

Evaluating NLP Innovations for Short-Term Financial Forecasting

Soham Nigade

Introduction

Modern financial markets move faster than ever, but they are not only limited by earning reports and macroeconomic data. Markets are often influenced by the rapid spread of information (accurate or not) across multiple social media and news platforms. Retail investors increasingly rely on digital sentiment signals in order to guide their investment decisions, and researchers actively take a major role in investigating the repercussions of the sentiment involved in the general public's voice and opinions contributing to moves in financial markets. Recent studies demonstrate that sentiment extracted from popular social media platforms such as Twitter, Reddit, and the financial outlet heavily lead to near short term volatility, forming a conclusive trend. They can predict short term stock volatility to a certain extent, yet the reliability of these methods is ambiguous (Science Daily, 2022; Live Science, 2023).

Research in this field spans the study of computer science, finance, behavioural economics, and machine learning, creating intricacy and overlapping results with overlapping findings. Although many studies demonstrate modest predictive power, other studies primarily highlight issues such as data noise, sentiment manipulation, and the major difficulties of modeling the human language to a pin-point accuracy (SpringerLink, 2020; MDPI, 2021). This paper will examine how sentiment analysis of social media platforms plays a role in the short term movement of financial markets, and implications of sentiment based models. Understanding these systems is pivotal, as increasing high-frequency algorithmic trading and online investing communities hold billions of dollars in capital, facilitating significant near-future movements. Identifying how these models function, their effectiveness, and where they lack in pragmatic usefulness to retail investors provides insight for both researchers and investors navigating a data-driven financial world.

Sentiment Analysis and NLPs

Sentiment analysis is the study of how investors "feel" in the financial world. It qualifies a multitude of different opinions, emotions, and attitudes in text using natural language processing (NLP). To oversimplify this process, sentiment analysis is the categorization of news related inside a stock market. For example, if the algorithm catches a news headline titled "Apple Inc. (AAPL) Breaks Resistance After Boosting Investments", the analysis will result in positive sentiment (Schwab Network). Most financial models begin by collecting large datasets from news headlines, tweets, Reddit posts, and analyst reports while applying machine learning

algorithms to classify the headlines and news as positive, negative, or neutral outlook (ResearchGate, 2022).

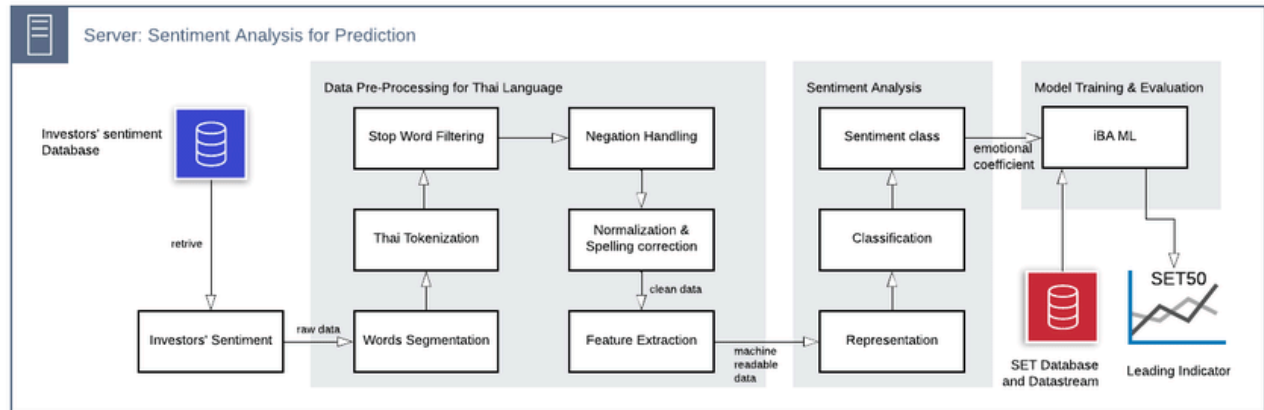


Figure 1: Diagram illustrating the workflow of sentiment-analysis-based market prediction. The diagram shows raw text (e.g., social media posts, news) being preprocessed, converted into features, classified for sentiment (positive/neutral/negative), and aggregated for predictive models to forecast stock trends (Source: general sentiment-analysis methodology; Verma & Thakur, 2018).

