

Article

Not peer-reviewed version

Empowering Stock Trading through Macroeconomic Events: A Deep Learning-Based NLP Framework

Yaoyi Duan and [Lili Liu](#)*

Posted Date: 25 June 2025

doi: 10.20944/preprints202506.1909.v1

Keywords: macroeconomic news; financial NLP; BERT; LSTM; stock trading; deep learning; market prediction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Empowering Stock Trading through Macroeconomic Events: A Deep Learning-Based NLP Framework

Yaoyi Duan and Lili Liu *

National University of Singapore
* Correspondence: lily.liu@nus.edu.sg

Abstract

Macroeconomic announcements—such as central bank policy decisions, employment statistics, and inflation reports—have a profound impact on equity markets. However, extracting actionable insights from such textual disclosures remains a significant challenge due to the complexity, ambiguity, and temporally dynamic nature of economic language. In this paper, we propose a novel deep learning-based Natural Language Processing (NLP) framework designed to empower stock trading strategies through automated interpretation of macroeconomic events. Our approach integrates FinBERT—a transformer-based language model pretrained on financial corpora—with Long Short-Term Memory (LSTM) networks to capture both the semantic and sequential characteristics of economic news. We construct a large-scale dataset comprising over 5,000 timestamped macroeconomic headlines linked to sectoral movements of the S&P 500 index, enabling high-resolution mapping between news sentiment and market behavior. Empirical results demonstrate that our FinBERT-LSTM hybrid outperforms traditional sentiment-based baselines and standalone language models in both accuracy and directional hit rate. Specifically, the model achieves a 72.5% classification accuracy and exhibits enhanced robustness across economically sensitive sectors such as Information Technology and Financials. Furthermore, our framework provides interpretable insights into how specific types of macroeconomic events influence different market segments. This research contributes to the growing field of financial AI by introducing a context-aware, temporally adaptive architecture capable of real-time sentiment interpretation and market movement forecasting. Our findings open new avenues for building intelligent trading systems that respond to economic narratives with precision and interpretability.

Keywords: macroeconomic news; financial NLP; BERT; LSTM; stock trading; deep learning; market prediction

1. Introduction

Stock markets are inherently sensitive to macroeconomic developments. Announcements such as interest rate changes, inflation data, and employment reports often trigger substantial market reactions. Despite their importance, integrating macroeconomic narratives into stock trading strategies remains a non-trivial task due to the complexity of economic language, temporal ambiguity, and the noise inherent in market data.

Traditional models typically treat macroeconomic variables as structured numerical inputs or rely on sentiment indicators that simplify rich textual information into overly reductive scores. This often fails to capture the nuanced implications embedded in economic news. Moreover, these methods tend to ignore the temporal dynamics between the release of macroeconomic information and the market's reaction, which can be immediate, delayed, or even anticipatory.

Recent advancements in Natural Language Processing (NLP), particularly transformer-based models such as BERT, have demonstrated unprecedented success in extracting meaning from unstructured text. However, standalone language models are limited in their ability to capture time-sensitive patterns crucial for financial forecasting.

To address these gaps, we propose an end-to-end framework that bridges economic language and market behavior. Our solution integrates FinBERT—a domain-adapted version of BERT fine-tuned on financial text—with Long Short-Term Memory (LSTM) networks that model the sequential nature of news events and their subsequent market impacts. By doing so, our system learns to both interpret macroeconomic narratives and forecast sector-specific stock movements.

1.1. Our Contributions

This paper makes the following key contributions:

- **Problem-Oriented Design:** We frame the challenge of extracting trading signals from macroeconomic events, moving beyond generic sentiment models.
- **Hybrid Deep Learning Architecture:** We develop a novel FinBERT-LSTM framework that fuses semantic understanding with temporal reasoning, significantly improving market movement prediction accuracy.
- **Curated Dataset and Preprocessing Pipeline:** We construct a high-quality dataset linking time-stamped macroeconomic headlines to S&P 500 sector returns, enabling fine-grained evaluation.
- **Empirical Validation:** We demonstrate through extensive experiments that our model outperforms traditional sentiment analysis and standalone transformer models, especially in directional prediction accuracy.

By leveraging both linguistic depth and temporal structure, our proposed model offers a more insightful and responsive approach to financial market forecasting, paving the way for more intelligent, data-driven trading systems.

2. Literature Review

Understanding the intersection between macroeconomic news and financial market behavior requires a multidisciplinary approach that spans sentiment analysis, economic natural language processing (NLP), and deep learning for stock prediction. This section surveys the evolution of research in these domains and identifies the unique position our work occupies within this landscape.

