FINS is your favor: A Task-Driven Data Evaluation Framework for Financial Sentiment Analysis

1st Liangzhou Qu
Shenzhen Audencia Financial Technology Institute
Shenzhen University
Shenzhen, China
liangzhou.qu@audencia.com

2nd Di Han
School of National Finance
Guangdong University of Finance
Guangzhou, China
dihan@gduf.edu.cn

3rd Junjie Mao
Shenzhen Audencia Financial Technology Institute
Shenzhen University
Shenzhen, China
junjie.mao@audencia.com

4th Qixian Li
School of National Finance
Guangdong University of Finance
Guangzhou, China
rambolqx@163.com

5th Zikun Guo
School of National Finance
Guangdong University of Finance
Guangzhou, China
zikunguo@foxmail.com

6th Weiquan Fan
School of National Finance
Guangdong University of Finance
Guangzhou, China
wqfan@gduf.edu.cn

***Abstract*—Currently, datasets for Financial Sentiment Analysis (FSA) often contain rich multimodal information, yet the task-oriented evaluation metrics for processing this sentiment remain singular. Consequently, performance comparisons across different datasets are challenging, which hinders researchers from selecting appropriate resources for specific tasks. To address this problem, we propose FINS (Financial Integration Nexus of Standards), a data evaluation framework that systematically summarizes existing studies and abstracts the main application tasks of FSA into five core dimensions: cycle, risk, causality, entity and wording. On this basis, we establish a unified quantitative evaluation system for data. FINS combines large language models with financial toolchains to achieve an automated evaluation process, and presents the adaptability of data across different task dimensions through multi-dimensional quantitative score maps. Empirical results and real financial scenario experiment on several representative data sources show that FINS can effectively capture structural differences, which provides systematic support for researchers in data selection and modeling practices for specific tasks.**

***Index Terms*—Financial Sentiment Analysis, Data Evaluation, Large Language Model, Task Adaptability, Quantitative Metrics.**

I. Introduction

Financial Sentiment Analysis (FSA) refers to the identification and quantification of opinions, attitudes, and emotions in financial texts [1]. It is commonly applied in market sentiment profiling [2], investment decision support [3], stock price prediction [4], and risk monitoring [5]. However, the further development of FSA is constrained by the limitations of data quality and task adaptability, and this contradiction highlights the urgent need for a systematic evaluation approach.

Existing studies have preliminarily established some evaluation systems within specific domains of FSA. For example, in causality extraction, FinCausal focuses on identifying causal relations in financial texts [6]. In multimodal expression, FinMME attempts to enhance sentiment analysis by integrating textual and visual information [7]. Although these studies have made contributions within their respective dimensions, their excessive focus on single tasks makes it difficult to provide researchers with cross-task selection and systematic evaluation criteria in practical applications. As a result, the field still lacks a unified fine-grained standard for overall FSA applications.

This paper presents FINS (Financial Integration Nexus of Standards), a task-driven data evaluation framework for FSA. FINS takes the task dimension as its entry point and, based on a systematic review and integration of existing studies, abstracts financial application requirements into five core dimensions: market cycle patterns, risk analysis, causality modeling, entity relationship modeling, and wording difference analysis. On this basis, it establishes a unified quantitative evaluation system for data. Across these dimensions, FINS combines large language models (LLMs), financial domain tools such as FinBERT [8], and structured graph algorithms to design automated evaluation processes for each dimension. Finally, the scores of each dimension are aggregated into a task adaptability profile under multi-task settings, which supports the fine-grained selection of FSA datasets. To validate the effectiveness of FINS, we conducted a systematic empirical evaluation on five representative financial text datasets. The real finance scenario experimental results show that FINS can effectively capture the structural differences of different datasets across the five task dimensions and provide systematic support for researchers in dataset selection and modeling practices for specific tasks.

The main contributions of this paper are as follows:

(1)We propose the FINS framework to systematically evaluate the structural characteristics and task performance of FSA datasets, providing a basis for dataset selection and model

This study is supported by the Natural Science Foundation of Guangdong Province (Grant No. 2025A1515011633) and the Key Program of the Department of Education of Guangdong Province (Grant No. 2024ZDZX1036).

design through result analysis.