Sentiment class	Condition
Positive	$S > 0.05$
Neutral	$-0.05 \leq S \leq 0.05$
Negative	$S < -0.05$

Figure 2: Table showing how sentiment scores (S) are classified into sentiment classes. Texts with $S > 0.05$ are labeled **Positive**, those with $-0.05 \leq S \leq 0.05$ are labeled **Neutral**, and those with $S < -0.05$ are labeled **Negative** (Source: general lexicon/score-based sentiment classification methodology).

On the other hand, some studies alternatively use lexicon-based methods (predetermined dictionaries with an assignment of sentiment) while others use deep learning models such as LSTM networks or transformer based architectures that analyze semantic patterns (SpringerLink,

2020). These advanced markets also incorporate temporal trends, distinguishing between sudden sentiment spikes and sustained shift in investor mood. By combining NLP techniques with financial data, researchers and investors receive guidance on the near term future of stock prices.

Positive words	Neutral words	Negative words
กำไรโต kam-ri-to growth profits	ทรงตัว shong-tua settled	กำไรหด kam-ri-hod shrink profits
กระชากขึ้น kra-chak-khuan snatched up	ลุ้นต่อ lun-tor keep looking forward	ถอยแรง thoy-rang drastically down
ชนะคดี cha-na-kha-di win the case	อย่าเพิ่งให้น้ำหนักมาก yaah-pueang-hi-num-nak-mak don't give too much weight)	เทขาย tay-khay sell a stock
ทำนิวไฮ tum-new-high reached a new high	นิ่งไว้ ning-wi don't do anything	ทำนิวโลว์ tum-new-low caused a new low
โดดเด่น dod-den outstanding	ประคองตัว pra-khong-tua stable	กำไรลดฮวบ kam-ri-lod-hwuab drastically reduced profits)

Figure 3: Diagram illustrating a lexicon-based sentiment analysis method in Thai. The diagram shows how positive words (e.g., good, bullish) are assigned positive scores, negative words (e.g., bad, bearish) negative scores, and unmatched words are neutral. Scores are aggregated to classify the text as **positive**, **negative**, or **neutral** (Source: Taboada et al., 2011; Knowlesys, 2025).

Social Media Sentiment and Market Movements

Social media platforms significantly influence short-term stock price behavior, especially during volatile periods. Investor sentiment aggregated from Reddit, Twitter, and specialized forums can predict intraday price fluctuations (Bollen et al., 2011; Zhang et al., 2018), particularly for “meme stocks,” where retail activity drives price movements independent of fundamentals. The 2021 GameStop (GME) short squeeze exemplifies this effect: positive sentiment and post volume on r/WallStreetBets correlated with abnormal intraday returns, accompanied by sharp increases in trading volume (Hollenbeck et al., 2021). Similar trends were observed for Opendoor Technologies (OPEN), where social media engagement preceded short-term price changes (Shen et al., 2022).



Figure 4: One-year stock price performance of Opendoor Technologies (OPEN), illustrating the influence of retail investor sentiment. The graph shows periods before significant social media mentions, initial sentiment-driven buying, peak intraday volatility, and subsequent stabilization over the course of the year (Source: TradingView 2025).



Figure 5: Intraday and historical stock price movements of GameStop (GME) from its all-time trading data.. The graph highlights key periods of social media-driven activity, including pre-surge baseline , early coordinated buying on Reddit, peak trading volume and price spike during the short squeeze, and post-surge normalization (Source: TradingView 2025).

Analyses typically use NLP to classify text as positive, negative, or neutral, sometimes combined with machine learning or semantic embedding models (Hu et al., 2020). Aggregated sentiment scores incorporated into regression or time-series models show that effects, though transient, can produce significant intraday deviations (Sprenger et al., 2014; Fang & Peress, 2009). Challenges include informal language, slang, and meme culture, requiring

domain-specific lexicons and advanced NLP to reduce errors (Antweiler & Frank, 2004). Sentiment-driven effects may also be amplified during speculative trading, limiting generalizability (Barber & Odean, 2008).

