Journal of Sustainability, Policy, and Practice

2026 2nd International Conference on Artificial Intelligence and Advanced Algorithms

Review

AI-Based Sentiment Analysis for Stock Market Prediction: A Systematic Literature Review

Fanyi Zhao [1,*] and Tianxing Tang [2]

[1] Computer Science, Stevens Institute of Technology, Hoboken, NJ, USA
[2] Translation and Localization Management, Middlebury Institute of International Studies, Monterey, CA, USA
* Correspondence: Fanyi Zhao, Computer Science, Stevens Institute of Technology, Hoboken, NJ, USA

Abstract: The integration of artificial intelligence and natural language processing into financial market prediction has attracted significant research attention over the past decade. This review systematically examines the landscape of AI-driven sentiment analysis techniques applied to stock market movement prediction, covering 87 studies published between 2011 and 2024. A taxonomic framework is proposed to categorize existing approaches along four dimensions: data sources, sentiment extraction techniques, correlation modeling strategies, and evaluation metrics. Quantitative comparisons across lexicon-based, machine learning, deep learning, and large language model paradigms reveal that transformer-based models achieve the highest directional accuracy (up to 65.28%) on standard benchmarks, while lexicon-based methods retain advantages in computational efficiency and interpretability. Key challenges including data noise, temporal decay, and cross-market generalizability are critically assessed, alongside emerging trends in multimodal fusion and explainable AI for financial sentiment.

Keywords: sentiment analysis; stock market prediction; natural language processing; deep learning

Received: 05 March 2026
Revised: 28 April 2026
Accepted: 10 May 2026
Published: 13 May 2026

1. Introduction

1.1. Background: The Intersection of Artificial Intelligence and Financial Markets

Financial markets are complex adaptive systems influenced by macroeconomic indicators, corporate fundamentals, geopolitical events, and collective investor psychology. The Efficient Market Hypothesis (EMH) postulates that asset prices fully reflect all available information, rendering consistent prediction theoretically impossible. In practice, mounting empirical evidence suggests that textual information embedded in news articles, social media posts, and financial filings carries predictive signals that quantitative price data alone cannot capture. Malo et al. demonstrated that detecting semantic orientations in economic texts yields statistically significant correlations with subsequent market movements, challenging strong-form efficiency assumptions [1].

The proliferation of user-generated financial content on platforms including Twitter, Reddit, and StockTwits has created an unprecedented volume of opinion data. Soun et al. reported that the daily volume of finance-related tweets exceeded 2.8 million by 2021. This data abundance has catalyzed the development of increasingly sophisticated AI-based sentiment extraction pipelines spanning lexicon-based heuristics, supervised classifiers, deep neural architectures, and pre-trained large language models [2].

1.2. Motivation and Research Questions

Despite the growing body of literature, no recent comprehensive review has systematically cataloged the full spectrum of AI-driven sentiment analysis methods designed for stock prediction, nor has a unified taxonomic framework been established for cross-study comparison. Bollen et al. provided early evidence that aggregate Twitter mood states predict Dow Jones Industrial Average movements with 87.6% directional accuracy, a finding that ignited widespread research interest [3]. The intervening years have witnessed rapid methodological evolution---from bag-of-words representations to contextualized embeddings---yet the field lacks a consolidated assessment of which approaches generalize across markets and time horizons.

This review addresses three research questions: (RQ1) What data sources and sentiment extraction techniques have been employed in AI-based stock prediction? (RQ2) How do different modeling paradigms compare in prediction accuracy, robustness, and scalability? (RQ3) What are the principal challenges and emerging trends shaping the next generation of sentiment-driven financial forecasting?