(2)We develop an automated evaluation pipeline that incorporates LLMs, addressing the consistency issues of manual evaluation and significantly reducing the required manpower and time costs.

II. Evolution of FSA Tasks

The data evaluation paradigm for FSA is not static. Instead, it has undergone a natural evolution from singular polarity classification to multi-dimensional signal parsing, accompanying the deepening demand for information processing in financial markets. To explore the latent structure governing financial text quality, we systematically reviewed classic literature over the past two decades. We found that the research focus of FSA has experienced a significant process from differentiation to deepening and revealed the critical theoretical attributes that high-quality financial data must possess.

Early FSA research primarily focused on validating the predictive power of textual information on markets. Classical quantitative analysis on Media Pessimism [9] revealed a key fact: textual sentiment alone cannot function independently. Its informational value is highly coupled with the market prevailing macro state, such as bull or bear markets. This indicates that textual data detached from the temporal dimension and market cycle context lacks explanatory power. Meanwhile, the construction of Domain-specific Lexicons [10] challenged the applicability of general dictionaries. This research explicitly pointed out that the characteristics of financial text lie in its Ambiguity and Uncertainty, rather than a simple accumulation of negative words. Research from this period demonstrates that high-quality data evaluation must include clear market cycle markers and precise wording characteristics.

As research deepened, scholars realized that macro indicators could not explain heterogeneity at the micro level, shifting the evaluation perspective to micro-subjects. The proposal of the Financial PhraseBank dataset [11] pushed research toward fine-grained, sentence-level sentiment analysis. Its core discovery was that financial sentiment possesses high object specificity, meaning the same text often contains opposing evaluations for different subjects. This requires data to possess precise entity recognition capabilities. Furthermore, empirical research on investor sentiment Network Contagion [12] [13] found that discussions on specific stocks exhibit significant topological clustering features on social media. This shift in focus marks that high-quality financial text must include clear entity granularity and their interconnecting network structures.

In recent years, the task objectives of FSA have upgraded from describing states to explaining mechanisms and predicting consequences. The establishment of the FinCausal shared task emphasized that simple sentiment classification is insufficient to assist complex decision-making; extracting the logical chains or Causal Chains implied in text is crucial for understanding market dynamics. Meanwhile, research such as News Implied Volatility (NVIX) and systemic risk modeling [14] [15] has pushed text analysis to the forefront of risk management. These works no longer focus on ordinary sentiment fluctuations but on the text ability to alert against extreme volatility and Tail Risk. This trend indicates that deep financial data should encompass clear deductive logic and acute sensitivity to extreme consequences.

With the rise of Generative AI represented by LLMs, FSA application scenarios have expanded to complex tasks like investment advice generation [16]. Although task forms are changing, we argue that the underlying evaluation logic presents a unification of the fundamental paradigm. Even for state-of-the-art generative tasks represented by FinGPT [17], the core capability remains the comprehensive invocation of the aforementioned elements: the model must understand market context to avoid spatiotemporal dislocation, identify trading entities to ensure specific referencing, warn of risk consequences to meet risk control requirements, construct reasoning logic to enhance persuasion, and employ professional wording to build trust.

Based on the above evolutionary logic, we propose following five FSA task dimensions, including market cycle patterns, wording difference analysis, entity relationship modeling, causal relationship modeling, and risk analysis—constitute the minimum complete basis for describing the quality of financial texts. Any high-order financial NLP task is essentially a linear combination of these five basis vectors.

III. Methodology

Guided by this theoretical convergence, we propose FINS, a task-driven data evaluation framework. Unlike previous works that focus on single metrics, FINS is designed to explicitly quantify these five orthogonal dimensions—Cycle, Risk, Causality, Entity, and Wording—thereby providing a comprehensive and structured assessment of financial datasets.

This section provides a detailed introduction to the overall design and implementation of the FINS framework (Fig.1).

