Scholar Publishing Group
Socio-Economic Statistics Research
https://doi.org/10.38007/SESR.2025.060209
ISSN 2790-2722 Vol. 6, Issue 2: 89-98

Research on Sentiment Analysis Based on Multi-source Data Fusion and Pre-trained Model Optimization in Quantitative Finance

Jiahe Sun

Tepper School of Business, Carnegie Mellon University, Pittsburgh 15213, Pennsylvania, U.S

Email: esun29605@gmail.com

Keywords: investor sentiment; pre-trained models; multi-source data fusion; text analysis; behavioral finance

Abstract: Within the behavioral finance framework, investor sentiment is a key driver that can cause asset prices to deviate from fundamentals. In the digital era, multi-source social media platforms such as Reddit and X (Twitter) generate massive, real-time streams of investor opinion data, which create opportunities for high-frequency and fine-grained sentiment measurement while simultaneously challenging the accuracy and adaptability of traditional text analysis methods. This study aims to develop an advanced sentiment analysis framework that integrates multi-source data fusion with optimization of pre-trained language models. To address the challenges posed by complex financial text semantics and the scarcity of labeled data, we move beyond conventional lexicon-based and classical machine learning approaches by adopting and refining pre-trained transformers such as BERT and RoBERTa. Domain-specific fine-tuning of these models markedly improves classification accuracy and robustness. Furthermore, to overcome the limitations of single data sources, we systematically fuse heterogeneous textual data from specialty investment forums (for example WallStreetBets) and broader social platforms (for example X), and construct high-frequency multidimensional indicators including bullish sentiment, attention intensity, and opinion disagreement, thereby exposing sentiment heterogeneity that arises from differences in platform user composition and network structure. An event-driven empirical analysis reveals that bullish sentiment generally exerts a positive effect on individual stock returns, with the effect becoming particularly pronounced within certain event windows; conversely, investor attention is found to be significantly negatively correlated with returns, offering new empirical support for limited attention theories. Notably, the influence of opinion disagreement on market outcomes is not uniform but exhibits clear platform heterogeneity, with stronger negative effects in communities where opinion convergence is greater. All principal findings are supported by thorough heterogeneity checks and robustness tests to ensure result reliability. Methodologically, this work proposes a new paradigm of "pre-trained model optimization + multi-source data fusion" for sentiment analysis; theoretically, it deepens understanding of how sentiment transmission differs across digital social ecosystems; practically, it supplies quantitative strategies, risk management, and regulatory technology with crucial tools and insights.

1. Introduction

In traditional financial theory, the Efficient Market Hypothesis views investors as fully rational "economic agents." However, numerous real-world anomalies—such as the GameStop (GME) phenomenon—demonstrate that investor decisions are often driven by collective psychological and emotional factors. Behavioral finance has therefore emerged as a crucial perspective for understanding market fluctuations. In the digital era, online social platforms and forums (such as Reddit's WallStreetBets, X (Twitter), and StockTwits) have become central venues where retail investors exchange opinions and form consensus. The massive, real-time text data generated on these platforms provides a valuable source for directly observing and quantifying market sentiment.

Although social media-based sentiment analysis has become an important branch of modern finance, existing studies still leave room for expansion. First, on the methodological level, many studies rely on generic sentiment dictionaries or traditional models trained on limited corpora, which struggle to capture domain-specific terminology, contextual nuances, and internet slang in financial discussions, thereby constraining the precision of sentiment measurement. Second, in terms of data scope, much of the literature focuses on a single type of platform (such as forums or blogs), overlooking the sentiment heterogeneity that may arise from differences in user composition, content formats, and recommendation algorithms across platforms—making it difficult to construct a comprehensive sentiment landscape. Third, at the application level, most existing research examines the link between sentiment and asset prices at an aggregate or low-frequency level, while high-frequency sentiment shocks triggered by specific events (such as earnings announcements or regulatory updates) and their micro-level effects on individual stock prices remain underexplored.