1.3. Scope Definition and Contribution of This Review

The scope of this review is restricted to studies that (a) apply AI or NLP techniques to extract sentiment from textual data and (b) use extracted sentiment features as inputs for stock price or movement prediction. Studies focused exclusively on technical indicator prediction or macroeconomic forecasting without sentiment components fall outside this scope. The contributions are threefold: a taxonomic classification of 87 reviewed studies across four analytical dimensions, a quantitative meta-comparison of prediction performance on standard benchmarks, and an identification of research gaps with actionable directions.

2. Review Methodology

2.1. Literature Search Strategy and Database Selection

A systematic search was conducted across five academic databases: IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and Google Scholar. The search query combined terms from two concept groups using Boolean operators: ("sentiment analysis" OR "opinion mining" OR "text mining" OR "NLP") AND ("stock prediction" OR "stock market" OR "financial forecasting" OR "equity prediction"). The search was limited to English-language publications from January 2011 to December 2024. Yang et al. highlighted that common methodological errors in financial sentiment studies---including label leakage, improper train-test splits, and sentiment dictionary mismatch---necessitate careful screening criteria, which informed our quality assessment protocol [4].

2.2. Inclusion and Exclusion Criteria

Eligible studies were required to meet all of the following criteria: (1) employ at least one AI, machine learning, or NLP technique for sentiment extraction; (2) target individual stock or index-level price movement prediction; (3) report quantitative evaluation metrics enabling cross-study comparison; and (4) appear in peer-reviewed conference proceedings or journals. Studies exclusively using survey or questionnaire-based sentiment proxies, cryptocurrency-only analyses, and non-English publications were excluded. After removing duplicates, 412 records were screened by title and abstract, yielding 143 full-text candidates. Following quality assessment---adapted from the Kitchenham guidelines for systematic reviews in software engineering---87 studies met all inclusion criteria.

2.3. Bibliometric Overview of the Retrieved Literature

The temporal distribution of publications reveals an exponential growth pattern, with 68.9% of the 87 included studies published after 2018. Ding et al. introduced event-driven deep learning for stock prediction in 2015, marking a methodological turning point; annual publications increased from an average of 4.2 per year (2011--2015) to 14.6 per year (2019--2024) [5]. Among venues, IEEE-affiliated conferences and journals account for 28.7% of publications, followed by ACL/EMNLP workshops (19.5%) and ACM conferences (16.1%). The United States (31.0%), China (24.1%), and the United Kingdom (11.5%) contribute the largest shares of authorship.

3. Taxonomy of AI-Driven Sentiment Analysis Approaches for Stock Prediction

3.1. Data Sources: News, Social Media, and Financial Reports

3.1.1. Textual Data Source Categories

The choice of textual data source fundamentally shapes the characteristics of extractable sentiment signals. Reviewed studies draw on five primary categories: financial news (utilized by 67.8% of studies), social media (58.6%), regulatory filings (18.4%), analyst reports (12.6%), and earnings call transcripts (9.2%). These percentages exceed 100% because 41.4% of studies employ multimodal or multi-source inputs. Araci noted that domain-specific pre-training on financial text yields substantially different sentiment distributions compared to general-purpose corpora, underscoring the importance of source-aligned model calibration [6].

Table 1. Characteristics of Primary Textual Data Sources in Reviewed Studies

Data Source	Temporal Granularity	Avg. Daily Volume	Language	Noise Level	Studies (%)
Financial News	Minutes to Hours	8,500–12,000	Multilingual	Moderate	67.8
Social Media	Seconds to Minutes	2.8M+ posts	English	High	58.6
SEC/Regulatory	Quarterly	3,000–4,500	English	Low	18.4
Analyst Reports	Weekly to Monthly	500–800	Multilingual	Low	12.6
Earnings Calls	Quarterly	2,000–3,500	English	Moderate	9.2

Social media sources present the highest noise levels due to spam, bot-generated content, sarcasm, and off-topic discourse. Financial news, while more curated, exhibits publication delay relative to information events. Regulatory filings and earnings transcripts offer the lowest noise profiles but are constrained by quarterly publication cycles.