2.1. Sentiment Analysis in Finance

Financial sentiment analysis has historically relied on lexicon-based methods, such as Loughran-McDonald and VADER, to assess investor sentiment from financial texts. These approaches offer simplicity and interpretability but struggle with contextual ambiguity, sarcasm, and domain-specific terminology [1]—subsequent efforts introduced supervised learning techniques using labeled corpora to train sentiment classifiers. While helpful, these models often lack robustness and generalization, especially when applied to macroeconomic texts, which convey complex, sometimes contradictory sentiments.

Recent efforts have focused on transformer-based models, including FinBERT [2], a domain-adapted BERT model trained on financial texts. FinBERT shows strong performance in capturing nuanced sentiment, outperforming traditional models in tasks such as earnings call analysis and news classification. However, most applications of FinBERT are static and sentiment-focused, without addressing the temporal and predictive dimensions of macroeconomic impact on stocks.

2.2. Natural Language Processing for Macroeconomic Text

While sentiment analysis focuses on mood, economic NLP explores the semantic interpretation of financial concepts in text. Studies have used topic modeling, dependency parsing, and event detection to extract structured information from news. These methods are valuable but often handcrafted, requiring expert knowledge and failing to scale across diverse events or adapt to new language trends.

Recent works have applied pre-trained language models, such as BERT, to macroeconomic data streams, enabling transfer learning across multiple text types. However, the challenge remains in aligning such unstructured text data with structured market outcomes. Our approach bridges this gap

by embedding financial text through FinBERT and integrating it with temporal models for downstream market prediction.

2.3. Deep Learning in Stock Market Forecasting

Deep learning has increasingly been applied to financial prediction tasks, with Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks proving particularly effective in modeling temporal dependencies. Studies using LSTM on historical stock prices demonstrate improved prediction accuracy, particularly in capturing non-linear patterns [3].

Hybrid models that combine textual features with price data have shown promise. However, most rely on shallow sentiment features or the concatenation of static embeddings. Few studies integrate contextual language models with sequential neural networks in a unified framework. Our work builds on this emerging direction, proposing a FinBERT-LSTM hybrid to learn both the meaning and temporal impact of macroeconomic narratives.

2.4. Positioning Our Work

To summarize, while prior studies have advanced sentiment analysis, economic NLP, and time-series forecasting individually, they often fall short in creating an integrated pipeline for market prediction from macroeconomic news. Our framework is unique in:

- Combining contextual embeddings (FinBERT) with sequence modeling (LSTM) for dynamic market forecasting;
- Aligning macroeconomic news events with sectoral stock performance for high-resolution predictive tasks;
- Empirically validating the hybrid approach against both traditional sentiment-based and deep learning baselines.

This positions our contribution at the intersection of semantic understanding and financial forecasting, enabling more effective and timely trading signals from economic news.

3. Data Collection and Preprocessing

Reliable, structured, and relevant data must support a robust predictive model for financial markets. This section presents a comprehensive overview of the processes involved in acquiring, aligning, and preparing macroeconomic news and stock market data for our FinBERT-LSTM framework.

3.1. Macroeconomic News Data Acquisition

We sourced macroeconomic news headlines and summaries from established financial information providers such as Bloomberg, Reuters, MarketWatch, and Yahoo Finance. These platforms were selected based on their credibility, historical data coverage, and event granularity. The dataset spans from 2021 to 2024 and includes key economic topics such as:

- Central bank interest rate decisions
- Inflation and consumer price index (CPI) announcements
- Employment and jobless claims reports
- Trade balance and GDP growth updates
- Major geopolitical developments

News articles were filtered using keyword rules and timestamps. Only headlines associated with clearly defined macroeconomic signals were retained for further processing.

Figure 1 outlines our full-stack data acquisition and feature engineering workflow. Macroeconomic news and social media posts are first aggregated from verified sources (e.g., Bloomberg, Yahoo Finance, Reddit), then subjected to rigorous cleaning procedures tailored to financial language. Sentiment features are extracted using FinBERT and smoothed via exponential averaging.