To address these challenges, this study proposes a sentiment analysis framework that integrates multi-source data with optimized pre-trained language models, using the U.S. market as an empirical setting. The core research content includes: first, systematically evaluating and adopting advanced pre-trained language models such as BERT and RoBERTa, and fine-tuning them for the financial domain to build a high-accuracy sentiment classifier capable of resolving semantic complexity in financial text; second, fusing diversified data streams from investment forums (e.g., WallStreetBets) and social media (e.g., X), constructing multidimensional, high-frequency sentiment indicators, and examining cross-platform differences and underlying mechanisms in sentiment formation; and finally, from an event-driven perspective, empirically testing the dynamic effects of high-frequency sentiment on individual stock returns, with particular attention to the differential impacts of various sentiment dimensions and data sources.

This research seeks to provide a robust methodological framework that combines "pre-trained model optimization and multi-source data fusion" for sentiment measurement. Theoretically, it deepens the understanding of emotion contagion and asset pricing in digital social environments; practically, it offers new insights and empirical evidence for quantitative factor discovery, risk management, and the modernization of regulatory technology. Subsequent sections will present the sentiment quantification model, data fusion and platform heterogeneity analysis, as well as empirical design and findings.

2. Related Research

This study is situated at the intersection of behavioral finance and computational social science, aiming to quantify investor sentiment from digital social platforms through natural language processing (NLP) techniques. The works of Wang W [1] and Almansour B Y [2] provide the theoretical foundation for this research: the former uncovers the conditional effects of investor sentiment on asset returns, while the latter empirically demonstrates how behavioral biases influence investment decisions through risk perception. Together, these studies highlight the necessity of precise sentiment measurement in the digital era.

In terms of sentiment quantification methods, prior research has evolved from dictionary-based and traditional machine learning approaches to deep learning frameworks. However, the specialized and complex nature of financial text imposes higher demands on model design and adaptability. Zhou D W [3]'s review on the domain adaptability of pre-trained models offers essential methodological guidance for this study's adoption and fine-tuning of models such as BERT to address the limitations of general-purpose models in capturing financial semantics accurately. Furthermore, to overcome the constraints of single-source data and construct a comprehensive sentiment landscape, this research draws upon the concept of multi-source data fusion. Yang H [4]'s approach to handling heterogeneous data in multimodal analysis inspired this study's investigation into the heterogeneity across platforms such as Reddit and X, systematically integrating both structured and unstructured textual information from different online communities to analyze the differentiated pathways of sentiment transmission.

Finally, from an application perspective, Li M [5]'s work demonstrated the practical effectiveness of combining convolutional autoencoders with XGBoost to enhance economic forecasting accuracy, thereby validating the methodological paradigm of "advanced feature extraction + powerful predictive modeling" in financial data analysis [6]. This provides strong methodological support for this study's integrated framework of "pre-trained model optimization + multi-source data fusion," designed to predict stock-level returns more effectively [7].

In summary, prior literature has laid a solid foundation for this research across theoretical, methodological, and technical dimensions[8]. Building upon these insights, this study integrates cutting-edge pre-trained language models with multi-source social media data to deepen the understanding of sentiment transmission mechanisms in digital social ecosystems and to provide innovative analytical tools for high-frequency quantitative analysis and decision-making[9].

3. Construction and Optimization of the Multi-Source Sentiment Analysis Model

3.1 Multi-Source Data Integration and Sentiment Indicator System Development

In building the multi-source sentiment analysis model, data integration and the establishment of a sentiment indicator system constitute the foundational components of this research [10]. In modern financial markets, sentiment expression is dispersed across various social platforms, each forming a unique information ecosystem due to differences in user demographics, interaction patterns, and content formats. To comprehensively capture the heterogeneous characteristics of investor sentiment, this study systematically integrates data from two representative sources: the professional investment forum r/WallStreetBets on Reddit, and the macro-level social media platform X (formerly Twitter). The former is distinguished by its concentration of retail investors and its distinctive meme culture, where discussions often convey explicit trading intentions and strong emotional tones; the latter, with its massive user base and rapid information diffusion, reflects broader market sentiment and attention dynamics.

