IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

How Fast Does the Indian Stock Market Price News? A Finbert Study of Five Years of Headlines and Exchange Filings

Dr. R. Y. Totare[1], Ashish Kharde[2], Chinmay Nakwa[3], Swaroop Patil[4]
Professor, Department of Information Technology[1]
Students, Department of Information Technology[2-4]
AISSMS Institute of Information Technology, Pune, India
[1] reshma.totare@aissmsioit.org, [2] ashnkharde@gmail.com,
[3] 29chinmaynakwa@gmail.com, [4] swarooppatil850@gmail.com

Abstract : *While FinBERT is extensively applied for sentiment analysis of finance news and for predicting stock pricing, few studies verify whether such news is publicly available in advance of the price a comparison is made with. We apply FinBERT on 5 years of market news data, comprising of 265860 headlines from Google News news from India and 146257 National Stock Exchange (NSE) corporate filings for 254 stocks, 2021-09 to 2026-09 and map each piece of news to the first trading session whose closing price date is after the release. Headlines and filings are positively correlated with prices on the day when the news is released, with the daily rank correlation between FinBERT sentences and the market-relative return of the stock being 0.026 (t = 4.2) in 2022–2024 for headlines and 0.020 (t = 5.0) for filings. The news gets absorbed by the market in the session in which it comes. Closed market filings are valued at the next opening price, and starting from the following session on, there is no predictive power in sentiment, on a held-out 2025-2026 period. If news is joined to prices by calendar date, however, instead of a next-day signal (t = 3.4) that a placebo test would not detect, one will be detected. 89.5% of the filings classified by FinBERT as negative are also routine notices of lost share certificate. On Indian equities FinBERT is a reliable measure of how the market reacted to news, not a forecast of what it will do next, and correct timestamps matter more than the choice of sentiment model.*

Keywords: BERT, FinBERT, financial news, Indian stock market, look-ahead bias, NSE, sentiment analysis, transformers.

I. INTRODUCTION

FinBERT [1] brings sentiment analysis—classifying the opinion in a text as positive, negative or neutral [2]—to financial language. It is based on the Transformer architecture [3] and the BERT language model [4] and is now a common method for converting news into numbers: News scores with FinBERT have been employed in management of US equity portfolios [5]; to predict the direction of US stocks [6]; as a benchmark for Chinese financial news [7]; and to investigate the price of cryptocurrencies [8].

A sentiment score can only assist a forecast when the news is made available prior to the price forecasted. If the public news is absorbed by prices during the time of its publication as the efficient-market view suggests [9], then FinBERT is able to explain price changes but not predict them. This requires the time that each piece was published, compared to the time for the market close, and knowing this time is very difficult. While free news feeds may release only a date, exchange news releases provide exact times but are predominantly delivered after trading hours; and news date

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, November 2026 Impact Factor: 8.2

attachment to the same day's closing price is a subtle way of using late-day news releases. How release timings were matched to prices is not mentioned in any of the above-mentioned FinBERT news studies.

This paper uses FinBERT to five years of market news from India, and correlates each piece of news with the first session of the day where the price of the market closed, and it measures the speed of the market closing the price of what FinBERT reads. The work is the news-sentiment module of our last year project called ARTHA. Its contributions are:

- A dataset of 412,117 timestamp-aligned news items for 254 NSE stocks, built from free sources and scored with a pinned FinBERT model;
- Evidence that FinBERT sentiment matches same-day price moves and that the market prices the news within one session, confirmed on a held-out period;
- A measurement of the false signal created by joining news to prices by calendar date; and
- An analysis of how FinBERT labels Indian exchange filings.

Section II reviews FinBERT and its use on news, Section III describes the data and method, Section IV presents the results, Section V discusses them, and Section VI concludes.

II. BACKGROUND AND RELATED WORK

A. Transformers, BERT and FinBERT

Early sentiment analysis relied on word lists and classical classifiers [2]. The Transformer replaced recurrent networks with self-attention, letting a model relate every word in a text to every other [3]. BERT pre-trains a bidirectional Transformer encoder on large general corpora and is then fine-tuned for a specific task [4]. FinBERT continues BERT's pre-training on financial news and fine-tunes it on the Financial PhraseBank, 4,840 sentences from English financial news labelled positive, negative or neutral by annotators with finance backgrounds [1], [10]. For each input it returns the probabilities of the three classes.