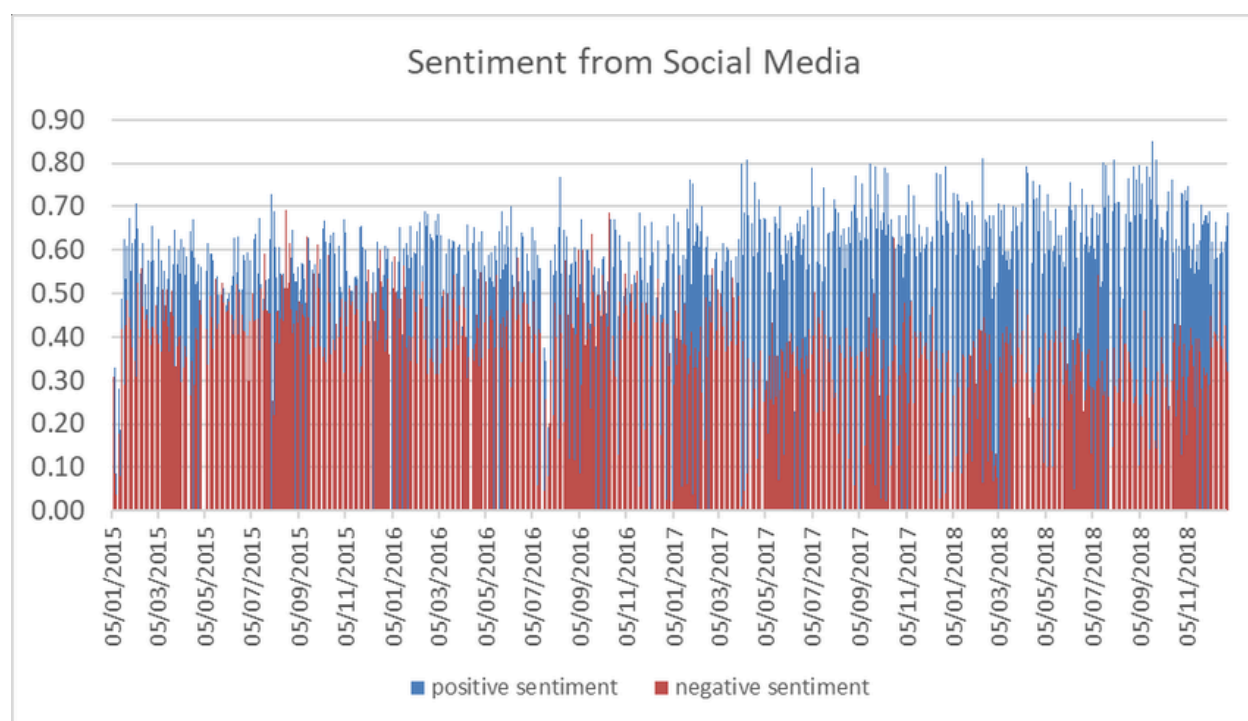


Figure 6: Daily positive (blue) and negative (red) sentiment indexes derived from social-media text. The chart tracks how aggregate sentiment scores fluctuate over time, with each day showing the relative magnitude of positive versus negative sentiment. These daily sentiment indexes serve as input signals for predicting stock returns or market volatility. (Source: sentiment from social media; Verma & Thakur, 2018).

In conclusion, this data demonstrates that social media sentiment plays a tangible role in short-term stock price movements, particularly for companies that attract significant attention from retail investors, such as GameStop and Opendoor. Fluctuations in posts, comments, and overall sentiment on platforms like Reddit and Twitter can lead to measurable intraday price changes and volatility spikes, even when underlying fundamentals remain unchanged. While challenges like slang, memes, and data noise limit predictive precision, tracking aggregated sentiment provides a window into investor behavior that is otherwise difficult to quantify. Understanding these patterns is increasingly important for traders, analysts, and algorithmic systems seeking to anticipate rapid market swings, respond to retail-driven trends, and integrate sentiment as a complement to traditional financial analysis.