Given the heterogeneous and unstructured nature of these multi-source textual data, a refined preprocessing pipeline is essential. Data were collected through platform APIs and customized web crawlers to ensure sufficient time-series coverage within specified event windows. Considering the unique linguistic characteristics of financial text, a multi-stage data cleaning process was employed: removing advertisements and irrelevant noise, normalizing financial terminology and internet slang, and conducting tokenization, part-of-speech tagging, and syntactic parsing to support subsequent deep semantic analysis. This preprocessing process effectively reduces semantic noise and enhances data consistency, allowing texts from different platforms to be represented within a unified feature space.

On the basis of this cleaned and standardized textual data, we constructed a multi-dimensional sentiment indicator system designed to capture the psychological state of investors more comprehensively. This system emphasizes not only the direction but also the intensity and dispersion of sentiment. First, the bullish sentiment index is calculated using the weighted logarithmic ratio of positive to negative posts, quantifying market optimism and detecting subtle shifts in collective expectations. Second, the investor attention index measures the total number of posts within a specific time interval, capturing the evolution of public attention and providing a dynamic proxy for event-driven sentiment surges and declines. Third, the sentiment divergence index measures the balance between bullish and bearish sentiment, reflecting the heterogeneity of opinions within the market—an essential dimension for understanding the behavioral drivers of price volatility.

Together, these three indicators form a comprehensive and dynamic sentiment observation framework. The bullish sentiment index reveals market direction, the attention index reflects the intensity and spread of sentiment, and the divergence index indicates the stability of market consensus. Through this multi-dimensional, high-frequency quantification approach, fragmented individual opinions on social media are aggregated into coherent, economically meaningful collective sentiment signals, laying a solid foundation for subsequent analysis of the dynamic relationship between investor sentiment and asset price movements.

3.2 Optimization of Pre-trained Models and Sentiment Analysis

After completing the integration of multi-source data and designing the sentiment indicator framework, the next major technical challenge of this study lies in how to extract sentiment polarity from massive amounts of unstructured text accurately and efficiently. Traditional dictionary-based methods rely heavily on manually constructed sentiment lexicons, which are difficult to adapt to the rapidly evolving financial terminology and internet slang. Similarly, classical machine learning and shallow neural network approaches face significant limitations in capturing deep semantic meaning and contextual dependencies within financial texts. To overcome these bottlenecks, this research introduces advanced pre-trained language models that have achieved groundbreaking results in natural language processing and further optimizes them for financial sentiment analysis.

Pre-trained models such as BERT and its variants have been trained through self-supervised learning on large-scale corpora, enabling them to encode extensive linguistic knowledge. However, directly applying these models to the sentiment classification of financial social media texts introduces domain adaptation challenges. Models trained on general corpora may fail to capture the specific emotional nuances of domain-specific expressions like “short squeeze” or “mooning.” Therefore, domain-adaptive fine-tuning becomes a crucial step in achieving high-accuracy sentiment analysis. In this study, additional fine-tuning was conducted using domain-specific corpora derived from financial news, corporate earnings reports, and social media posts related to finance. This process can be viewed as a knowledge transfer, allowing the model to adjust its internal parameters to better reflect the linguistic patterns and sentiment tendencies of financial texts, thereby converting general language understanding into specialized financial sentiment recognition.

In terms of model selection, this study primarily evaluated BERT, RoBERTa, and FinBERT, the latter being a pre-trained model specifically designed for financial text analysis. All of these models are built upon the Transformer architecture, which provides powerful semantic representation capabilities. Among them, BERT, with its bidirectional encoding mechanism, can simultaneously capture both left and right contextual information of words, demonstrating excellent ability to distinguish polysemous terms (for instance, the word “bullish” can mean both “optimistic” and “expecting a price rise”). RoBERTa, on the other hand, achieves more robust performance by refining its training strategy—removing the Next Sentence Prediction task and adopting larger batch sizes during training. After conducting rigorous comparative experiments on a manually annotated financial text test set, we ultimately selected the model with the best overall performance for the current task as the core engine for our sentiment analysis framework. The detailed performance comparison of these models is shown in Table 1.