B. FinBERT for Financial News in Other Markets

Mantshimuli averages FinBERT scores of news on S&P 500 companies over 30 days and adds them to the state of a deep-reinforcement-learning portfolio rebalanced monthly, which beats an equal-weight benchmark [5]. Shrestha et al. merge each day's headlines for ten US tickers, score them with FinBERT and combine the scores with technical indicators, reporting direction accuracies of up to 0.90 [6]. Chen et al. use a Chinese FinBERT as a baseline against prompted large language models for classifying Chinese financial news [7]. Liu et al. review BERT, FinBERT and newer language models for financial sentiment and, in a case study of cryptocurrency news from November 2021 to November 2023, find a modest correlation between news sentiment and Bitcoin prices [8]. Table I compares these studies with ours.

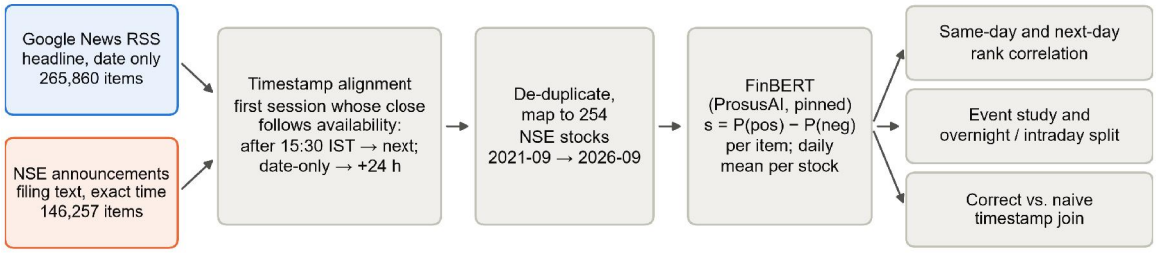


Fig. 1. Pipeline from news collection to FinBERT scores and the tests applied to them. Item counts are after de-duplication, September 2021 to September 2026.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

TABLE I
FinBERT news studies and this work

Study	Market, text	Use of FinBERT	Evaluation	Release time
[5]	US (S&P 500), news	30-day mean in RL state	Portfolio return	Monthly dates
[6]	US (10 tickers), headlines	Daily score + indicators	Direction accuracy	Live use only
[7]	China, news	Baseline classifier	Labelled F1	No prices
[8]	Crypto, news	Reviewed; case study	Correlation	Not stated
This work	India (254 NSE), headlines + filings	Item score, daily mean	Same- vs. next-session	First close after release

"Release time": how the paper matches each text's publication time to the price data in its backtest. RL = reinforcement learning.

C. Research Gap

These studies use FinBERT either as a classifier scored against human labels [7] or as an input to a forecasting or trading model [5], [6]. In the second use the timing of the news decides the result, yet none states how the release time of each item was matched to the price data in its backtest; Shrestha et al. specify collecting headlines before the open only for live use [6]. We found no study of FinBERT on Indian news with release-time alignment, and none on Indian exchange filings.

III. DATA AND METHODOLOGY

Fig. 1 summarises the pipeline: two news sources, a timestamp-alignment step, FinBERT scoring, and three groups of tests.

A. Stocks and Prices

The sample is the 254 NSE stocks that are current Nifty 500 members, have traded since January 2014, and appear in our news archive. Daily open, high, low, close and volume data come from Yahoo Finance, adjusted for splits and dividends, with the Nifty 500 index as the market. A stock's market-relative return in session t is its close-to-close return minus the index return; 16 one-day moves above 40% (bad ticks and demergers) are removed. Each session's return is also split into an overnight part, from the previous close to the open, and an intraday part, from the open to the close; 11,409 stock-days (1.2%) whose open equals the previous close or whose volume is zero are left out of this split.