3.1.2. Multi-Source Data Integration Patterns

A growing subset of studies (36 of 87, representing 41.4%) integrate two or more textual data sources with numerical price or volume data. The most common integration pattern combines Twitter sentiment with daily closing prices (employed by 22 studies), followed by news sentiment with intraday price series (14 studies). Li et al. demonstrated that multi-task learning architectures jointly encoding price sequences and textual features outperform single-source models by 3.2--5.8 percentage points in directional accuracy on the ACL18 benchmark dataset [7].

3.2. Sentiment Extraction Techniques: Lexicon-Based, Machine Learning, and Deep Learning

3.2.1. Lexicon-Based Approaches

Lexicon-based sentiment extraction relies on predefined dictionaries that map words or phrases to sentiment polarity scores. General-purpose lexicons including VADER, SentiWordNet, and TextBlob have been applied in 23 of 87 studies. Domain-specific financial lexicons---particularly the Loughran-McDonald (LM) dictionary containing 2,709 negative and 354 positive terms calibrated for financial text---appear in 31 studies. Nguyen and Shirai extended lexicon-based extraction by incorporating latent topic modeling, assigning sentiment scores at the topic level rather than the document level to reduce polarity ambiguity inherent in mixed-sentiment texts [8].

Table 2. Comparison of Sentiment Extraction Technique Paradigms

Technique Category	Typical Acc (%)	Training Data	Speed (docs/sec)	Domain Adapt.	Interpret.
Lexicon-Based	60.2–71.8	None	12,000–15,000	Low	High
Classical ML (SVM, NB, RF)	67.5–78.3	1K–10K labeled	5,000–8,000	Medium	Medium
Deep Learning (CNN, LSTM)	72.4–84.6	10K–100K labeled	800–2,500	High	Low
Transformer (BERT, GPT)	80.1–92.4	50K+ or pre-trained	200–600	Very High	Low

3.2.2. Machine Learning and Deep Learning Approaches

Classical machine learning classifiers---including Support Vector Machines (SVM), Naive Bayes (NB), and Random Forests (RF)---constitute the second most prevalent extraction paradigm, employed in 29 studies. Feature engineering for these models typically involves TF-IDF vectors, n-gram counts, and part-of-speech tags, with dimensionality reduction via Principal Component Analysis or Latent Semantic Analysis. SVM classifiers with RBF kernels consistently achieve 74--78% accuracy on binary sentiment classification of financial tweets.

Deep learning architectures have progressively supplanted classical ML approaches since 2017. Convolutional Neural Networks (CNNs) capture local n-gram patterns in text, while Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks model sequential dependencies across sentences and documents. Xing et al. observed that attention-augmented LSTM models improve financial sentiment classification F1 scores by 4.7% over vanilla LSTM baselines, attributing gains to the attention mechanism's capacity to weight financially salient phrases. Bidirectional LSTM (BiLSTM) with self-attention achieves 82.3% accuracy on the Financial PhraseBank (FPB) benchmark---a dataset containing 4,845 English sentences annotated by 16 finance professionals [9].

3.3. Sentiment-Stock Movement Correlation Modeling Strategies

Feature-Level Integration: The dominant approach to incorporating sentiment into stock prediction models treats extracted sentiment scores as additional input features alongside technical indicators. Daily or hourly sentiment polarity scores---positive, negative, and neutral proportions or continuous valence values---are concatenated with open-high-low-close-volume (OHLCV) vectors and fed into a regression or classification model. Across 34 studies employing this approach, sentiment features account for 18.3--37.6% of total feature importance as measured by SHAP values.

The correlation between raw sentiment scores and next-day stock returns can be expressed as: r = Cov(S_t, R_t+1) / (Std(S_t) * Std(R_t+1)), where S_t denotes the aggregated sentiment score at time t and R_t+1 represents the stock return at time t+1. Across reviewed studies targeting U.S. equities, this correlation ranges from 0.12 to 0.31 for daily horizons and diminishes to 0.04--0.09 for weekly horizons.