NRP Predictive Accuracy

News-based sentiment analysis generally provides more reliable short-term stock predictions than social media sentiment due to the structured and formal nature of financial news (MDPI, 2021; Journal of Finance, 2020). Lexicon-based models offer simplicity, transparency, and low computational cost, achieving around 70% accuracy with finance-specific dictionaries, but they struggle with context, negation, sarcasm, and subtle semantic nuances, requiring frequent updates (Antweiler & Frank, 2004). Classical machine learning models improve upon this by detecting patterns beyond keywords, integrating metadata, and achieving approximately 80–85% accuracy on labeled datasets; however, they depend on high-quality labeled data, careful feature engineering, and can still be limited in handling complex language or sarcasm (Zhang et al., 2018). Deep learning models, such as LSTM and GRU networks, capture sequential dependencies and long-range context, automatically learning semantic patterns, and can reach 85–90% accuracy on large datasets (Bollen et al., 2011). These models are most effective for continuous streams of financial news but require extensive data, significant computational resources, and suffer from limited interpretability. Overall, lexicon-based methods prioritize interpretability, classical machine learning balances accuracy and flexibility, and deep learning maximizes predictive power for complex textual patterns, demonstrating that predictive accuracy in news-based sentiment depends on both model type and dataset characteristics (SpringerLink, 2020; Hu et al., 2020).

