Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Predicting Nifty 50 Opening Direction with Sentiment Analysis and Logistic Regression

Mr. Sahal Patel[1], Dr. Neha Vora[2]

[1]MSC Student, Usha Pravin Gandhi College Of Arts, Science And Commerce
[2]Assistant Professor, Usha Pravin Gandhi College Of Arts, Science And Commerce

Abstract

This study explores a hybrid approach to forecasting the opening direction of the Nifty 50 index by combining financial sentiment analysis and binary logistic regression. Recognizing the growing influence of investor sentiment on market movements, transformer-based language models such as FinBERT and RoBERTa are employed to extract sentiment scores from pre-market financial news and social media sources. These sentiment indicators are then integrated with historical market data to train a logistic regression model that predicts whether the index will open positively or negatively. The model is evaluated on accuracy and precision using recent Nifty 50 data. Results suggest that incorporating sentiment significantly improves predictive performance compared to models relying only on historical price patterns. This research highlights the potential of combining natural language processing with statistical models to improve short-term market forecasting in India.

Keywords: Nifty 50, Stock Market Forecasting, Sentiment Analysis, Logistic Regression, FinBERT, RoBERTa, Natural Language Processing (NLP), Financial Analytics, Investor Sentiment, Machine Learning in Finance

1. Introduction

1.1 Background

The Indian stock market, represented prominently by the Nifty 50 index, has become one of the most dynamic and closely observed markets globally. Predicting the short-term movement of the index is a challenging task due to its sensitivity to global events, investor behavior, and economic indicators. Traditional time-series forecasting methods often fail to account [8], [8], [10], [10] for the qualitative aspects of market sentiment. With the rise of Natural Language Processing [9] (NLP), particularly Large Language Models (LLMs), sentiment analysis from news and social media has emerged as a valuable tool for enhancing market prediction. This integration provides a holistic view [8] by combining numerical and textual data.

1.2 Objectives

The primary objectives of this research are as follows:

1. To design and implement a predictive framework for forecasting the opening direction of the Nifty 50 index using binary logistic regression.

2. To leverage advanced Large Language Models (LLMs), specifically FinBERT and RoBERTa, for performing sentiment analysis on financial news articles, expert commentary, and social media posts.

3. To combine sentiment-derived variables with quantitative historical data such as past index values, volatility measures, and trading volumes, thereby creating a hybrid feature set.

4. To evaluate the effectiveness of the proposed hybrid model against conventional statistical models (e.g., ARIMA) and machine learning models (e.g., LSTM).

5. To investigate the contribution of sentiment in enhancing short-term predictive accuracy of Indian equity indices compared to purely numerical approaches.

6. To establish a replicable methodology that can be extended beyond the Nifty 50 to other indices, asset classes, or emerging market contexts.

7. To provide actionable insights for traders, investors, and analysts by identifying patterns in sentiment that strongly influence index behavior.

8. To contribute academically to the intersection of econometrics, natural language processing, and financial analytics, thereby filling existing research gaps.

1.3 Purpose, Scope, and Applicability

1.3.1 Purpose

The purpose of this research is to bridge the gap between conventional quantitative models and qualitative sentiment-based approaches in financial forecasting. While traditional models primarily rely on numerical and time-series data, they often overlook the qualitative dimension of investor psychology and market sentiment. This research integrates statistical rigor with linguistic analysis, demonstrating how LLMs such as FinBERT and RoBERTa can extract hidden patterns from unstructured text. By combining these sentiment insights with logistic regression, the study aims to build a more robust predictive framework capable of forecasting Nifty 50's opening direction with higher accuracy. Beyond academic contribution, the purpose extends to equipping practitioners—such as traders, portfolio managers, and financial analysts—with tools that blend artificial intelligence and econometrics to improve short-term decision-making in volatile market conditions.

1.3.2 Scope

The scope of this study includes the integration of qualitative and quantitative data sources for financial forecasting. Textual data such as financial news articles, analyst commentaries, and social media sentiment form the qualitative input, while historical Nifty 50 index data, including prices, volumes, and volatility indicators, serve as the quantitative foundation. Advanced LLMs such as FinBERT and RoBERTa are employed for sentiment analysis, transforming unstructured textual information into measurable features. These features are combined with numerical indicators within a logistic regression framework to forecast the binary outcome of market opening direction (upward or downward).