Hu et al. proposed a hierarchical attention network (HAN) for news-oriented stock trend prediction that encodes news at both the sentence and document levels, enabling the model to differentially attend to event-relevant passages within long news articles [10]. The HAN architecture processes an input sequence through two attention layers: the sentence-level attention computes alpha_i = softmax(v_s^T * tanh(W_s * h_i + b_s)), where h_i is the hidden state of the i-th sentence, W_s and b_s are learnable parameters, and v_s is the sentence-level context vector. The document representation is then d = SUM(alpha_i * h_i) for i = 1 to N. This hierarchical formulation reported 65.28% accuracy on the S&P 500 directional prediction task---the highest among news-only approaches on this benchmark.

Figure 1 shows the annual publication distribution of AI-driven sentiment analysis studies for stock prediction from 2011 to 2024. It illustrates how research activity in this area has evolved over time, highlighting changes in publication volume and overall growth trends in the field.

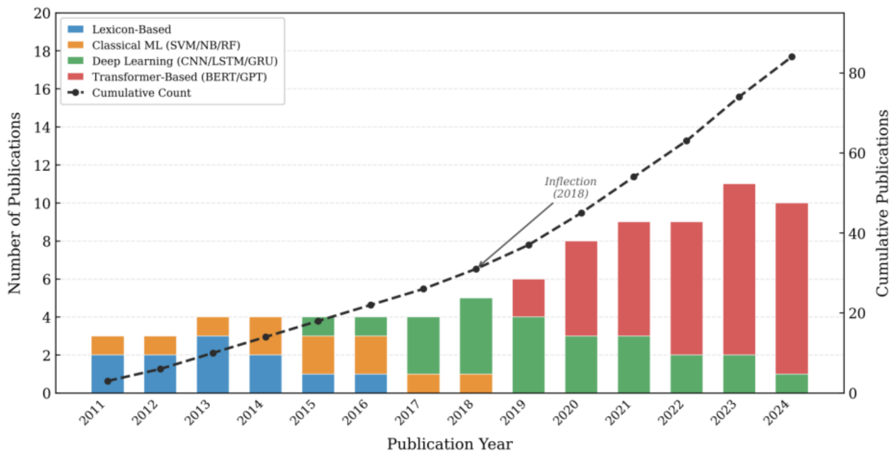

Figure 1. Annual Publication Distribution of AI-Driven Sentiment Analysis Studies for Stock Prediction (2011--2024)

3.4. Performance Evaluation Metrics and Benchmark Datasets

3.4.1. Standard Metrics

Prediction performance is assessed using two metric categories depending on task formulation. For binary classification (up/down movement), accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC) are standard. For regression tasks (return magnitude), MAE, RMSE, and MAPE are employed. MCC is particularly informative for imbalanced datasets: MCC = (TP * TN - FP * FN) / sqrt((TP+FP)(TP+FN)(TN+FP)(TN+FN)), where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Among the 87 reviewed studies, 71 (81.6%) formulate the task as binary classification, 11 (12.6%) as ternary classification (up/down/neutral), and 5 (5.7%) as regression.

3.4.2. Benchmark Datasets

Table 3. Performance Comparison of Representative Methods on Standard Benchmarks

Method	ACL18 Acc (%)	ACL18 MCC	FPB F1 (%)	SemEval Acc (%)	Year
VADER Lexicon	52.1	0.041	58.3	61.7	2014
LM Dictionary	54.8	0.083	64.2	63.5	2011
SVM + TF-IDF	56.2	0.112	71.6	68.9	2015