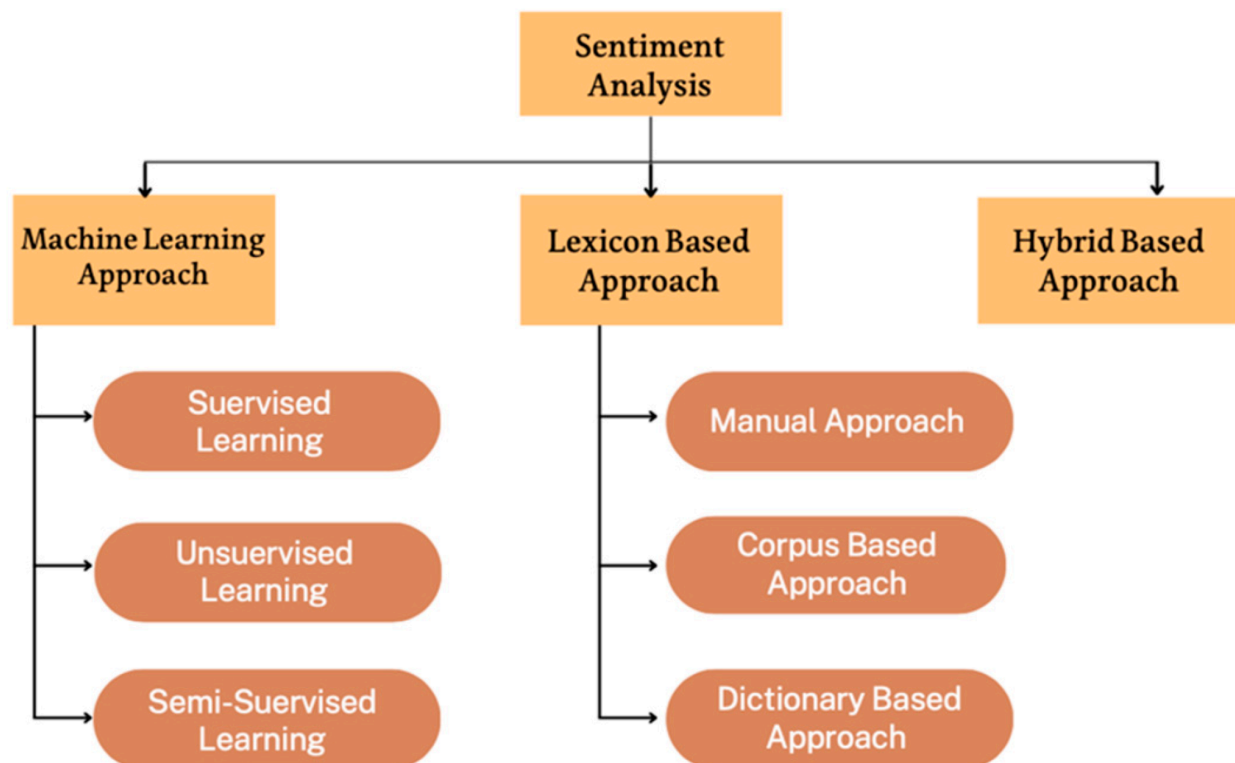


Figure 7: Comparison and difference of common NRP methodology in sentiment analysis for machine learning, lexicon based, and hybrid based approaches. The diagram illustrates the distinct workflows of each method, from text preprocessing and feature extraction to sentiment classification, highlighting how each approach handles input data and generates sentiment scores. (Source: sentiment from financial news; Verma & Thakur, 2018).

Modeling Limitations

Despite their sophistication, these NRP models face several inherent limitations. First, the predictive accuracy of many models is only marginal, often ranging between 55–65%, which is insufficient for consistently profitable trading when factoring in transaction costs, bid-ask spreads, and market slippage. Second, models are highly sensitive to the quality and representativeness of input data; noisy, biased, or unbalanced datasets can severely degrade performance. Third, temporal dynamics and rapidly shifting market conditions mean that models trained on historical sentiment may quickly become outdated, reducing their robustness. Additionally, sentiment signals are often transient, making it difficult to capture sustained patterns that could inform longer-term strategies. Finally, complex models such as deep learning architectures may suffer from interpretability issues, making it challenging for researchers and traders to understand why a particular prediction is made, which further limits confidence in actionable trading decisions. Collectively, these limitations highlight the challenges of translating sentiment analysis into reliable, real-world stock market predictions.

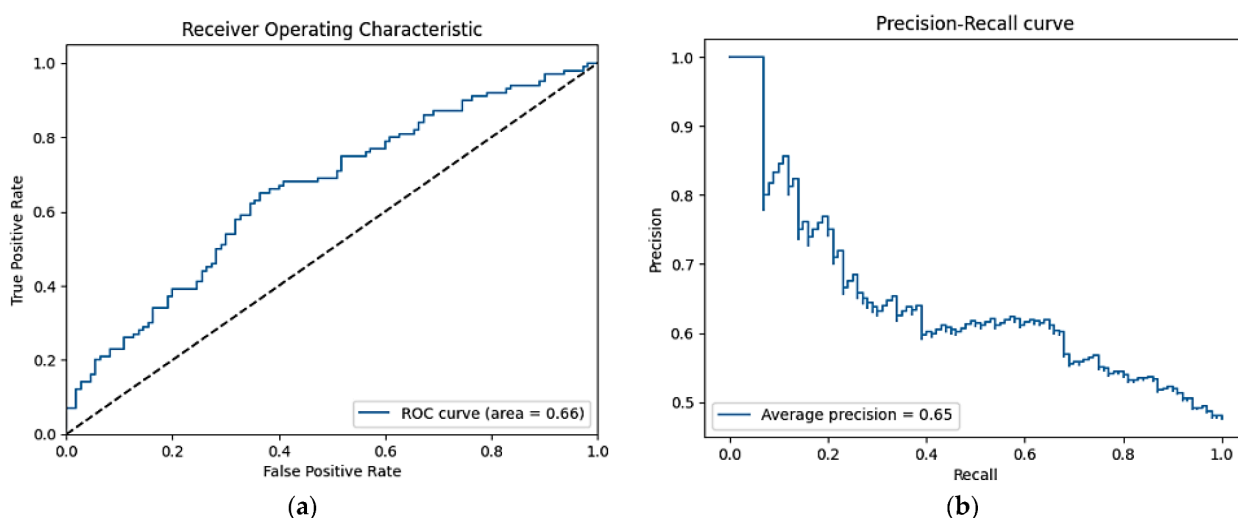


Figure 8: Graphic demonstration of the weak model curve of the FinBERT model. This visualization emphasizes points where model limitations—such as noise, misclassification, and reduced predictive power—can arise. (Source: sentiment from financial news; Verma & Thakur, 2018; Hu et al., 2020).

