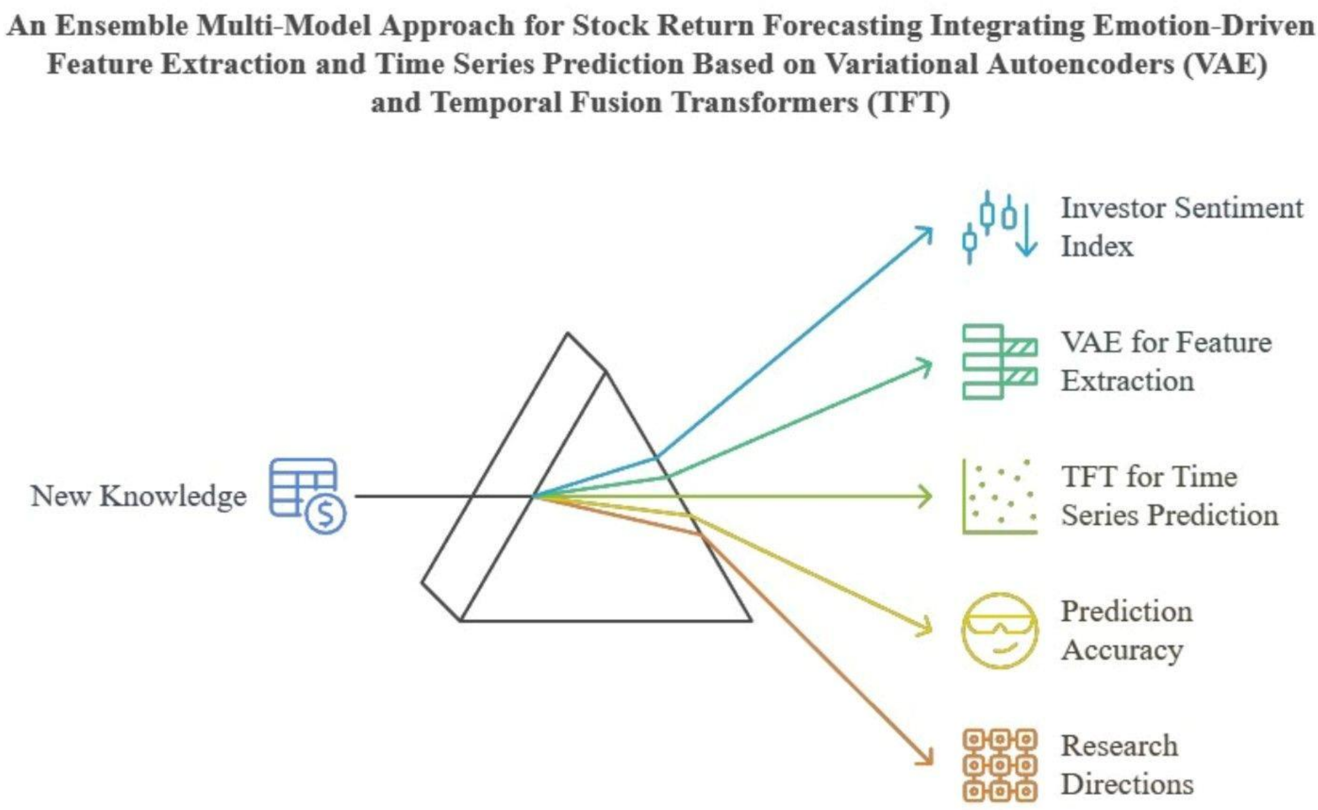

Figure 4 New Knowledge

Conclusions

The ASTRM model proposed in this study achieves deep integration of investor sentiment information and stock market return data by combining multi-head attention mechanism and deep learning network. The ASTRM model has shown significant advantages in capturing complex data features and suppressing noise, which not only makes it much more accurate than traditional LSTM models, but also significantly accelerates the convergence speed of the model. This model can effectively utilize emotional data, enhance sensitivity to market fluctuations, and provide more accurate earnings forecasts. In addition, the multiple information processing capabilities of the ASTRM model enable it to exhibit excellent robustness and adaptability when facing complex time-series data in financial markets. With the continuous development of data analysis technology, the ASTRM model is expected to play a key role in a wider range of financial applications, promoting further innovation and progress in intelligent finance.

Recommendations

1. Policymaking Recommendations

Based on the findings of this study, which demonstrate the significant impact of investor sentiment on stock market returns through the ASTRM model, we propose the following policy recommendations:

1.1 Corporate-Level Recommendations: Companies should establish comprehensive information disclosure frameworks that ensure timely, complete, and reliable communication with investors. This includes developing dedicated digital communication departments that can effectively manage online public relations, guide market narratives during critical events, and correct misinformation that may lead to excessive market volatility. Companies should also integrate sentiment analysis tools into their investor relations strategies to better understand market perceptions and proactively address negative sentiment.

1.2 Investor-Level Recommendations: Individual investors should enhance their financial literacy by incorporating sentiment analysis into their investment decision-making processes. Investors should develop the capability to distinguish between sentiment-driven market movements and fundamental value changes, utilizing both technical and sentiment indicators to make more informed decisions while avoiding emotional investment traps.

1.3 Regulatory Recommendations: Financial regulators should recognize the increasing importance of sentiment data in market dynamics and develop regulatory frameworks that account for sentiment-driven volatility. This includes establishing monitoring systems that can detect unusual sentiment patterns and developing guidelines for the use of AI-driven sentiment analysis in investment strategies to ensure fair market practices.

2. Recommendations for Future Research

While the ASTRM model demonstrates superior performance in handling complex data relationships, several areas warrant further investigation:

2.1 Cross-Industry Applications: Given the automotive industry focus of this study, future research should explore the generalizability of the ASTRM model across different industries and geographical markets. Comparative studies examining sentiment impact variations across sectors would provide valuable insights for developing industry-specific investment strategies.

2.2 Model Architecture Integration: Future studies should investigate hybrid architectures that combine the ASTRM model with other advanced machine learning techniques and explore the integration of alternative data sources to create more comprehensive prediction models for intelligent financial systems.

References

Andleeb, R., & Hassan, A. (2023). Impact of investor sentiment on contemporaneous and future equity returns in emerging markets. *Sage Open, 13*(3), 21582440231193568.

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of financial Markets, 7*(3), 271-299.

Chen, C. (2022). *Investor's sentiment and stock return: an empirical study on the Thai Stock Market*. (Master of Science, Chulalongkorn University).

CSMAR. (2025). *Company financial Analysis*. Retrieved from https://data.csmar.com/.

Mahmud, T., Karim, R., Chakma, R., Chowdhury, T., Hossain, M. S., & Andersson, K. (2024). A benchmark dataset for cricket sentiment analysis in bangla social media text. *Procedia Computer Science, 238*, 377-384.

McGurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance, 44*(3), 458-485.

Shleifer, A. (2000). *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford: Oxford University Press.

Yasser, F., Hatem, S., Ayman, L., Mohamed, L., & Mohammed, A. (2024). Data augmentation for sentiment analysis enhancement: A comparative study. In *2024 6th International Conference on Computing and Informatics (ICCI)* (pp. 53-58). New York: IEEE.

Zhou, Y., Tu, F., Sha, K., Ding, J., & Chen, H. (2024). A survey on data quality dimensions and tools for machine learning invited paper. In *2024 IEEE International Conference on Artificial Intelligence Testing (AITest)* (pp. 120-131). New York: IEEE.