Leveraging Transformer-Based Sentiment Analysis for Financial Market Insights

Anany Sachan

September 18, 2025

In recent decades, sentiment analysis has emerged as one of the most widely studied applications of natural language processing (NLP). At its core, sentiment analysis refers to the automatic detection of opinions, attitudes, and emotions expressed in text. Foundational work by Pang, Lee, and Vaithyanathan established that classical machine learning methods such as Naïve Bayes, Maximum Entropy, and Support Vector Machines could be successfully applied to classify movie reviews as positive or negative [7]. This research established sentiment classification as a formal computational task and inspired a wave of subsequent studies across domains ranging from politics to product reviews.

As the field matured, researchers began to move beyond polarity detection toward the identification of richer affective states. Mohammad surveyed advances in detecting valence, discrete emotions such as joy, anger, or fear, and other affectual categories, while also highlighting issues of bias, fairness, and societal consequences [6]. Around the same time, deep learning models began to dominate sentiment analysis tasks. Surveys such as Zhang, Wang, and Liu documented how convolutional and recurrent neural networks dramatically improved performance compared to traditional approaches, especially when trained on large labeled datasets [10].

The most significant methodological breakthrough came with the introduction of transformer-based language models, particularly BERT and its domain-specific variants. Transformers rely on attention mechanisms that capture long-range dependencies in text, enabling much more accurate contextual representations than bag-of-words or recurrent architectures. Subsequent surveys and applied studies demonstrated that transformer models outperform previous baselines across a wide range of sentiment tasks [2, 8]. In the financial domain, fine-tuned transformer models such as FinBERT have been shown to significantly improve sentiment detection in technical texts like earnings reports and analyst briefings [3, 1].

The significance of these advances lies in the growing recognition that public sentiment is a powerful driver of financial markets. Traditional quantitative models, which rely primarily on historical prices and trading volume, fail to account for the behavioral and psychological forces that often dictate short-term volatility. Events such as the 2021 GameStop short squeeze—fueled largely by sentiment on Reddit's r/wallstreetbets—illustrate the real-world consequences of collective investor mood [4]. Incorporating sentiment signals into financial analysis has the potential to improve forecasting, risk management, and decision-making [5]. Recent work has begun to test this hypothesis, showing correlations between sentiment

derived from social media or news and subsequent asset price movements [9]. However, these methods remain far from standardized, and questions remain regarding reliability, data bias, and robustness across market conditions.

The goal of this project is to design and implement a finance-focused sentiment analysis system that builds on the trajectory of prior research while addressing gaps in applied usage. Specifically, I will develop a **Finance Sentiment Dashboard**: a software application that ingests financial text data (news headlines, social media posts, or discussion forum content) and outputs sentiment classifications and visualizations in near real-time. The system will leverage a transformer-based model fine-tuned for financial text, such as FinBERT, to provide ticker-specific sentiment scores and trends.

The proposed software will have three main capabilities. First, users will be able to enter a stock ticker symbol, at which point the backend will collect relevant recent text data from configured sources. Second, the system will preprocess the data and apply the sentiment model to classify texts as positive, negative, or neutral with respect to financial outlook. Third, the frontend will display results in a clear, interactive dashboard, including aggregate sentiment metrics, temporal trend charts, and confidence estimates. The pipeline flow is illustrated in Figure 1.

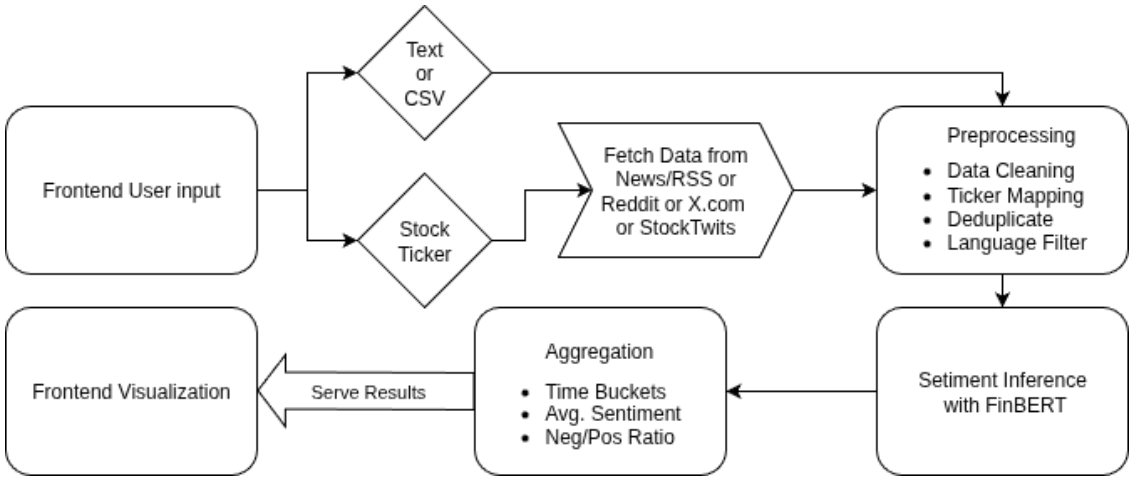

Figure 1: System architecture for the Finance Sentiment Dashboard.