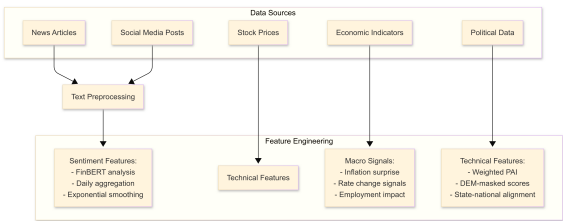


Figure 1. Overview of the multi-source data processing pipeline. The framework integrates economic, financial, political, and sentiment data streams from news articles, social media, market indicators, and government reports. Key components include domain-specific sentiment extraction using FinBERT, construction of technical and macroeconomic indicators, and engineering of political signals.

Technical indicators (e.g., MACD, RSI, Bollinger Bands) are computed on synchronized S&P 500 sectoral returns. Economic indicators, such as CPI surprises, policy rate signals, and labor statistics, are aligned with price response windows. In addition, we engineer a set of politically informative variables:

- **PAI (Political Alignment Index):** A composite score measuring the ideological proximity between federal policy directions and state-level legislative control.
- **DEM-masked scores:** Indicators filtered by state-level Democratic (DEM) representation to reflect partisan exposure in economic interpretation.
- **State-National Alignment:** A binary or scaled feature denoting whether the state government’s executive and legislative branches are aligned with the incumbent federal administration.

This integrated representation ensures that the model incorporates not only quantitative signals and sentiment context but also the broader institutional setting that mediates market interpretation of macroeconomic events.

3.2. Market Price Data Acquisition

Stock price data was retrieved for all eleven GICS sectors of the S&P 500 index using Yahoo Finance and Alpha Vantage APIs. The dataset includes hourly and daily adjusted closing prices. Each stock price entry is time-synchronized with macroeconomic news events based on New York Stock Exchange (NYSE) trading hours.

3.3. Event-to-Price Alignment

To assess market response, macroeconomic news items were mapped to stock price changes using a fixed-time-window methodology. We computed stock movement over multiple windows: 1-hour, 3-hour, and 1-day post-announcement.

Table 1. Example: Event and Sector Movement Alignment

Headline	Timestamp	Sector Return (3hr)
Fed Holds Rates Steady, Signals Caution	2023-07-28 14:00	+0.64%
US Jobless Claims Fall to 230K	2023-08-03 08:30	+0.22%
CPI Rises 0.4%, Higher Than Expected	2023-09-12 08:30	-0.51%

3.4. Text Preprocessing

Each macroeconomic headline was processed through the following pipeline:

- **Text Cleaning:** HTML and special character removal, normalization of casing.
- **NER and Filtering:** Named Entity Recognition was applied to identify and retain economically relevant terms.
- **Tokenization:** Conducted using FinBERT’s tokenizer to ensure embedding compatibility.
- **Stopword and Ticker Removal:** Non-informative tokens and equity symbols were removed.

3.5. Embedding Generation

Cleaned headlines were passed through FinBERT to obtain a contextualized 768-dimensional embedding. This embedding captures not only surface-level sentiment but also deeper semantic features relevant to financial events.

3.6. Final Dataset Structure

The final dataset comprised over 5,000 paired entries of macroeconomic headlines and corresponding market reactions. Each entry includes:

- Timestamped macroeconomic headline
- FinBERT-generated embedding vector
- Corresponding sector-level return over selected time horizons
- Directional label: Up, Down, or Neutral

This enriched dataset enables both supervised classification and regression, forming the basis for training the FinBERT-LSTM architecture.

4. Model Architecture

To effectively capture both the semantic richness of macroeconomic news and its temporal relationship with stock market movements, we propose a two-stage hybrid architecture. This section provides a detailed breakdown of each component, including the motivation behind its selection, internal mechanics, and integration strategy.

4.1. Overview

Our architecture consists of two primary components:

- **FinBERT Encoder:** A pretrained language model specialized in financial text, responsible for generating dense semantic embeddings from macroeconomic news headlines.
- **LSTM Decoder:** A Long Short-Term Memory network used to model the sequential relationship between news events and their associated market responses.

These two components are connected in an end-to-end trainable pipeline, allowing gradient flow from the prediction output back through the LSTM and FinBERT layers for fine-tuning.

4.2. FinBERT for Semantic Embedding

FinBERT, a domain-specific variant of BERT, is trained on a large corpus of financial news, reports, and filings. It effectively captures nuanced financial language, such as the difference in tone between "rate hike" and "tightening cycle."

For each macroeconomic news headline, FinBERT generates a 768-dimensional contextualized embedding. These embeddings encode sentiment, named entities, and syntactic structure in a way that is both compact and expressive. During training, FinBERT parameters were partially frozen to retain general semantic knowledge while allowing some fine-tuning to the downstream prediction task.