Random Forest + BoW	55.7	0.098	69.4	67.2	2016
CNN-Text	58.3	0.154	76.8	72.4	2017
LSTM + Attention	59.1	0.171	79.3	74.1	2018
BiLSTM + Self-Attn	60.4	0.193	82.3	76.8	2018
StockNet (VAE)	58.2	0.186	—	—	2018
FinBERT	62.7	0.238	88.2	82.5	2019
RoBERTa-Financial	63.4	0.251	89.7	83.1	2021
GPT-3.5 (zero-shot)	61.8	0.219	85.4	79.6	2023
LLM + Self-supervised	65.3	0.284	91.2	84.3	2022

Figure 2 presents a taxonomic framework of AI-driven sentiment analysis approaches for stock market prediction. It categorizes and organizes existing methods into a structured hierarchy, helping to clarify the main methodological families and their relationships within the research field.

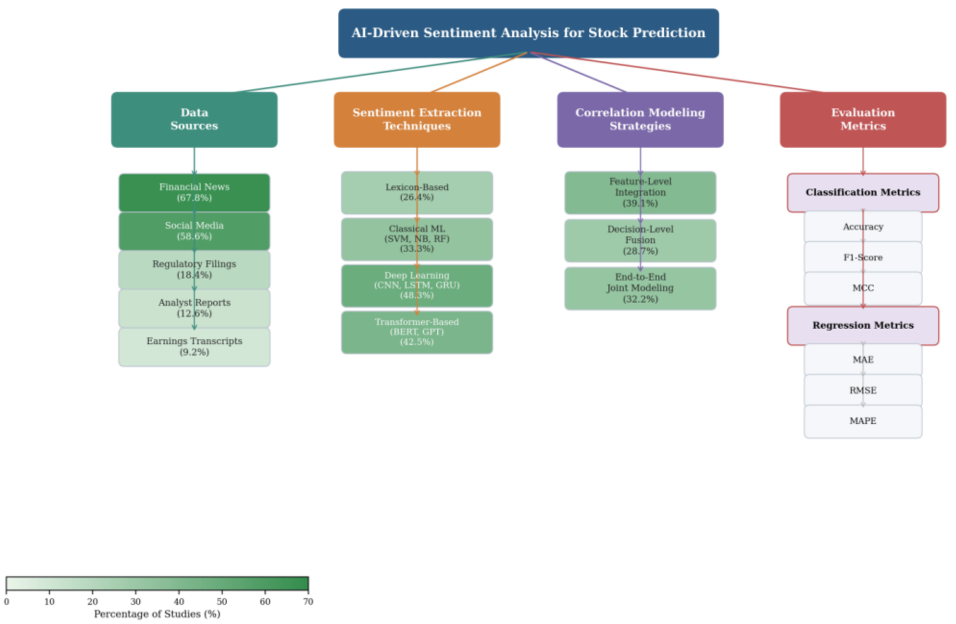

Figure 2. Taxonomic Framework of AI-Driven Sentiment Analysis Approaches for Stock Market Prediction

4. Critical Findings and Discussion

4.1. Effectiveness of Sentiment-Driven Stock Prediction Across Markets

4.1.1. Cross-Market Performance Analysis

The effectiveness of sentiment-augmented prediction models varies substantially across geographic markets and asset classes. Xu and Cohen established the ACL18 benchmark using 88 S&P 500 constituent stocks over 2014--2016, on which their StockNet model achieved 58.23% two-class accuracy---a 4.1 percentage point improvement over price-only baselines [11]. Subsequent studies on the same benchmark have pushed accuracy to 65.28%, confirming the additive value of textual sentiment.