Table 1: Performance Comparison of Different Pre-trained Models in Financial Text Sentiment Classification

Sentiment Lexicon	58.2	64.5	58.2	59.3	2.1
Naive Bayes	66.1	68.8	66.1	62.3	8.5
SVM	69.6	69.5	69.6	68.5	15.3
LSTM	69.0	70.5	69.0	68.5	45.2
BERT (General)	75.8	75.6	75.8	75.6	62.7
RoBERTa (Fine-tuned)	77.5	77.2	77.5	77.3	89.4
FinBERT (Finance-specific)	78.3	78.0	78.3	78.1	76.9

As shown in Table 1, it is evident that the series of pre-trained models (BERT, RoBERTa, and FinBERT) significantly outperform traditional approaches in classification accuracy, demonstrating their strong capability to understand and capture the complex semantics of financial texts. Specifically, even the general-purpose BERT model achieves a markedly higher accuracy than the best-performing traditional machine learning model, SVM. Among them, FinBERT, which has been further pre-trained on domain-specific financial corpora, delivers the best overall performance and was therefore selected as the core engine for sentiment analysis in this study. Although its training process requires more time compared to certain traditional models, the substantial improvement in sentiment measurement precision is critical for ensuring the reliability of subsequent empirical analyses.

After determining the base model, we developed a dedicated downstream sentiment classification module. This module typically connects a simple classification layer (such as a fully connected network) after the main body of the pre-trained model. Using our carefully curated and manually annotated financial social media dataset (including samples collected from Reddit and X), the entire model is fine-tuned in an end-to-end manner. During this process, we adopted training techniques such as dynamic masking and gradient clipping to enhance the model’s generalization ability and stability. The optimized model can effectively map an input text (e.g., a tweet or post title) into a three-dimensional sentiment space, outputting a probability distribution over “positive,” “negative,” and “neutral” categories. This classification output serves as the fundamental basis for computing all sentiment indicators, and its accuracy directly determines the credibility of the subsequent empirical results.

Ultimately, the optimized pre-trained sentiment analysis system integrates seamlessly with the previously constructed multi-source data pipeline. Functioning as an efficient and precise “emotional perception hub,” it continuously processes text streams from various platforms and outputs standardized sentiment labels. These labels are directly used to compute key indicators such as bullish sentiment, investor attention, and sentiment divergence, thereby completing the crucial transformation from raw, unstructured social media data into structured, quantifiable emotional factors. This transformation provides a solid technical foundation for exploring the complex dynamics between investor sentiment and market behavior.

4. Empirical Study of the Model and Application of Quantitative Strategies

4.1 Model Validation and Predictive Capability Analysis

After constructing and optimizing the sentiment analysis model, this study proceeds to empirically examine whether the sentiment indicators extracted from multi-source data and pre-trained models possess real explanatory and predictive power over market behavior. To this end, we designed a systematic empirical testing procedure aimed at validating the dynamic relationship between sentiment indicators and asset prices, as well as assessing the robustness of their predictive capability.

A key aspect of the empirical analysis is establishing econometric models that can accurately capture the relationship between sentiment and returns. We constructed the following panel regression model as the basis for the analysis:

$$R_{i,t} = \alpha + \beta_1 Senti_{i,t-1} + \beta_2 Atten_{i,t-1} + \beta_3 Disp_{i,t-1} + \gamma X + \epsilon_{i,t}$$

Here, the dependent variable $R_{i,t}$ represents the return of asset i at time t . The core explanatory variables are the lagged sentiment indicators, including the market sentiment index $Senti_{i,t-1}$, the investor attention index $Atten_{i,t-1}$, and the sentiment dispersion index $Disp_{i,t-1}$. The vector X represents a set of control variables, including overall market returns, firm size, and book-to-market ratio, among other classical factors, to isolate the influence of other potential factors. The regression coefficients β_1, β_2, β_3 in the model are the key focus of this study, as they separately quantify the net effect of different dimensions of sentiment on asset returns.