A. Market Cycle Patterns

Financial market modeling aims to characterize the performance of data across different market environments by combining economic cycle data with textual data. If the data adequately cover both bull and bear market cycles, they are more applicable in subsequent analysis and modeling. Thus, we propose the Macro Cycle Coverage Rate (MCCR) indicator to quantify the coverage of data within economic cycles:

$$MCCR = \sum_{i=1}^{n} \frac{|D \cap B_i \cap T|}{|B_i \cap T|} \tag{1}$$

Here, D is the dates where dataset actually contains documents, B_i represents the i-th bull or bear market cycle, defined using economic cycle data provided by Yardeni Research [1], and T refers to the dataset's time window. MCCR measures the dataset's coverage ability by summing the coverage ratio for each bull and bear market cycle.

[1] www.yardeni.com

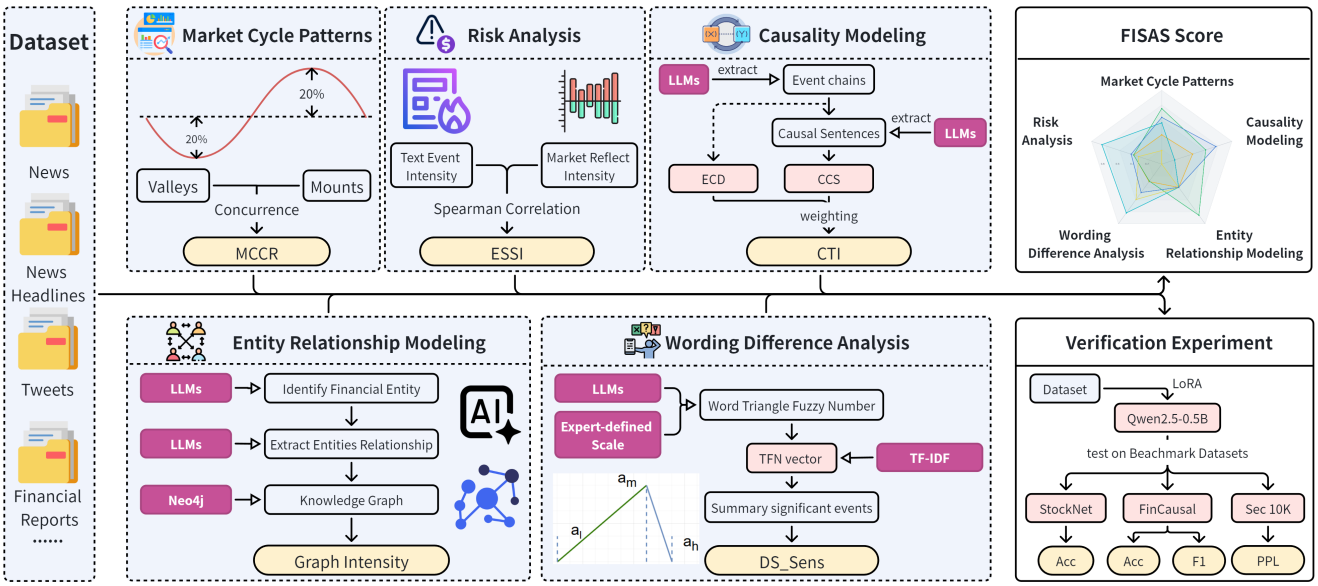

Fig. 1. FINS ingests datasets simultaneously across five tasks: market cycle patterns, entity relationship modeling, risk analysis, causality modeling, and wording difference analysis. Each task evaluates datasets using quantitative metrics. Finally, result gives each dataset scores across these task dimensions, providing assessment of its performance and adaptability.

B. Risk Analysis

Risk analysis task focuses on identifying early signals of extreme negative sentiment and potential volatility. Studies have shown that a sudden surge in company-related texts often corresponds to abnormal changes in market sentiment and may trigger risk events [18]. To quantify the amplification effect of discursive shocks on market behavior, we propose the Event Sentiment Shock Index (ESSI). This index evaluates the risk impact implied by text surges by measuring the statistical correlation between the volume of event-related texts and the intensity of market reactions.