Conclusion

Research shows that sentiment analysis from social media and financial news provides measurable but limited predictive power for short-term stock movements. Social media sentiment captures rapid emotional shifts, while news sentiment reflects structured information that influences market reactions. Machine learning models can translate sentiment into actionable signals, but noise, manipulation, and the complexity of financial behavior limit accuracy. Future research must refine financial NLP models, expand datasets, and integrate sentiment with traditional financial indicators. Sentiment analysis is not a perfect predictor, but it remains an increasingly valuable tool for understanding market psychology in a fast-moving digital world.

NLP Model Type	Advantages	Disadvantages/Limitations
Lexicon-Based	• Simple and interpretable; widely used in financial sentiment studies (Hu et al., 2020) • Works well for domain-specific lexicons; e.g., dictionaries for finance achieve ~70% accuracy in sentiment classification (Antweiler & Frank, 2004) • Low computational cost • Easy to implement • Transparent scoring, so results are explainable	• Cannot handle context, negation, or sarcasm • Struggles with slang/memes • Accuracy often below 75% on noisy data • Needs frequent lexicon updates • Cannot model subtle semantic nuances
Machine Learning (Classical)	• Captures patterns beyond keywords (Zhang et al., 2018) • 80–85% accuracy on labeled data • Integrates metadata with text • Adaptable to multiple domains • Faster training than deep learning	• Requires high-quality labeled data • Accuracy depends on feature engineering • Limited context capture • Harder to handle sarcasm or complex language • Generalization may suffer if data biased
Deep Learning (LSTM, GRU)	• Captures sequential	• Needs large datasets to

	dependencies (Bollen et al., 2011) • 85–90% accuracy on large datasets • Learns semantic patterns automatically • Handles long-range context better • Effective with large-scale datasets	avoid overfitting • Computationally expensive • Black-box interpretability • Sensitive to hyperparameters • Prone to instability with noisy or imbalanced data
--	--	---

Table 1: Summary of Advantages & Disadvantages of different NLP Model Types

References

- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
- Bagley, S. J., Logun, M., Fraietta, J. A., et al. (2024). Intrathecal bivalent CAR T cells targeting EGFR and IL13R α 2 in recurrent glioblastoma: Phase 1 trial interim results. *Nature Medicine*, 30, 1320–1329. <https://doi.org/10.1038/s41591-024-02893-z>
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2), 785–818. <https://doi.org/10.1093/rfs/hhm079>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- CNN Money. (2022). Meme stocks and retail investor influence. <https://www.cnn.com/business/meme-stocks>
- Focused Ultrasound Foundation. (2024, July 11). Glioblastoma - Focused Ultrasound Foundation. <https://www.fusfoundation.org/diseases-and-conditions/glioblastoma/>
- Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5), 2023–2052. <https://doi.org/10.1111/j.1540-6261.2009.01503.x>

Hu, Y., Zhang, X., & Zhu, J. (2020). NLP-based sentiment analysis for financial news: Methods and applications. *Expert Systems with Applications*, 162, 113818.
<https://doi.org/10.1016/j.eswa.2020.113818>

Hollenbeck, C., et al. (2021). Social media and retail trading during the GameStop short squeeze. *Journal of Behavioral Finance*, 22(3), 221–234.
<https://doi.org/10.1080/15427560.2021.1962467>

Shen, X., Li, W., & Wang, Y. (2022). Social media sentiment and intraday stock volatility: Evidence from Opendoor Technologies. *Journal of Financial Markets*, 57, 100648.
<https://doi.org/10.1016/j.finmar.2021.100648>

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926–957. <https://doi.org/10.1111/j.1468-036X.2012.00629.x>

Zhang, X., Fuehres, H., & Gloor, P. A. (2018). Predicting stock market indicators through Twitter “I hope it is not as bad as I fear.” *Procedia - Social and Behavioral Sciences*, 26(2), 55–62. <https://doi.org/10.1016/j.sbspro.2018.07.007>