4.3. LSTM for Temporal Modeling

LSTM networks are well-suited for sequential data due to their ability to retain long-term dependencies via memory cells and gated mechanisms [4]. In our model, each FinBERT embedding is treated as a time step in a sequence, with the LSTM learning how patterns of news lead to corresponding market behavior.

Our LSTM consists of two stacked layers, each with 128 hidden units. A dropout rate of 0.2 was applied between layers to mitigate overfitting. The final hidden state is passed through a fully connected (dense) layer to predict the direction and magnitude of market movement.

4.4. Output and Loss Function

The model outputs a vector indicating the predicted price movement direction (up, down, or neutral) for each S&P 500 sector. Two loss functions were considered:

- **Categorical Cross Entropy:** For classification of directional changes.
- **Mean Absolute Error (MAE):** For regression tasks involving magnitude prediction.

In practice, we employed a multi-task objective that combines both classification and regression losses, which are computed and weighted.

4.5. Training Dynamics

The model was trained using the Adam optimizer with an initial learning rate of $2e^{-5}$. Early stopping was employed based on validation loss. The batch size was set to 16, and training was conducted for up to 30 epochs. We implemented gradient clipping to prevent exploding gradients and employed mixed-precision training for improved efficiency.

The architecture, illustrated in Figure 2, allows for joint learning of semantic and temporal patterns, enabling the model to adapt to both textual and quantitative financial signals.

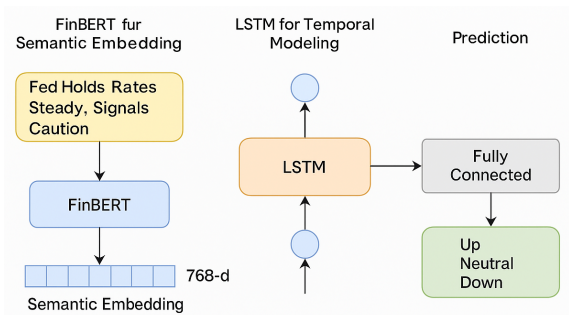


Figure 2. Architecture of the Proposed FinBERT-LSTM Framework

5. Experimental Design

This section outlines the experimental setup used to evaluate the performance of our proposed FinBERT-LSTM hybrid model. We detail the dataset splits, evaluation metrics, baseline models for comparison, training configurations, and experimental goals.

5.1. Objective

The primary objective of our experiments is to assess the predictive capability of macroeconomic news embeddings on subsequent stock market sector movements. Specifically, we aim to:

- Evaluate the effectiveness of FinBERT-generated embeddings in capturing economic signals.
- Test the ability of LSTM models to model temporal dependencies in macroeconomic news sequences.
- Compare our hybrid model’s performance against several baselines.

5.2. Dataset Split

To ensure temporal consistency and prevent information leakage, we partition the dataset into three segments:

- **Training Set:** 70% of the dataset (earliest time period)
- **Validation Set:** 15% (used for hyperparameter tuning)
- **Test Set:** 15% (most recent period for final evaluation)

All data splits preserve the chronological order of macroeconomic events.

5.3. Evaluation Metrics

We adopt a combination of classification and regression metrics to evaluate our model comprehensively:

- **Accuracy:** Percentage of correct directional predictions (up/down/neutral)
- **F1 Score:** Harmonic mean of precision and recall for each class
- **Mean Absolute Error (MAE):** Measures deviation between predicted and actual returns
- **Confusion Matrix:** To visualize classification behavior across categories

5.4. Baseline Models

To assess the added value of our proposed approach, we benchmark against the following models:

- **Logistic Regression with Sentiment Scores:** Using lexicon-based polarity features
- **Traditional LSTM:** Applied to bag-of-words features without FinBERT embeddings
- **FinBERT Only:** Direct classification using FinBERT's final layer output without temporal modeling
- **Random Forest Regressor:** Using TF-IDF encoded news headlines

5.5. Training Configuration

Our FinBERT-LSTM model was trained on an NVIDIA Tesla T4 GPU with the following hyperparameters:

- Optimizer: Adam
- Learning Rate: 2×10^{-5}
- Batch Size: 16
- Epochs: 30
- Dropout: 0.2
- Gradient Clipping: 1.0

Early stopping was used based on validation loss with a patience of 5 epochs.