B. News Sources

Google News RSS was searched one company name at a time in fixed seven-day windows, and only items whose headline names the company were kept; each item has a headline and a publication date. The NSE corporate-announcements service gives each filing's category, text and dissemination time to the second. The same headline for the same stock in the same session counts once. Table II describes the data. Google coverage grows over time—20,790 items in 2022 against 114,460 in January–September 2026—while NSE filings stay at about 27,000–34,000 a year, so all tests compare stocks with each other on the same day rather than over time.

TABLE II: News data for 254 NSE stocks

	Google News	NSE filings
Items collected	276,156	163,886
Items after de-duplication	265,860	146,257
Stocks covered	254	251
Text scored	Headline	Filing text
Time stamp	Date only (95.1%)	Exact (100%)
Released while market closed	—	72.1%

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

	Google News	NSE filings
FinBERT positive / neutral / negative (%)	30.8 / 55.1 / 14.2	4.0 / 83.6 / 12.4
Mean |s|	0.418	0.172

1 Sep. 2021–17 Sep. 2026. Labels are FinBERT's most probable class; s is defined in (2).

C. Timestamp Alignment

Let aj be the time at which item j became public and Ct the 15:30 IST close of session t. The item is assigned to its news session

$$\tau_j = min\{ t : C_t \geq a_j \} \qquad (1)$$

so a filing released at 17:00 belongs to the next session. Of the Google items, 95.1% carry a 07:00 or 08:00 GMT stamp, which is midnight in US-Pacific time: a date without a time. Read literally it becomes 12:30 IST and would place every such item before that day's close. We therefore set aj to the stamp plus 24 hours, assuming only that the item was public by the end of its date. For comparison, the naive alignment places each item on the first session on or after its calendar date, which is what a simple join of news dates to price dates does.

D. FinBERT Scoring

We use ProsusAI/finbert from the Hugging Face hub, pinned to one model revision so that the scores can be reproduced. For Google items FinBERT reads the headline, as the feed carries no article text. For NSE filings the headline field holds only a category such as "Updates", so FinBERT reads the filing text, after removing the template opening "<Company> has informed the Exchange regarding", found in 92.4% of filings, so that it neither dilutes the text nor puts the company name in front of the model. Texts are short (median 75 characters for headlines, 57 for filings); they are tokenised with truncation at 512 tokens, grouped by length into batches of 64 and run in full 32-bit precision, and a softmax gives the three class probabilities. Each distinct text is scored once and cached by its hash: the 654,657 collected items contain 459,879 distinct texts. Scoring ran at about 23 texts per second on a laptop CPU, so it was moved to a cloud GPU; when the GPU scores are imported, 200 random texts are re-scored locally and the import is rejected if any probability differs by more than 10−3.

We chose FinBERT over a recent large language model for a reason specific to historical studies: a model trained on 2023–2026 text may already know how the stories in our sample turned out, which is a form of look-ahead. FinBERT's fine-tuning data, the Financial PhraseBank [10], predates our sample. The score of item j and the daily tone of stock i in session t are

$$s_j = P(pos) - P(neg), \; S_{i,t} = \frac{1}{n_{i,t}} \Sigma_j s_j \quad (2)$$

where the sum runs over the ni,t items of stock i assigned to session t. Tone is defined only on stock-days with news.

E. Evaluation

The main measure is the daily rank information coefficient (IC): the Spearman correlation, across the stocks with news on day t, between tone and the return h sessions later,

$$IC_t^{(h)} = \rho_S(S_{\cdot,t}, r_{\cdot,t+h}) \qquad (3)$$

averaged over days, with a t-statistic computed on every h-th day so that returns do not overlap. h = 0 is the news session itself, which cannot be traded on; h = 1 is the first return a trader could earn after learning the news at the close. Because the IC ranks stocks within a day, it is unaffected by the market's common move. An event study follows the market-relative return of positive (S > 0.2) minus negative (S < −0.2) news stock-days from five sessions before to five after the news session. A placebo shuffles tone across stocks within each day, 20 times. The development period is 2022–2024 (736 sessions); January 2025 to September 2026 (427 sessions) is held out and every IC is repeated on it. Rebuilding the daily features from items known up to 30 June 2023 reproduces the full build exactly, so no feature uses later news.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