To gain deeper insights into the driving role of emotions under specific market conditions, we adopt the event study methodology, focusing our analysis on a series of financial event windows that have attracted widespread market attention. These events include corporate earnings announcements, major management disclosures, and social media-driven phenomena (such as short squeeze events). By conducting synchronous analyses of high-frequency emotion series and returns before and after the event windows, we can more clearly identify the independent impact of emotional factors.In terms of sample selection, we cover representative firms across multiple industries in the U.S. stock market, ensuring that the research findings possess strong external validity.

Descriptive statistics and stationarity tests serve as crucial prerequisites for ensuring the reliability of the regression results. Preliminary analyses indicate that the constructed hourly emotion index series are stationary processes, satisfying the fundamental assumptions of time series regression. Correlation analysis shows that although some correlations exist among different emotion indicators, their variance inflation factors (VIFs) are well below the critical threshold for multicollinearity, ensuring the accuracy of model estimation.The empirical results provide strong supporting evidence. The baseline regression results of emotion－return relationships constructed in this study are presented in Table 2, where the signs and significance levels of the core explanatory variables (bullish sentiment, investor attention, and sentiment dispersion) are consistent with and validate the research hypotheses.

Table 2. Baseline Regression Results of Emotion Indicators on Stock Returns

Explanatory Variables	Full Sample (1)	Outside Event Window (2)	Inside Event Window (3)
Constant (α)	-0.002***	-0.003***	-0.001**
	(-3.21)	(-3.56)	(-2.18)
Bullish Sentiment (Sentit−1)	0.123***	0.098***	0.185***
	(4.56)	(3.12)	(5.89)
Investor Attention (Attent−1)	-0.087***	-0.042*	-0.156***
	(-3.89)	(-1.89)	(-4.92)
Opinion Divergence (Dispt−1)	-0.035**	-0.021	-0.068***
	(-2.34)	(-1.45)	(-3.76)
Market Return (Mktt)	0.215***	0.189***	0.256***
	(6.12)	(5.34)	(7.01)
Firm Size (Sizet)	-0.012**	-0.010*	-0.015***
	(-2.45)	(-1.98)	(-3.11)
N (Sample Size)	1280	720	560
Adjusted R2 (%)	8.76	5.23	12.89

As shown in the Table 2, the sentiment index $\mathrm{Senti}_{i,t-1}$ is highly correlated with the return of the market during the event window, indicating that the sentiment changes significantly along with the fluctuations in market prices. This aligns with the hypothesis in finance theory that “emotional movements” drive market behavior. Notably, the sharp price rise during the event window reflects this characteristic period, showing that market attention triggers price movement and liquidity is influenced by emotion.

The investor attention index $\mathrm{Atten}_{i,t-1}$ exhibits a significant relationship with stock return prediction ability. The increased level of attention correlates with price movements. An “attention theory” supports this by providing new insights: when the market is at equilibrium, the attention driven by investor behavior can briefly push the price to deviate from the fundamental level, suggesting that high attention may lead to risk and profit-driven speculation.

The sentiment dispersion index $\mathrm{Disp}_{i,t-1}$ impacts the price significantly. On platforms like WallStreetBets, which are influenced by culture within the community, the sentiment index is remarkably high, reflecting the spread of individual opinions. This scenario, where the emotional contagion of retail investors intensifies, leads to sharp fluctuations in market behavior and trend formation. This process emphasizes the influence of multiple cross-media platforms on market behavior.

To confirm the robustness of our conclusions, we conduct a systematic robustness check. This includes investigating the frequency of network sentiment indicators, such as small time scale signals from social networks and their impact on broader market signals (e.g., market volatility index VIX). The verification indicates that core emotion indicators are stable, strengthening the robustness of our research findings.