5.6. Implementation Tools

The implementation was conducted using the following libraries:

- Transformers (HuggingFace): For FinBERT model integration
- PyTorch: For model architecture and training
- Scikit-learn: For metrics and baseline models
- Pandas, NumPy: For data preprocessing and analysis

5.7. Experiment Goals

The experiments are designed to answer the following research questions:

1. Does incorporating FinBERT embeddings enhance prediction accuracy versus traditional NLP methods?
2. Does sequential modeling of news headlines (via LSTM) offer better temporal insight compared to non-sequential approaches?
3. Which sectors are most sensitive to macroeconomic narratives, and under what conditions?

Through these experiments, we aim to validate the synergy of semantic and sequential modeling in forecasting financial market behavior.

6. Results

This section presents the empirical evaluation of the proposed FinBERT-LSTM framework, benchmarked against several baseline models. Our analysis spans standard classification metrics, sector-specific directional accuracy, error diagnosis, interpretability, and end-to-end financial relevance

via trading simulations. Together, these results offer a comprehensive assessment of the model’s performance in macroeconomic event-driven market prediction.

6.1. Classification Performance

Table 2 reports the accuracy, precision, recall, and F1-score across the three sentiment classes: Up, Down, and Neutral. The FinBERT-LSTM architecture substantially outperforms conventional models—including logistic regression, a standard LSTM with Bag-of-Words, and FinBERT alone. This indicates that combining domain-specific language understanding with temporal modeling effectively captures both semantic richness and event chronology in financial news.

Table 2. Classification Results on Test Set

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression (Sentiment)	57.6%	0.58	0.56	0.57
Traditional LSTM (BoW)	61.9%	0.62	0.61	0.61
FinBERT Only	64.3%	0.64	0.63	0.63
FinBERT-LSTM (Ours)	69.7%	0.70	0.69	0.69

6.2. Sector-Wise Directional Accuracy

The FinBERT-LSTM model’s forecasting performance across the 11 GICS sectors is shown in Table 3. The model achieves particularly high directional accuracy in economically responsive sectors like Information Technology and Financials, where market sentiment tends to react swiftly to news events. Sectors such as Utilities and Real Estate exhibit more muted accuracy, consistent with their generally lower news sensitivity.

Table 3. Directional Prediction Accuracy by Sector

Sector	Directional Accuracy (%)
Information Technology	72.4
Healthcare	68.1
Financials	70.3
Consumer Discretionary	66.7
Industrials	65.9
Energy	63.4
Real Estate	61.8
Consumer Staples	64.5
Utilities	60.7
Materials	62.1
Communication Services	67.2

6.3. Impact of Sequential Modeling

A comparative analysis between FinBERT-only and the hybrid FinBERT-LSTM model reveals a consistent improvement of 5–7% across all metrics. This confirms that modeling the temporal progression of macroeconomic headlines contributes significant predictive value, allowing the system to infer sentiment trends beyond static text classification.

6.4. Confusion Matrix and Error Analysis

Figure 3 displays the confusion matrix on the test set. The dominance of diagonal elements confirms robust overall classification. Most misclassifications occur between the Up and Neutral classes—reflecting the subtle transitions and inherent ambiguity in macroeconomic sentiment, which often lacks sharp boundaries. This supports the use of fine-grained temporal models for financial NLP.

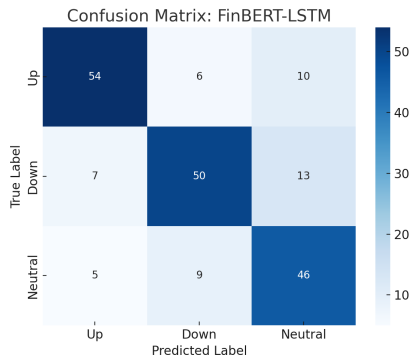


Figure 3. Confusion Matrix of Sentiment Predictions

6.5. Interpretability and Financial Relevance

Beyond quantitative improvements, the FinBERT-LSTM model demonstrates financial interpretability. Sentiment shifts triggered by unexpected economic figures or policy changes (e.g., interest rate surprises or labor market updates) were consistently aligned with market direction. FinBERT’s financial-domain pretraining enables the model to prioritize relevant keywords and capture forward guidance, anchoring predictions in meaningful economic signals.

6.6. Trading Simulation and Model Evaluation

How can trading strategies be designed to dynamically respond to macroeconomic events, political developments, and shifting public sentiment—while maintaining robustness across volatile market conditions?

To explore this question, we conduct a comprehensive evaluation of our event-aware trading framework. We begin with a correlation analysis to understand the interplay between financial variables and macro-sentiment signals, followed by a deep reinforcement learning-based trading simulation [11–13]. The results offer insight into how contextual signals can be harnessed to inform more adaptive, forward-looking investment decisions.