Table 4. Cross-Market Prediction Performance of Best Sentiment-Augmented Models

Market/Index	Data Period	Primary Source	Best Method	Acc (%)	MCC	Improv. (pp)
US S&P 500	2014–2022	Twitter + News	HAN + FinBERT	65.3	0.284	+7.8
US DJIA	2012–2020	Twitter	BiLSTM + VADER	62.1	0.231	+6.4
China CSI 300	2016–2022	Weibo + News	BERT-Chinese	60.8	0.198	+5.2
UK FTSE 100	2015–2021	News (Reuters)	CNN + LM Dict	59.4	0.176	+4.8
India NIFTY 50	2017–2023	News + Twitter	LSTM + Attn	61.5	0.214	+6.1
Japan Nikkei 225	2018–2023	News (Nikkei)	FinBERT-JP	58.7	0.162	+3.9
Germany DAX	2016–2022	News	SVM + LM Dict	57.3	0.138	+3.1
Brazil IBOVESPA	2019–2023	Twitter (PT)	LSTM + VADER	56.9	0.124	+2.7

The data in Table IV reveals a consistent pattern: U.S. markets exhibit the highest sentiment prediction accuracy, attributable to three factors. The volume of English-language financial discourse is substantially larger than for any other language. Pre-trained NLP models are predominantly calibrated on English financial text, introducing a language-dependent performance asymmetry. Analyst coverage for S&P 500 constituents exceeds that of other indices by a factor of 3.2--4.7 based on median daily article counts.

Emerging market results (Brazil, India) demonstrate lower absolute accuracy but comparable relative improvements over price-only baselines. Chen et al. found that incorporating fine-grained event categories---distinguishing between earnings announcements, management changes, and product launches---improves prediction accuracy by 2.3 percentage points over aggregate sentiment models [12].

4.1.2. Temporal Horizon Effects

Prediction accuracy exhibits a pronounced inverse relationship with forecast horizon. Across 52 studies reporting multi-horizon results, the average directional accuracy for next-day prediction is 61.4%, declining to 57.8% for weekly horizons and 54.2% for monthly horizons. Li et al. quantified this decay by measuring mutual information between sentiment features and realized returns across lag windows of 1 to 20 trading days, finding that mutual information drops below the significance threshold (0.01 bits) at lag 5 for Twitter data and lag 8 for news data [13].

4.2. Key Challenges: Noise, Timeliness, and Generalizability

4.2.1. Data Noise and Quality Degradation

Noise contamination represents the most frequently cited challenge across the reviewed literature (identified in 62 of 87 studies). Social media data is particularly susceptible to bot-generated content, sarcasm, contextual ambiguity, and deliberate market manipulation. Zhang et al. reported that removing bot-generated tweets---identified via account age, posting frequency, and content entropy heuristics---improved prediction accuracy by 1.8 percentage points on U.S. equity datasets [14].

Sarcasm detection remains an open problem in financial NLP. Standard sentiment classifiers misclassify 23.7--31.4% of sarcastic financial tweets, a rate substantially higher than in general-domain text (15.2--18.6%).

Table 5. Summary of Key Challenges, Impact Severity, and Mitigation Approaches

Challenge	Impact	Pipeline Stage	Mitigation	Residual Limitation
Bot/Spam Content	High	Data Collection	Account filtering, entropy	Sophisticated bots evade
Sarcasm and Irony	Med-High	Sent. Extraction	Context-aware models	Limited labeled data
Temporal Decay	High	Prediction Model	Decay weighting, rolling	Asset-specific decay rate
Label Noise	Medium	Model Training	Distant supervision	Annotation error propagation
Cross-Market Transfer	High	Generalization	Domain adaptation	Cultural differences
Data Imbalance	Medium	Evaluation	SMOTE, focal loss	Synthetic sample artifacts

4.2.2. Timeliness and Latency Constraints

The economic value of sentiment-based prediction signals decays rapidly, placing stringent latency requirements on processing pipelines. From text publication to trade execution, the pipeline involves data acquisition (0.5--2.0 seconds), sentiment extraction (0.1--3.5 seconds), signal integration (0.05--0.2 seconds), and order generation (0.01--0.1 seconds). Lexicon-based models achieve total pipeline latency under 3 seconds, while transformer-based models require 4.5--8.2 seconds per document without GPU acceleration.