IV. RESULTS

A. FinBERT Agrees With Same-Day Prices

Table III shows that FinBERT tone is ranked in line with the market-relative return of the news session for both sources: IC 0.026 (t = 4.2) for headlines and 0.020 (t = 5.0) for filings in 2022–2024, against a placebo bar of |t| = 2.2 that holds 95% of shuffled draws. FinBERT's scores therefore carry market-relevant meaning. Headlines also correlate strongly with the five sessions before the news session (IC 0.087, t = 14.4): for date-only items this window contains the publication day, and many headlines report a move that has already happened. In the held-out period the headline result is unchanged (0.027, t = 4.9); the filing result is weaker (0.012, t = 1.9), although filings released outside trading hours are still priced at the open (Section IV-B).

TABLE III
Rank IC of FinBERT tone with relative returns

Source, period	Prior 5	News session	Next 1	Next 2
Google, 2022–24	8.69 (14.4)	2.56 (4.2)	0.11 (0.2)	0.61 (0.6)
Google, 2025–26	14.29 (25.7)	2.69 (4.9)	−0.01 (0.0)	0.45 (1.0)
NSE, 2022–24	1.47 (3.5)	2.00 (5.0)	−0.10 (−0.2)	−0.09 (−1.6)
NSE, 2025–26	0.83 (1.3)	1.18 (1.9)	−0.37 (−0.6)	−0.87 (−1.0)

Mean daily Spearman IC ×100 (t-statistic), stock-days with news only. "Prior 5" = sessions −5 to −1; "Next 1" is the first return that can be traded after the news.

B. The Market Prices News Within One Session

From the first tradable close onward the IC is indistinguishable from zero for both sources and both periods (Table III, |t| ≤ 1.6). Fig. 2 shows the same in returns. The gap between positive- and negative-news stocks is 84 basis points (bp) on the publication day of date-only headlines and 24 bp in the news session, then 8 bp and 7 bp in the next two sessions (t = 2.0 and 1.9) and nothing after that. For filings it is 28 bp in the session before and 32 bp in the news session, then 1.5 bp (t = 0.4). A round trip in Indian delivery trades costs roughly 40 bp in taxes, fees and slippage, so the small remaining drift in headlines cannot be traded.

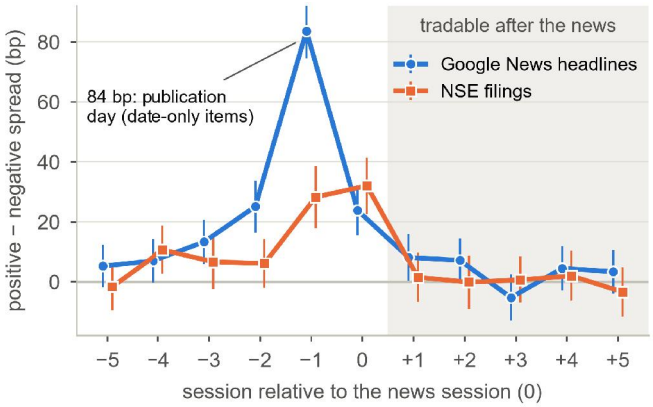


Fig. 2. Positive-minus-negative market-relative return around the news session (session 0), 2022–2024, with 95% intervals. Session 0 cannot be traded; the shaded sessions can. Google items carry only a date, so their publication day is session −1.

Fig. 3 shows when the reaction happens for NSE filings, whose times are exact. The 72.1% released while the market is closed are priced between the previous close and the next open (overnight IC 0.057, t = 11.2; held-out t = 6.5); the intraday return that follows has a slightly negative IC (t = −1.2) and the next day shows nothing (t = −0.7). Filings released during trading hours are priced within the same session (intraday t = 3.0). Headlines behave the same way:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

overnight IC 0.078 (t = 12.6), intraday −0.010 (t = −1.6). By the time a trader at the close could act, the news is already in the price.