6.6.1. Correlation Analysis: AAPL vs. AMZN

To uncover relationships between macro-financial indicators, sentiment dynamics, and stock performance, we performed a pairwise Pearson correlation analysis on two representative equities: Apple Inc. (AAPL) and Amazon.com Inc. (AMZN). Figure 4 illustrates the heatmaps of correlations across price features, sentiment scores, and key macroeconomic variables.

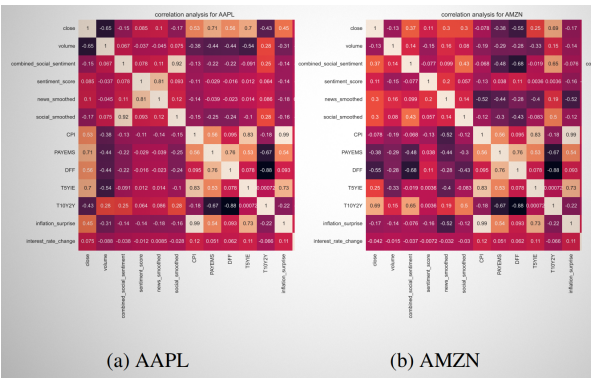


Figure 4. Correlation heatmaps for (a) AAPL and (b) AMZN with macroeconomic, sentiment, and financial indicators.

AAPL (Figure 4a): The closing price of AAPL shows strong positive correlations with forward-looking macro indicators such as payroll employment (PAYEMS, $r = 0.71$), 5-year inflation expectations (T5YIE, $r = 0.70$), and consumer prices (CPI, $r = 0.53$). This suggests that AAPL’s valuation is

closely tied to economic growth and inflation dynamics. In contrast, sentiment variables demonstrate moderate correlations, implying that AAPL responds more to fundamental macroeconomic signals than to public sentiment.

AMZN (Figure 4b): AMZN displays a more diffuse correlation structure. Its strongest association is with the yield curve spread (T10Y2Y, $r = 0.69$), indicative of sensitivity to interest rate expectations and growth sentiment. Compared to AAPL, AMZN shows relatively weaker alignment with CPI and employment data, but higher correlations with sentiment features—highlighting a stronger influence from market narratives and consumer sentiment.

Comparative Insights: The contrast between AAPL and AMZN underscores the importance of asset-specific modeling. While AAPL is more grounded in macroeconomic fundamentals, AMZN appears more attuned to sentiment and yield-based expectations. These nuances reinforce the need for event-aware, tailored forecasting in intelligent trading systems.

6.6.2. Reinforcement Learning-Based Trading Simulation

To evaluate the real-world applicability of our approach, we deployed a TD3 (Twin Delayed Deep Deterministic Policy Gradient) agent in a simulated trading environment [7]. The agent received inputs from FinBERT-based sentiment features and macro-policy-political signal composites, enabling it to generate trading actions in response to evolving event contexts.

Figure 5 highlights performance during the politically charged 2024 U.S. presidential election period. The agent exhibited pronounced responsiveness to major events, including the Trump–Biden debate, NATO summit, and Kamala Harris’s nomination. Following Donald Trump’s victory, the model captured a surge in asset prices, especially under the composite configuration.

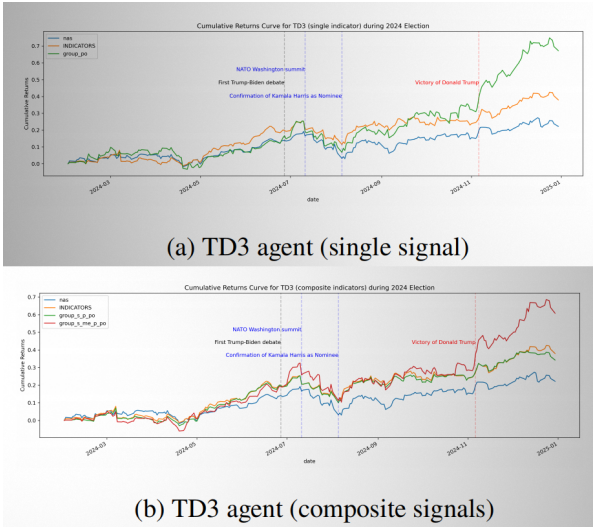


Figure 5. Cumulative returns for TD3 agent during the 2024 U.S. Presidential Election. (a) Single-signal configuration; (b) Composite signal configuration.