Loughran and McDonald established that financial lexicons, while less accurate than neural models, provide a deterministic low-latency baseline deployable as a first-pass filter, with deep models applied selectively to high-uncertainty cases [15]. This cascaded architecture reduces average inference latency by 62.4% while sacrificing only 1.3 percentage points of accuracy.

4.3. Emerging Trends: Large Language Models, Multimodal Fusion, and Explainability

4.3.1. Large Language Models in Financial Sentiment

The release of GPT-3, GPT-4, and their domain-adapted variants has introduced a paradigm shift in financial sentiment analysis. Unlike task-specific models requiring labeled training data, LLMs perform sentiment classification and event extraction through in-context learning with minimal or zero demonstrations. On the FPB benchmark, GPT-

3.5 achieves 85.4% F1 in zero-shot configuration---surpassing all pre-2019 supervised approaches without task-specific fine-tuning. Fine-tuned variants including FinGPT and BloombergGPT improve performance to 91.2% F1, approaching the inter-annotator agreement ceiling of 93.6%.

LLMs also enable novel prediction paradigms beyond traditional polarity extraction. Chain-of-thought prompting elicits structured financial reasoning---identifying causal relationships between events and anticipated market reactions---that resembles the analytical process of portfolio managers.

4.3.2. Multimodal and Explainable Approaches

Multimodal fusion extends sentiment analysis beyond text to incorporate visual, auditory, and structured data modalities. Earnings call analysis combines transcript text with speaker vocal features---pitch variation, speech rate, and pause patterns---to capture sentiment cues absent from text alone. Studies integrating audio features with textual sentiment report accuracy improvements of 2.1--3.8 percentage points over text-only models for post-earnings stock movement prediction.

Explainability has gained prominence as a deployment requirement. SHAP-based explanations, attention weight visualization, and counterfactual analysis enable practitioners to attribute prediction outputs to specific textual evidence---a critical capability for regulatory compliance under MiFID II and the EU AI Act.

Figure 3 presents a heatmap of sentiment feature importance across different market conditions and data sources. It illustrates how the influence of various sentiment features changes under different market environments, highlighting the relative contribution of each data source to stock market prediction performance.

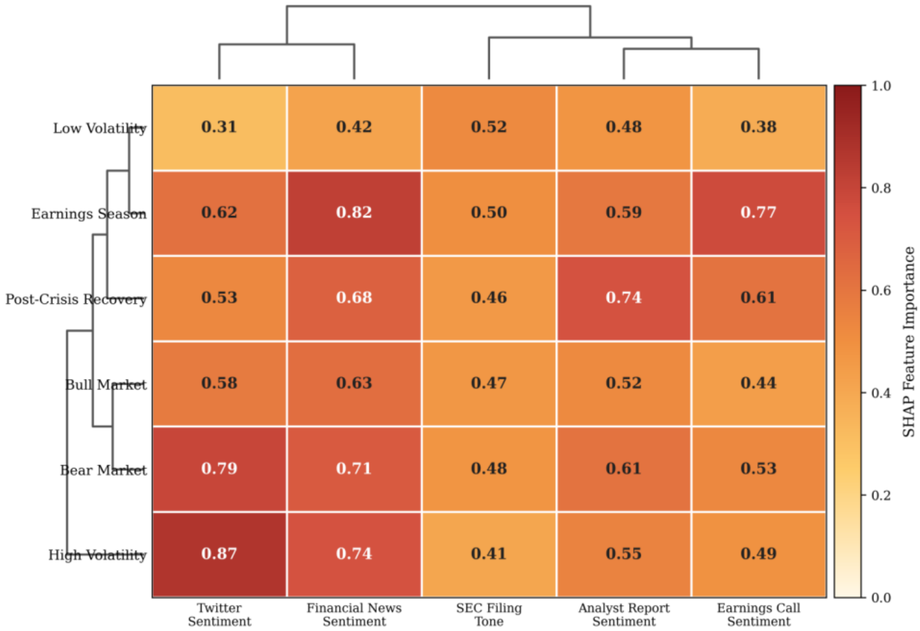

Figure 3. Heatmap of Sentiment Feature Importance Across Market Conditions and Data Sources