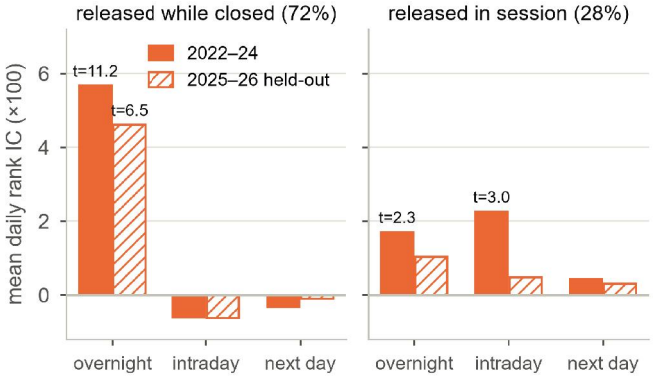


Fig. 3. NSE filings split by release time: rank IC of FinBERT tone with the overnight (previous close to open) and intraday (open to close) parts of the news-session return, and with the next day's return. Labels give t-statistics where |t| ≥ 2.

C. A Naive Date Join Creates a False Signal

Table IV and Fig. 4 repeat the analysis with the naive alignment. Placing items on their calendar date raises the news-session IC of headlines fourfold (0.108, t = 18.8) and, more importantly, creates a next-session IC of 0.020 (t = 3.4) in 2022–2024 and 0.018 (t = 3.4) in 2025–2026, positive in every year; filings show the same in the development period (t = 3.0). The cause is simple: an item published after the close is dated to that session but priced in the next one, so FinBERT appears to predict the next session using news a trader did not yet have.

TABLE IV
Correct vs. naive alignment

Source, period	Session: correct	naive	Next: correct	naive
Google, 2022–24	2.56 (4.2)	10.82 (18.8)	0.11 (0.2)	2.03 (3.4)
Google, 2025–26	2.69 (4.9)	16.17 (27.8)	−0.01 (0.0)	1.81 (3.4)
NSE, 2022–24	2.00 (5.0)	2.12 (5.1)	−0.10 (−0.2)	1.22 (3.0)
NSE, 2025–26	1.18 (1.9)	1.30 (2.1)	−0.37 (−0.6)	0.46 (0.7)

Mean daily Spearman IC ×100 (t-statistic) with the news-session and next-session relative return. Naive = item placed on the first session on or after its calendar date.

This error is not detected by a placebo test. The naive signal will quickly pass over the bar, as it is a real correlation in the data, not a result of chance; shuffled tone will not pass it, in 95% of draws |t| will be below 1.8. It is only by alignment that the two are separated.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

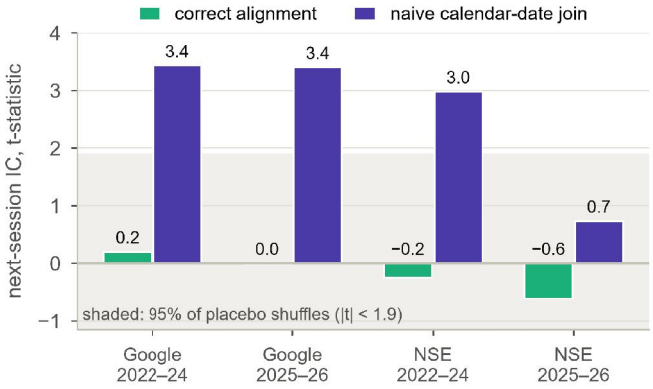


Fig. 4. Next-session IC t-statistic under correct and naive alignment. The naive join yields t ≈ 3 in three of four cases; the shaded band holds 95% of placebo shuffles.

D. What FinBERT Reads in Exchange Filings

Table V shows the distribution of FinBERT's labels by the NSE filing category. The vast majority of categories are neutral most of the time, such as financial-result updates, which state that results were filed but do not actually report the results. The 41% of the time, press releases are positive. FinBERT labels 12.4% of filings negative, but 89.5% of those are notices that a shareholder has lost a share certificate—an administrative event that FinBERT reads as a financial loss. Removing these notices leaves the news-session IC essentially unchanged (0.022, t = 5.0) and the next-session IC still insignificant (t = 1.2): they add noise rather than bias, but they make a count of "negative filings" misleading.