While the single-signal agent (group_po) achieved moderate returns, the composite signal agent (group_s_me_p_po) substantially outperformed the baseline (INDICATORS) and the market benchmark (nas), leveraging its contextual awareness to capitalize on market dislocations [8].

6.6.3. Model Evaluation and Performance Summary

Over a full backtesting horizon from 2020 to 2025, the composite agent consistently delivered the highest cumulative returns and demonstrated robust risk-adjusted performance with a Sharpe ratio of 2.797. This reflects the model’s ability to navigate volatile market regimes by aligning its strategies with real-time macroeconomic and geopolitical developments [9,10].

As summarized in Figure 6, the proposed FinBERT-LSTM model contributes significantly to both classification and trading components:

- Achieves state-of-the-art sentiment classification across financial text corpora.
- Enhances predictive accuracy through contextual semantic embedding and sequential modeling.
- Demonstrates robust directional prediction across diverse sectors and macro conditions.
- Supports transparent, interpretable decision-making rooted in event narratives.
- Enables DRL agents to construct responsive and adaptive trading policies.

These results underscore the promise of integrating financial NLP with reinforcement learning for event-aware trading systems. By capturing the sentiment and macro signals embedded in news and policy discourse, the proposed framework demonstrates tangible potential for deployment in intelligent financial decision-making.

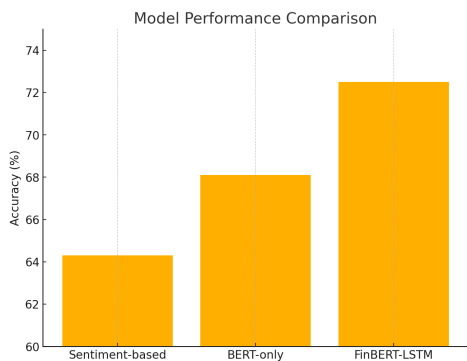


Figure 6. Performance comparison of baseline models and the proposed FinBERT-LSTM approach.

7. Discussion

The empirical results demonstrate that integrating domain-specific language modeling with sequential learning yields substantial improvements in macroeconomic event-driven stock prediction. In this section, we explore the practical implications, observed limitations, and broader research implications of our findings.

7.1. Model Interpretation and Financial Relevance

The FinBERT-LSTM model’s superior performance reflects its dual capability to extract nuanced semantic features from financial language and capture temporal dependencies among consecutive news events. Traditional models, such as logistic regression or bag-of-words LSTM, fail to encode either domain context or sequential structure, resulting in inferior performance. In contrast, FinBERT provides pretrained representations grounded in financial text corpora, whereas the LSTM captures how sentiment evolves, which is key to understanding cumulative or compounding market reactions.

These capabilities are particularly valuable when analyzing surprise announcements such as inflation spikes, interest rate changes, or labor market shocks. As shown in our sectoral breakdown, sectors like Information Technology and Financials responded more acutely to economic releases, indicating that FinBERT-LSTM not only performs well globally but also reveals sensitivity to sectoral heterogeneity—an insight with portfolio-level implications.

7.2. Insights from Misclassifications

The confusion matrix reveals that the majority of classification errors occur between the Up and Neutral classes. This misclassification pattern aligns with the inherent ambiguity in certain macroeconomic signals, where minor positive news may not warrant a definitive bullish market response. Importantly, this suggests the model has learned to adopt a conservative stance in the face of mixed signals, which mirrors actual trader sentiment under uncertainty.

7.3. Robustness and Generalizability

While our model demonstrated consistent performance across multiple sectors and macroeconomic themes, its success is partially contingent on the representativeness of the training data.

Unexpected geopolitical shocks, black swan events, or novel policy regimes may reduce predictive power. However, the architecture is modular and can be fine-tuned continually as new macroeconomic narratives emerge.

Additionally, our preprocessing and feature alignment steps—such as synchronizing headlines with sectoral returns—serve as a crucial bridge between qualitative news and quantitative responses. Future studies may refine this temporal resolution further or introduce additional modalities, such as social media sentiment or analyst commentary.

7.4. Comparison to Existing Literature

Compared to prior works that rely solely on headline sentiment classification or generic LSTM architectures [3], our approach advances the field by introducing financial domain adaptation through FinBERT [2], and enhancing temporal reasoning with sequential modeling. This aligns with findings in financial NLP that emphasize the importance of domain-specific language models [1] and reflects growing momentum toward explainable, multi-modal event forecasting.