We define the event-driven text volume N_t. Using the DeepSeek-V3 model to enrich company labels, we count the number of strongly event-related texts for companies at time t. The market reflect intensity M_r, following Cheng et al. [19], is measured as the absolute deviation of trading volume from its mean, standardized by its standard deviation:

$$M_r = \frac{|V_t - \mu_V|}{\sigma_V} \tag{2}$$

We first measure the Spearman rank correlation coefficient ρ_s between text volume N_t and market reflect intensity M_r. To comprehensively capture the relationship, we decompose it into three dimensions: mean correlation ρ_{mean} represents the overall strength of association, effective event coverage (COV) indicates the proportion of events with significant associations, and stability of association (STA) reflects the robustness of the correlation over the time series. Finally, ESSI is defined as the product of the three dimensions:

$$\rho_{mean} = \frac{1}{K}\sum_{k=1}^{K} \rho_s^{(k)} \tag{3}$$

$$COV = \frac{1}{K}\sum_{k=1}^{K} \mathbb{I}\left(\rho_s^{(k)} > \tau\right) \tag{4}$$

$$STA = e^{\sigma_{\rho_s}} \tag{5}$$

$$ESSI = \rho_{mean} \times COV \times STA \tag{6}$$

Here, K denotes the total number of event windows, $\rho_s^{(k)}$ represents the Spearman rank correlation coefficient in the k-th event window, $\mathbb{I}$ is the indicator function, τ is the significance threshold, and σ_{ρ_s} is the standard deviation of the Spearman coefficients. This multiplicative form ensures that ESSI effectively filters out random noise and enhancing reliability.

C. Causality Modeling

Causality modeling aims to identify and quantify the causal transmission mechanisms driven by sentiment signals in unstructured financial texts [20]. Since financial events often affect prices and risk behaviors through complex chains, capturing causal semantics is crucial for uncovering the relationship between sentiment and market dynamics [21]. However, traditional rule-based methods that rely on dependency syntax and causal verbs are difficult to generalize in financial texts characterized by diverse expressions and ambiguous contexts [22]. Therefore, we introduce LLMs in combination with financial domain lexicons to achieve causal extraction and chain construction across sentences and documents. We propose the Causal Training Index (CTI) to quantify the adaptability of data in the causality dimension:

$$CTI = \alpha \cdot EDC_{avg} + (1-\alpha) \cdot CCS \tag{7}$$

Here, α is a weighting parameter that balances the importance of causal chain complexity and cross-document causal density. Causal chain complexity EDC_{avg} and cross-document causal density CCS are respectively used to characterize the depth of causal relations in texts and the extent of cross-document coverage:

$$ECD_{avg} = \frac{1}{|C|}\sum_{k=1}^{|C|} L\left(c_k\right) \tag{8}$$

$$CCS = \frac{\left|\left\{(s_i, s_j) \mid s_i \in doc_a, s_j \in doc_b, s_i \rightarrow s_j\right\}\right|}{(|D_e| \times S)} \tag{9}$$

Here, C denotes the set of event chains, and the length $L\left(c_k\right)$ of a single chain $c_k \in C$ represents the number of causal relations it contains. D_e is the number of documents, and S is the average number of sentences per document. CTI can capture both the complexity and the coverage of causal expressions in financial texts, providing a comprehensive assessment of data adaptability in the causality modeling dimension.

D. Entity Relationship Modeling

Financial texts often involve multiple entities and their interrelations, with sentiment propagating along these relational chains and exerting varying degrees of impact [23]. Thus, it is necessary to examine whether it contains sufficient financial entity information and the connections among those entities.

We design a method of entity density and relation parsing. First, we use DeepSeek-V3 model to identify entity names such as financial institutions, individuals, indicators, and stock codes [24]. Next, we extract semantic relations between entities and inject the results into a Neo4j [2] knowledge graph to construct entity nodes and relational networks. Finally, we calculate the graph density ρ_{graph} to measure the distribution of financial entity relations in the dataset:

$$\rho_{graph} = \frac{|E|}{(|V|(|V|-1))} \tag{10}$$

Here, V is the number of nodes in the graph, and E denotes the number of edges. A higher value of ρ_{graph} indicates that entities and relations in the dataset are more densely connected, which helps support sentiment modeling and propagation analysis at the entity level.