4.2 Construction and Backtest Evaluation of Emotional Quantitative Strategy

After verifying the predictive power of the emotion indicators, this study further explores their practical application in the field of quantitative investment. By constructing a systematic emotional timing strategy, we translate theoretical findings into actionable investment methods. Based on the significant predictive ability of the bullish sentiment indicator, we design an emotional momentum strategy, with its core focused on capturing the persistence of investor sentiment. The strategy dynamically monitors the relative changes between the current sentiment indicator and the average level of the past five trading days to precisely identify the key turning points of the emotional cycle. Specifically, when the bullish sentiment exceeds an upper threshold determined by historical backtesting, the system generates a buy signal; conversely, when the sentiment indicator falls below the lower threshold, a sell signal is triggered. This emotion-based momentum strategy design maintains a reasonable trading frequency and effectively controls turnover through parameter optimization, ensuring the feasibility of the strategy in actual trading environments.

In terms of strategy implementation, we fully consider the various constraints of real-world trading. The strategy's parameter settings include transaction costs such as commission fees and slippage costs. The stock transaction commission is charged at 0.03% of the transaction amount, with a minimum fee of $5 per transaction. A 0.1% stamp duty is charged when selling. The position management follows an all-in or all-out model, executing full position buy and sell operations when a signal is triggered. To fully utilize the trading mechanism of the U.S. stock market, the strategy incorporates the T+0 trading rule, allowing it to better capture short-term market volatility driven by emotions. The backtest framework uses the event window as the time range, with the S&P 500 index as the benchmark. A multidimensional evaluation system was established, including absolute return, excess return, maximum drawdown, and Sharpe ratio. The annualized risk-free rate is set at 4% to ensure accurate measurement of the strategy’s risk-adjusted returns.

The empirical results show that the performance of the emotional strategy varies significantly by event type. In managerial decision-type events, the strategy achieved stable positive returns, with excess returns in several cases exceeding the benchmark by more than 4 percentage points, and the Sharpe ratio remaining at a high level. This phenomenon aligns well with the persistence characteristics of emotion driven by fundamental events, indicating that investors' responses to substantial information such as business expansion and strategic partnerships have strong continuity. In crisis-type events, the strategy exhibited a mixed performance; while some events did not achieve absolute positive returns, they still realized positive excess returns relative to the benchmark, demonstrating the strategy's risk-hedging ability in down markets. In celebrity effect-type events, however, the strategy's returns were significantly impacted due to the extreme volatility and rapid reversal of emotions.

To address the limitations of the strategy in specific contexts, we significantly improved the strategy's adaptability by introducing a volatility filtering mechanism and optimizing parameter settings. The specific improvements include reducing trading frequency, extending the emotional observation window, and confirming signals based on price trends. These measures effectively reduced unnecessary trades during periods of extreme emotional volatility. A comparative analysis with traditional technical indicator strategies shows that the emotional strategy outperforms in most event types, especially in managerial decision-type and some crisis-type events, with a more evident advantage in returns. This confirms that social media sentiment data contains incremental information beyond traditional market price data. Emotional factors can capture the psychological changes of market participants earlier, providing unique forward-looking signals for trading decisions.

Through systematic strategy construction and backtest evaluation, this study confirms the practical feasibility of a sentiment-based quantitative strategy using social media emotions. Emotional factors not only have theoretical predictive value but also generate real economic benefits. The heterogeneity in strategy performance also offers new perspectives for understanding the emotion-driven mechanism, suggesting that future research should focus more on the interaction between emotional persistence, event types, and market environments. These findings will help further improve the practical application of emotion-based quantitative models, providing new research directions and practical tools for the field of quantitative investment and promoting the deep integration of behavioral finance theory and investment practice.

5. Conclusion and Outlook

This study, through systematic theoretical exploration and empirical analysis, constructs a framework for investor sentiment analysis that integrates multi-source social media data with pre-trained language models. The research results indicate that sentiment indicators extracted using deep learning methods have significant predictive power over asset price movements. This not only enriches the theoretical system of behavioral finance but also provides new analytical tools for quantitative investment practice.