By combining academic rigor with practical functionality, this project will contribute both to scholarly discussions and to applied finance practice. It demonstrates how the evolution of sentiment analysis — from early machine learning approaches [7], through emotional nuance [6], to modern transformer architectures [2, 3] — can be operationalized into a real-world tool. At the same time, it aims to provide users with an accessible interface for exploring how public mood influences financial markets, reinforcing the significance of sentiment as both a computational and economic phenomenon.

Appendix

A concise list of features / user stories in the order in which they will be built.

- Set up the foundational project environment for backend (Flask) and frontend (React), including code quality tools, base tests, and optional CI workflow.
- Configure environment variables and secrets for API keys and external service access.
- Define and enforce a common document schema for all ingested and uploaded texts across adapters and APIs.
- Implement a preprocessing pipeline for text cleaning, deduplication, and mapping to ticker symbols.
- Integrate the FinBERT model to perform financial sentiment inference on text data.
- Add a backend endpoint for analyzing pasted text and returning sentiment results.
- Add a backend endpoint for analyzing uploaded CSV files and returning sentiment results.
- Develop a news ingestion adapter to fetch recent financial headlines for given tickers.
- Develop a Reddit/StockTwits ingestion adapter to collect ticker-specific posts.
- Create an aggregation module to compute sentiment metrics (mean score, percent positive, volume, positive/negative ratio).
- Implement an explainability endpoint that returns token-level sentiment highlights for model predictions.
- Develop REST API endpoints for metrics, posts, explanations, and system health monitoring.
- Build the core React dashboard with ticker input, source toggles, KPI displays, sentiment chart, and top posts list.
- Add frontend flows for paste-text and file upload sentiment analysis.
- Implement user experience states such as loading, empty view, and error handling in the frontend.
- Containerize both backend and frontend using Docker and supply a Compose file for local deployment.
- (**Stretch Goal**) Implement a scheduler and caching mechanism to periodically refresh and store sentiment results in memory.
- (**Stretch Goal**) Add a persistent storage layer for texts, sentiment scores, and aggregated metrics.

- (**Stretch Goal**) Add a live stock price overlay to the sentiment time-series chart in the dashboard.
- (**Stretch Goal**) Develop a prototype predictive model that combines sentiment features with price baselines for short-term forecasting.

References

[1] Araci, D. Finbert: Financial sentiment analysis with pre-trained language models, 2019.

[2] Das, R., and Singh, T. D. Multimodal sentiment analysis: A survey of methods, trends, and challenges. *ACM Comput. Surv. 55*, 13s (July 2023).

[3] Deng, X., Bashlovkina, V., Han, F., Baumgartner, S., and Bendersky, M. Llms to the moon? reddit market sentiment analysis with large language models. In *Companion Proceedings of the ACM Web Conference 2023* (New York, NY, USA, 2023), WWW '23 Companion, Association for Computing Machinery, pp. 1014–1019.

[4] Desiderio, A., Aiello, L. M., Cimini, G., and Alessandretti, L. The dynamics of the reddit collective action leading to the gamestop short squeeze. *npj Complexity 2*, 1 (Feb. 2025).

[5] Jiang, T., and Zeng, Q. Financial sentiment analysis using finbert with application in predicting stock movement, 2025.

[6] Mohammad, S. M. Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In *Emotion Measurement (Second Edition)*, H. Meiselman, Ed. Elsevier, 2021.

[7] Pang, B., Lee, L., and Vaithyanathan, S. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10* (USA, 2002), EMNLP '02, Association for Computational Linguistics, pp. 79–86.

[8] Tang, J. Leveraging transformer neural networks for enhanced sentiment analysis on online platform comments. In *Proceedings of the 2023 4th International Conference on Machine Learning and Computer Application* (New York, NY, USA, 2024), ICMLCA '23, Association for Computing Machinery, pp. 256–260.

[9] Xing, F. Z., Cambria, E., and Welsch, R. E. Intelligent asset allocation via market sentiment views. *Comp. Intell. Mag. 13*, 4 (Nov. 2018), 25–34.

[10] Zhang, L., Wang, S., and Liu, B. Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery 8*, 4 (2018), e1253.