E. Wording Difference Analysis

Wording difference analysis is used to measure the impact of variations in wording across different events or time periods on sentiment expression [25]. In the financial domain, even subtle differences in wording may lead to significant sentiment shifts, making its quantification highly important.

First, we use FinBERT to assign sentiment scores to each text, outputting $\{negative, neutral, positive\}$ to represent the probability distribution of the three sentiment categories [26]. Then, we define the uncertainty expansion factor ε, which maps the sentiment probabilities into a fuzzy triangular number (a_l, a_m, a_h). a_m represents the median level of positivity, while the interval width $\Delta\omega = a_l - a_h$ reflects the uncertainty of the model's prediction [27]. For the same event c, we define the maximum knowledge deviation Δe_c to characterize the maximum wording difference within the group:

$$\Delta e_c = \max_{i,j} |a_{m,i} - a_{m,j}| + \frac{\Delta\omega_i + \Delta\omega_j}{2} \tag{11}$$

We propose the DS_{sens} metric to quantify the sensitivity of sentiment distribution caused by wording differences:

$$DS_{\text{Sens}} = \frac{1}{|C_{\text{valid}}|}\sum_{c \in C_{\text{valid}}} \Delta e_c \tag{12}$$

IV. Experiment and Result

A. Datasets Description

We selected five representative datasets, including news, news headlines, financial reports and tweets, effectively supporting multi-dimensional evaluation tasks (see in Table1).

TABLE I
DATASETS OUTLINE

Dataset[3]	Type	Timestamp	Company label
KrossKinetic	News, Headlines	2006–2024	✓
JanosAudran	Financial reports	1993–2020	✓
Bloomberg	News, Headlines	2006–2013	×
Financial Tweets	Tweets	2018	✓
S&P 500	News headlines	2008–2024	×

Note: The source of datasets can be viewed in the GitHub repository we created: https://github.com/PeterQLz/FINS-data-evaluation-framework

B. FINS Scores

The empirical results in Table 2 (with visualization in Fig.2) show significant differences across the five quantitative dimensions, revealing the varying adaptability and performance advantages of different corpora.

TABLE II
PERFORMANCE OF DATASETS ON FIVE EVALUATION METRICS

Dataset	**Metrics**				
	MCCR	**ESSI**	**CTI**	ρ_{graph}	DS_{sens}
KrossKinetic	0.3480	0.0086	1.0149	0.0003	0.7921
JanosAudran	0.5154	0.0040	1.0148	**0.0075**	**0.8502**
Bloomberg	**0.7385**	0.0114	**5.1064**	0.0002	0.7924
Financial Tweets	0.0759	0.0003	1.0150	0.0041	0.7843
S&P 500	0.5931	**0.0140**	1.0090	0.0027	0.7962

Note: In the ESSI and DS_{sens} metric, the significance threshold τ is set to 0.3. In the CTI metric, the parameter α is set to 0.5.

(1) **MCCR**: Bloomberg dataset achieves the highest score (0.7385), demonstrating its advantage in time span and synchronization with market cycles.
(2) **ESSI**: S&P 500 dataset records the highest score (0.0140), indicating that its related texts provide immediate responses to major market events, which leads to higher volatility and stronger correlation with trading volume. Bloomberg (0.0114) focuses on offering comprehensive market analysis and forecasts rather than instant reactions, making it relatively less