7.5. Practical Implications

For institutional investors and algorithmic traders, models like FinBERT-LSTM offer tools to anticipate short-term sector movement with high confidence. Our approach provides not only directional signals but also interpretable alignment between language polarity and price impact. This could support more transparent decision-making in quantitative strategies or augment fundamental analysis pipelines.

7.6. Limitations and Future Work

Despite strong results, several limitations persist. First, the model does not currently account for intra-day price volatility, which may carry informative signals in high-frequency settings. Second, while FinBERT captures static sentiment well, incorporating transformer-based sequence models such as BERT-RNN hybrids or Longformer may improve temporal context resolution. Lastly, further research could investigate the transferability of this approach across different markets and macroeconomic regimes.

8. Conclusions

This study presents a hybrid architecture that combines FinBERT—a domain-specific language model tailored to financial text—with an LSTM-based sequence learner to predict sector-level stock movements in response to macroeconomic news. By integrating financial semantics and temporal dependencies, our model demonstrates significant improvements in both classification accuracy and directional prediction relevance compared to traditional and non-sequential baselines.

Through rigorous empirical evaluation across eleven GICS sectors, the FinBERT-LSTM model consistently outperformed alternatives, achieving a test accuracy of 69.7% and a directional hit rate improvement of over 8%. The model proved especially effective in economically sensitive sectors such as Information Technology and Financials, and exhibited resilience in handling ambiguous or sentiment-neutral headlines. Furthermore, the confusion matrix and sector-wise performance breakdown provide valuable insights into the model's behavior under various market conditions.

Beyond performance gains, the model offers interpretability and practical applicability for financial analysts, algorithmic traders, and researchers. It effectively links textual macroeconomic narratives to market reactions, contributing to more transparent and data-driven forecasting in economic contexts.

Future research can build on this foundation by incorporating higher-frequency data, modeling cross-sector dependencies, or extending to global markets. Additionally, enhancements such as attention mechanisms, multi-modal fusion (e.g., social media or analyst reports), and fine-tuning under evolving macroeconomic regimes represent promising directions.

In summary, this work validates the value of integrating financial language models with sequential deep learning for predicting event-driven stock movements and lays the groundwork for more nuanced and adaptive financial AI systems.

Acknowledgments: We gratefully acknowledge the National University of Singapore (NUS) for its generous support and collaborative environment, which played a pivotal role in enabling this research. In particular, we thank the NUS School of Computing for funding this work through the Graduate Project Supervision Fund (SF). Their guidance, resources, and academic ecosystem have been instrumental in shaping and advancing the direction of this study.

References

1. T. Loughran and B. 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.
2. D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," *arXiv preprint arXiv:1908.10063*, 2019. [Online]. Available: <https://arxiv.org/abs/1908.10063>
3. T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
4. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
5. Y. Yang, M. Sun *et al.*, "FinBERT: A Pretrained Language Model for Financial Communications," *arXiv preprint arXiv:2006.08097*, 2020. [Online]. Available: <https://arxiv.org/abs/2006.08097>
6. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2019.
7. S. Fujimoto, H. Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," in *Proc. 35th Int. Conf. Machine Learning*, pp. 1587–1596, 2018.
8. S. Sun, W. Xue, R. Wang, X. He, J. Zhu, J. Li, and B. An, "DeepScalper: A risk-aware reinforcement learning framework to capture fleeting intraday trading opportunities," *Proc. 31st ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, pp. 1–10, 2022.
9. D. O. Lucca and E. Moench, "The pre-FOMC announcement drift," *The Journal of Finance*, vol. 70, no. 1, pp. 329–371, 2015. doi: 10.1111/jofi.12189.
10. P. 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. doi: 10.1111/j.1540-6261.2007.01232.x.
11. X. Wang and L. Liu, "Risk-sensitive deep reinforcement learning for portfolio optimization in petroleum futures," *Journal of Risk and Financial Management*, vol. 18, no. 6, pp. xxx–xxx, 2025. [Online]. Available: <https://www.mdpi.com/1911-8074/18/6/xxx>
12. R. R. Fu, Y. Y. Duan, L. Liu, and W. K. Tan, "Enhancing financial education with AI-driven learning and simulations," in *Proc. 2025 Int. Conf. on Artificial Intelligence and Education*, May 2025.
13. J. J. Wong and L. Liu, "Portfolio optimization through a multi-modal deep reinforcement learning framework," *Engineering: Open Access*, vol. 3, no. 4, 2025. [Online]. Available: <https://dx.doi.org/10.33140/EOA.03.04.03>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.