TABLE V
FinBERT labels by NSE filing category

Category	Filings	Pos. %	Neu. %	Neg. %
Analyst / investor meet updates	20,236	0.0	99.8	0.2
Loss of share certificate(s)	19,092	0.0	14.9	85.1
Updates	16,014	6.3	91.7	2.1
Press release	9,759	40.9	58.7	0.4
Copy of newspaper publication	8,564	0.0	99.6	0.4
Outcome of board meeting	4,699	0.4	99.6	0.0
Financial result updates	3,003	0.0	100.0	0.0
All NSE filings	146,257	4.0	83.6	12.4

Largest categories after de-duplication; labels are FinBERT's most probable class. The two spellings of the certificate category are combined.

V. DISCUSSION

A. What the Results Mean

The way FinBERT works on Indian news is that it's used for measurement - the tone of the news is used to rank stocks in the same way as the market reacted to the same news, both headlines and filings. It cannot be used as a forecast as NSE prices will reflect the news within one session, news released after the close of the previous session will be reflected on the next opening session of the trading day. This is the prediction of the efficient-market view for public news [9].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

The results are not in conflict with the findings of the FinBERT studies in other markets [5, 6], they reveal potential areas of opportunities. A model to learn from the reaction could be constructed by correlating a news date with the date of the price, so that they could be used to explain that day's price change. On this data, this would yield a t-statistic of 3.4. The studies that have been conducted with FinBERT as a predictive tool should therefore make sure that release times are captured, that each item is labeled at the first time that they were known, and that their results are split between those achieved in the same session and those achieved in the next session, and that they do not use a placebo test to check for look-ahead.

B. Limitations

Our prices are daily, as such "priced at the open" indicates the price range from the previous close to the opening auction, intraday data will reflect how many minutes the reaction will take. Google items only have a date; some news items during the day may be a session late due to the +24-hour rule, which is not conducive to finding predictability. Google news isn't an archive, so some news is lost and coverage is added over time. This is because, FinBERT was trained on English news about the Finnish companies and not on Indian texts. The stocks are existing members of the Nifty 500, which is a survivorship bias that has a greater impact on return levels than it does on rankings during the day. Lastly, many correlations are reported; the held-out period and the placebo limit remove the chance of a false finding, but do not remove it.

VI. CONCLUSION AND FUTURE WORK

We find that the sentiment associated with price moves on news is a good representation of the Indian price data, and it has no predictive ability—after-hours filings during the next session after the data appear, and free news during the next session after the data appear, are what the market absorbs, in both the development and held-out periods. The biggest danger in these studies is the clock, rather than the sentiment model: a naive date join generates a plausible false signal that a placebo test detects. Future studies will refine the model by tuning FinBERT on labelled headlines from Indian newspapers, bypass the administrative filing categories and read the numbers included in results filing and compare the reaction speed with intraday prices.

REFERENCES

[1] D. Araci, “FinBERT: Financial sentiment analysis with pre-trained language models,” 2019, arXiv:1908.10063.
[2] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.
[3] A. Vaswani et al., “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008.
[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics (NAACL-HLT), 2019, pp. 4171–4186.
[5] L. A. Mantshimuli, “Sentiment-aware portfolio optimization: CVaR-based diversification with deep reinforcement learning,” IEEE Access, vol. 13, pp. 182690–182704, 2025, doi: 10.1109/ACCESS.2025.3624652.
[6] S. Shrestha et al., “A hybrid ML algorithm trading model combining sentiment, technical indicator and risk management,” Procedia Comput. Sci., vol. 270, pp. 2858–2867, 2025, doi: 10.1016/j.procs.2025.09.408.
[7] W. Chen, W. Liu, J. Zheng, et al., “Leveraging large language model as news sentiment predictor in stock markets: A knowledge-enhanced strategy,” Discover Comput., vol. 28, Art. no. 74, 2025, doi: 10.1007/s10791-025-09573-7.
[8] C. Liu, A. Arulappan, R. Naha, A. Mahanti, J. Kamruzzaman, and I.-H. Ra, “Large language models and sentiment analysis in financial markets: A review, datasets, and case study,” IEEE Access, vol. 12, pp. 134041–134061, 2024, doi: 10.1109/ACCESS.2024.3445413.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, November 2026
Impact Factor: 8.2

[9] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," J. Finance, vol. 25, no. 2, pp. 383–417, 1970.
[10] 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.