[2]https://neo4j.com/

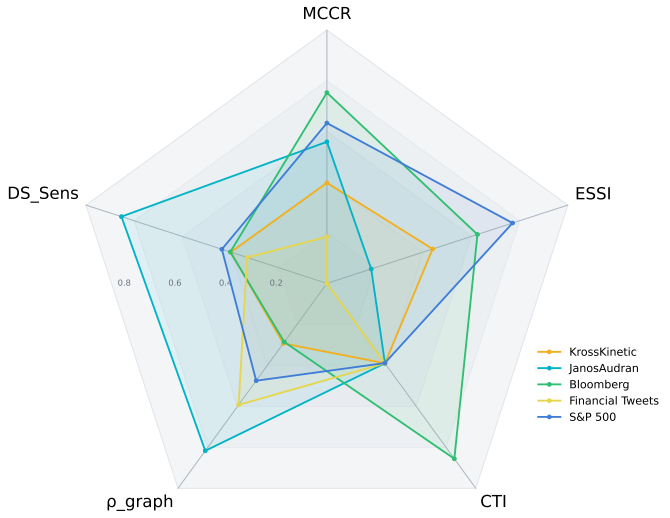

Fig. 2. Radar Chart of Dataset Performance Across Five Metrics (after processing with z-score and sigmod)

sensitive to market shocks and less effective in capturing short-term fluctuations.

(3) **CTI**: Bloomberg dataset achieves 5.1064, which is significantly higher than other datasets. This indicates that its texts are longer and contain richer causal narrative structures, enabling effective causal relationship modeling. In comparison, the CTI values of the Tweets and KrossKinetic are close to 1, reflecting their limited capacity for causal semantic modeling due to shorter texts that lack sufficient context and causal logic.

(4) ρ_{graph}: JanosAudran financial reports dataset obtains the highest score (0.0075), which is clearly higher than news datasets. Since financial reports typically contain structured information that explicitly specifies entities and their interconnections, they are more suitable for fine-grained entity relationship modeling.

(5) DS_{sens}: JanosAudran financial reports also lead with a score of 0.8502, indicating that their sentiment polarity aligns more consistently with actual market reactions. In contrast, news and social media texts are influenced by wording variations and media selectivity, which results in relatively lower sentiment consistency.

C. Verification Experiment

1) Experimental Objective: This section is to empirically validate the effectiveness of the quantitative evaluation metrics proposed by the FINS. While FINS provides multi-dimensional static scores for financial datasets, it is imperative to verify whether these scores possess significant predictive validity regarding the actual performance of models in downstream applications. Since MCCR is essentially a descriptive statistical metric measuring the cycle coverage , its validity is established through explicit data distribution analysis.

2) Experimental Setup and Pipeline: To validate the effectiveness of FINS metrics, this study established a standardized source training and downstream testing evaluation pipeline. We utilized Qwen-2.5-0.5B as the base model and employed Low-Rank Adaptation technology (LoRA) to fine-tune it under identical hyperparameters on three corpora with distinct FINS characteristics, specifically Bloomberg (High-FINS data), Financial Tweets (Low-FINS data), and JanosAudran (high specificity data). The models were subsequently evaluated on three benchmark datasets:

- StockNet [28] is a comprehensive dataset that aligns crowd-sourced social media texts with historical stock price movements to support binary classification tasks. StockNet validates the ESSI metric by examining the capacity of the model to extract effective market signals from social media noise to measure data impact.
- FinCausal is a specialized corpus derived from the Financial Document Causality Detection Task, featuring annotated complex cause-effect chains extracted from financial news. FinCausal can validate CTI and ρ_{graph} by assessing the ability of the model to construct deep causal chains and dynamic knowledge structures.
- SEC 10-K Snippets [3] comprise a collection of text segments extracted from annual financial reports filed with the U.S. Securities and Exchange Commission, characterized by rigorous and specialized legal-financial terminology. SEC 10-K Snippets for domain adaptation validates DS_{sens} by calculating Perplexity, also known as PPL, to measure the generalization robustness of the model when adapting to difficult professional terminology. The PPL is calculated as follows:

$$\text{PPL}(X) = e^{-\frac{1}{t}\sum_{i=1}^{t} \ln P_\theta(x_i|x_{<i})} \tag{13}$$

where $P_\theta(x_i|x_{<i})$ represents the predicted probability assigned by the model to the i-th token, and a lower PPL indicates superior domain adaptation.

3) Results and Analysis: Table 3 reports the result of performance evaluation on downstream tasks.