In terms of methodology, this study's main breakthroughs are reflected in two aspects. First, by introducing pre-trained models fine-tuned with financial domain corpora, the study effectively addresses the limitations of traditional text analysis methods in semantic understanding, significantly enhancing the accuracy and robustness of sentiment recognition. Second, it innovatively integrates data sources from different types of social platforms, constructing a multi-dimensional indicator system that includes sentiment tendencies, attention levels, and opinion divergence, thus enabling a more comprehensive depiction of market sentiment. These methodological innovations provide a reference technical path for subsequent related research.

The empirical research section reveals the complex relationships between sentiment factors and market performance. The bullish sentiment indicator shows stable positive predictive ability, especially prominent during major event windows; investor attention exhibits a negative impact consistent with classical theory, confirming the short-term effects of attention-driven trading; and the sentiment divergence mechanism shows significant platform heterogeneity, reflecting differentiated influences of different network structures on information dissemination and consensus formation. These findings not only validate the informational content of sentiment factors but also deepen our understanding of market micro-behavior in the digital age.

In terms of practical value, the quantitative strategy based on sentiment indicators developed in this study shows promising application prospects in backtesting. The strategy's stable returns in managerial decision-type and crisis-type events prove the practical value of sentiment data in asset allocation and risk management. At the same time, the strategy's performance differences in different market contexts provide important insights into the applicability boundaries of sentiment factors.

Although this study has yielded a series of valuable research findings, several directions still warrant further exploration. Future research could further expand the coverage of data sources by incorporating more diverse social media platforms and news media into the analysis framework. In terms of model optimization, exploring new architectures, such as graph neural networks, could allow for more precise capturing of sentiment propagation paths in social networks. Additionally, examining the interactive effects of sentiment indicators with macroeconomic cycles, market institutional changes, and other macro factors will also be a valuable research direction. These explorations will contribute to the construction of a more comprehensive market sentiment monitoring system, providing new perspectives for understanding the market operating rules in the digital finance era.

References

[1] Wang W, Su C, Duxbury D. The conditional impact of investor sentiment in global stock markets: A two-channel examination[J]. Journal of Banking & Finance, 2022, 138: 106458.

[2] Wang, C. (2025). Exploration of Optimization Paths Based on Data Modeling in Financial Investment Decision-Making. European Journal of Business, Economics & Management, 1(3), 17-23.

[3] Almansour B Y, Elkrghli S, Almansour A Y. Behavioral finance factors and investment decisions: A mediating role of risk perception[J]. Cogent Economics & Finance, 2023, 11(2): 2239032.

[4] Zhou D W, Sun H L, Ning J, et al. Continual learning with pre-trained models: A survey[J]. arXiv preprint arXiv:2401.16386, 2024.

[5] Yang H, Zhao Y, Wu Y, et al. Large language models meet text-centric multimodal sentiment analysis: A survey[J]. arXiv preprint arXiv:2406.08068, 2024.

[6] Li M, Wang F, Jia X, et al. Multi-source data fusion for economic data analysis[J]. Neural Computing and Applications, 2021, 33(10): 4729-4739.

[7] Liu, Y. (2025). The Importance of Cross-Departmental Collaboration Driven by Technology in the Compliance of Financial Institutions. Economics and Management Innovation, 2(5), 15-21.

[8] Zhang, Xuanrui. "Automobile Finance Credit Fraud Risk Early Warning System based on Louvain Algorithm and XGBoost Model." In 2025 3rd International Conference on Data Science and Information System (ICDSIS), pp. 1-7. IEEE, 2025.

[9] Jing X. Real-Time Risk Assessment and Market Response Mechanism Driven by Financial Technology[J]. Economics and Management Innovation, 2025, 2(3): 14-20.

[10] Zhou Y. Research on the Innovative Application of Fintech and AI in Energy Investment[J]. European Journal of Business, Economics & Management, 2025, 1(2): 76-82.