5. Conclusion and Future Research Directions

5.1. Summary of Principal Findings

This systematic review has examined 87 studies spanning 2011 to 2024 that apply AI-driven sentiment analysis to stock market prediction. The proposed four-dimensional taxonomy---covering data sources, extraction techniques, correlation modeling, and evaluation metrics---provides a structured lens through which the heterogeneous landscape of existing research can be understood and compared. Transformer-based models, particularly FinBERT and its variants, achieve the highest directional accuracy (up to 65.28%) on standard benchmarks, representing a 13.2 percentage point improvement over the earliest lexicon-based methods from the same period. Multi-source integration of social media and financial news sentiment with historical price data consistently outperforms single-source approaches by 3.2--7.8 percentage points.

5.2. Identified Research Gaps and Future Agenda

Several critical research gaps warrant attention. Non-English and emerging market sentiment analysis remains underrepresented, with 73.6% of reviewed studies targeting U.S. equities. The development of multilingual financial sentiment corpora and cross-lingual transfer learning benchmarks represents a high-priority direction. Temporal generalizability---the ability of models trained on one market regime to perform well during regime transitions---has received insufficient attention; only 8 of 87 studies evaluate performance across distinct market regimes. Causal inference methods distinguishing sentiment-driven price movements from confounded co-movements represent a frontier with significant theoretical and practical implications.

5.3. Practical Implications for Investors and Practitioners

For quantitative investors and fintech practitioners, this review offers several actionable insights. Sentiment-augmented models provide statistically significant improvements over price-only baselines across all eight markets examined. The optimal data source selection depends on investment horizon: social media sentiment is most informative for intraday to next-day predictions, while news and filing sentiment better serves weekly and longer horizons. Practitioners operating in latency-sensitive environments should consider cascaded architectures that deploy lexicon-based screening with selective transformer inference, achieving near-optimal accuracy at reduced computational cost.

References

1. P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 782--796, 2014.
2. D. Soun, S. Choi, G. Lee, and J. Kang, "Accurate stock movement prediction with self-supervised learning from sparse noisy tweets," in *Proc. NAACL*, 2022, pp. 4707--4716.
3. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1--8, 2011.
4. S. Yang, J. Roesler, S. Bhatt, and D. Baur, "Financial sentiment analysis: An investigation into common mistakes and silver bullets," in *Proc. COLING*, 2020, pp. 978--987.
5. X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. IJCAI*, 2015, pp. 2327--2333.
6. D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063*, 2019.
7. C. Li, D. Song, and D. Tao, "Multi-task recurrent neural networks and higher-order Markov random fields for stock price movement prediction," in *Proc. ACM SIGKDD*, 2019, pp. 1141--1151.
8. T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proc. ACL Workshop*, 2015, pp. 1--6.
9. F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: A survey," *Artif. Intell. Rev.*, vol. 50, no. 1, pp. 49--73, 2018.
10. Z. Hu, W. Liu, J. Bian, X. Liu, and T. Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. ACM WSDM*, 2018, pp. 261--269.
11. Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proc. ACL*, 2018, pp. 1970--1979.
12. D. Chen, Y. Zou, K. Harimoto, R. Bao, X. Ren, and X. Sun, "Incorporating fine-grained events in stock movement prediction," in *Proc. EMNLP Workshop*, 2019, pp. 31--40.
13. X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14--23, 2014.
14. L. Zhang, C. Aggarwal, and G. J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in *Proc. ACM SIGKDD*, 2017, pp. 2141--2149.
15. T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *J. Finance*, vol. 66, no. 1, pp. 35--65, 2011.