TABLE III
PERFORMANCE EVALUATION ON DOWNSTREAM TASKS

Task	Metric	Training Dataset Source		
		Bloomberg	Tweets	JanosAudran
StockNet	Accuracy	**69.60%**	63.25%	65.72%
FinCausal	Accuracy	**60.49%**	8.79%	7.50%
	F1 Score	**18.24%**	13.95%	11.87%
SEC 10-K	PPL (↓)	**7.6204**	7.6954	9.1430

Note: PPL (Perplexity) measures model uncertainty, thus a lower score reflects better domain adaptation.

On the StockNet benchmark, the Bloomberg model achieved the highest accuracy (69.60%), surpassing even the in-domain Tweets model (63.25%). This demonstrates that high-ESSI data captures intrinsic market event-driving signals and sentiment shock that facilitate superior cross-domain generalization, whereas the low-ESSI Tweets dataset suffers from significant noise interference that hampers predictive performance despite domain consistency.

[3]https://huggingface.co/datasets/DerivedFunction/sec-filings-snippets-10K

In the FinCausal task, the Bloomberg model dominated with an F1 score of 18.24%, while baselines degraded to below 14%. This confirms that high CTI equips models with the deep logical reasoning necessary to avoid mode collapse. Moreover, the underperformance of JanosAudran, despite its high static graph density (ρ_{graph}), proves that true financial intelligence relies on dynamic causal logic to activate entity connectivity rather than the rote memorization of static structures.

On the SEC 10-K benchmark, the Bloomberg model achieved the optimal perplexity (7.6204), significantly outperforming the report-based JanosAudran model (9.1430). This reveals that the extremely high DS_{sens} of JanosAudran leads to overfitting and "brittle" adaptation to specific wording styles. In contrast, the Bloomberg dataset strikes a critical balance, maintaining professional adaptability while ensuring robust generalization to unseen standardized filings.

V. Conclusion and Future Work

To address the limitations of singular evaluation metrics, we propose FINS task-driven data evaluation framework. FINS establishes the feature engineering mechanism and mathematical expression for the evolution of FSA from simple polarity classification to multi-dimensional deep semantic features (cycle, risk, causality, entity, wording). Empirical results demonstrate that FINS can reveal differences in task alignment among datasets. Specifically, downstream validation on StockNet, FinCausal, and SEC 10-K confirms that High-FINS datasets achieve superior performance in market prediction, logical reasoning, and domain adaptation. Future work will extend to three-dimensional evaluation involving datasets, tasks, and large language models, in order to enhance the practicality and interpretability of financial NLP systems.

References

[1] Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria, "Financial sentiment analysis: Techniques and applications," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–42, 2024.

[2] Allen H Huang, Hui Wang, and Yi Yang, "Finbert: A large language model for extracting information from financial text," *Contemporary Accounting Research*, vol. 40, no. 2, pp. 806–841, 2023.

[3] J Liu, F Xu, and X Liu, "Financial market sentiment analysis and investment strategy formulation based on social network data," *J. Electr. Syst*, vol. 20, pp. 655–660, 2024.

[4] Ali Mehrabian, Ehsan Hoseinzade, Mahdi Mazloum, and Xiaohong Chen, "Mamba meets financial markets: A graph-mamba approach for stock price prediction," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[5] Pengyu Chen and Mingjun Ji, "Deep learning-based financial risk early warning model for listed companies: A multi-dimensional analysis approach," *Expert Systems with Applications*, p. 127746, 2025.

[6] Dominique Mariko, Hanna Abi Akl, Estelle Labidurie, Stephane Durfort, Hugues De Mazancourt, and Mahmoud El-Haj, "Financial document causality detection shared task (fincausal 2020)," *arXiv preprint arXiv:2012.02505*, 2020.

[7] Junyu Luo, Zhizhuo Kou, Liming Yang, Xiao Luo, Jinsheng Huang, Zhiping Xiao, Jingshu Peng, Chengzhong Liu, Jiaming Ji, Xuanzhe Liu, et al., "Finmme: Benchmark dataset for financial multi-modal reasoning evaluation," *arXiv preprint arXiv:2505.24714*, 2025.

[8] Dogu Araci, "Finbert: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063*, 2019.

[9] Paul C Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[10] Tim Loughran and Bill McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *The Journal of finance*, vol. 66, no. 1, pp. 35–65, 2011.

[11] Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *Journal of the Association for Information Science and Technology*, vol. 65, no. 4, pp. 782–796, 2014.

[12] Thomas Renault, "Intraday online investor sentiment and return patterns in the us stock market," *Journal of Banking & Finance*, vol. 84, pp. 25–40, 2017.

[13] Kelvin Du, Rui Mao, Frank Xing, and Erik Cambria, "A dynamic dual-graph neural network for stock price movement prediction," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[14] Asaf Manela and Alan Moreira, "News implied volatility and disaster concerns," *Journal of Financial Economics*, vol. 123, no. 1, pp. 137–162, 2017.

[15] Paola Cerchiello and Paolo Giudici, "Big data analysis for financial risk management," *Journal of Big Data*, vol. 3, no. 1, pp. 18, 2016.

[16] Di Han, Jiaqi Chen, Zikun Guo, Qixian Li, and Bocheng Wang, "A novel exchange rate forecasting paradigm based on multi-agent collaboration and multimodal big data-driven methods," *Big Data Mining and Analytics*, 2025.

[17] Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha, "Fingpt: Democratizing internet-scale data for financial large language models," *arXiv preprint arXiv:2307.10485*, 2023.

[18] Hiroyasu Inoue and Yasuyuki Todo, "Market reaction to news flows in supply chain networks," *arXiv preprint arXiv:2409.06255*, 2024.

[19] Wai Khuen Cheng, Khean Thye Bea, Steven Mun Hong Leow, Jireh Yi-Le Chan, Zeng-Wei Hong, and Yen-Lin Chen, "A review of sentiment, semantic and event-extraction-based approaches in stock forecasting," *Mathematics*, vol. 10, no. 14, pp. 2437, 2022.

[20] Brahim Gaies, Mohamed Sahbi Nakhli, Rim Ayadi, and Jean-Michel Sahut, "Exploring the causal links between investor sentiment and financial instability: A dynamic macro-financial analysis," *Journal of Economic Behavior & Organization*, vol. 204, pp. 290–303, 2022.

[21] Keane Ong, Wihan Van Der Heever, Ranjan Satapathy, Erik Cambria, and Gianmarco Mengaldo, "Finxabsa: Explainable finance through aspect-based sentiment analysis," in *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2023, pp. 773–782.

[22] Marie-Catherine De Marneffe and Christopher D Manning, "The stanford typed dependencies representation," in *Coling 2008: proceedings of the workshop on cross-framework and cross-domain parser evaluation*, 2008, pp. 1–8.

[23] Yixuan Tang, Yi Yang, Allen Huang, Andy Tam, and Justin Tang, "Finentity: Entity-level sentiment classification for financial texts," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 15465–15471.

[24] Xingkun Yin, Da Yan, Abdullateef Almudaifer, Sibo Yan, and Yang Zhou, "Forecasting stock prices using stock correlation graph: A graph convolutional network approach," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[25] Naimeh Sadeghi, Nima Gerami-Seresht, Witold Pedrycz, and Aminah Robinson Fayek, "Fuzzy granular computing for evaluating average uncertainty in machine learning models," *Engineering Applications of Artificial Intelligence*, vol. 160, pp. 111723, 2025.

[26] Chi-Han Du, Ming-Feng Tsai, and Chuan-Ju Wang, "Beyond word-level to sentence-level sentiment analysis for financial reports," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1562–1566.

[27] Di Han, Yijun Chen, and Shuya Zhang, "Implicit social recommendation algorithm based on multilayer fuzzy perception similarity," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 2, pp. 357–369, 2022.

[28] Yumo Xu and Shay B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Iryna Gurevych and Yusuke Miyao, Eds., Melbourne, Australia, July 2018, pp. 1970–1979, Association for Computational Linguistics.