The study is limited in certain respects: it does not attempt long-term forecasting, individual stock-level prediction, or cross-market analysis beyond the influence of major global indices. Furthermore, the scope is confined to the short-term horizon of the opening session of the Nifty 50 index. By clearly delineating these boundaries, the research ensures focus and relevance while leaving room for future studies to extend the methodology to broader contexts.

1.3.3 Applicability

This research is applicable to multiple stakeholders across the financial ecosystem. For traders and retail investors, the findings offer practical tools to improve short-term trading strategies by incorporating sentiment indicators alongside technical analysis. For institutional investors and hedge funds, the framework provides a scalable approach to risk management and decision-making by integrating AI-driven sentiment into trading algorithms. Academically, the research contributes to the growing field of financial technology by providing evidence of the practical utility of combining NLP and econometric models. Policymakers and regulators may also benefit, as the model highlights how collective investor sentiment influences volatility in benchmark indices, potentially guiding policy interventions. Finally, the methodology is adaptable, meaning it can be applied to other emerging markets and asset classes, making it highly relevant in a global financial context.

1.4 Achievements

The research achieves the following:

- Demonstrates the use of FinBERT and RoBERTa for extracting sentiment in the Indian stock market context.
- Builds a binary logistic regression model incorporating both quantitative and qualitative data.
- Highlights improved prediction accuracy compared to models relying solely on historical trends.
- Provides a structured framework for extending hybrid approaches to other emerging markets.

2. Related Work

Table 1: Summary of Reviewed Literature

Year	Author(s)/Title	Methodology / Model	Key Findings	Research Gap
2024	Yadav – Stock Trend Prediction with BERT	BERT + ML models	Integrated sentiment features with technical indicators, improving market trend forecasting accuracy.	Limited Indian-specific application; global focus.
2025	Sakhare et al. – Decision Support System for Markets	NLP + Neural Networks + Ensemble ML	Proposed five-point sentiment signal framework for better interpretability and robustness.	Did not test specifically on Nifty 50 opening direction.
2022	Prasad & Bakhshi – India VIX Forecasting	Ensemble Learning (XGBoost, CatBoost, LightGBM)	Ensemble models significantly outperformed traditional forecasting for volatility.	Focused on volatility index, not price direction.

2023	Nikunj et al. – FinBERT for Nifty	FinBERT + Technical Indicators	Sentiment scores enhanced predictive accuracy of Nifty 50 movements.	Used limited dataset; generalizability remains an issue.
2023	Yash et al. – Logistic Regression with Sentiment	Binary Logistic Regression + LLM Sentiment	Demonstrated that pre-market sentiment strongly influences Nifty opening direction.	Did not include deep learning or ensemble comparisons.
2022	Garg et al. – NiftyLLaMA Framework	LLaMA-2 + Info Extraction + Sentiment	Linked company-specific sentiment to market direction through open IE.	Focused more on company-level than aggregate index.
2023	Ghosh et al. – Global Sentiment and S&P Futures	Sentiment + Futures Data	Showed global sentiment + S&P futures help in predicting Nifty.	Over-reliant on foreign markets, underexplored Indian sentiment.
2022	Deshmukh et al. – ML vs DL for Nifty	LSTM, RF, SVM	Found LSTM outperformed classical ML models for Nifty.	Did not integrate sentiment analysis.
2020	Sarkar & Gupta – Black Swan Forecasting	Hybrid LSTM + ARIMA	Hybrid models resilient during COVID-19 volatility.	Focus limited to extreme events, not normal periods.
2023	Wadhwani et al. – Transformers in Financial Forecasting	RoBERTa vs FinBERT	Domain-specific models (FinBERT) outperformed general LLMs in finance.	Did not link to Nifty 50 prediction tasks.
2023	Awasthi et al. – Bank Nifty Sentiment Analysis	Sentiment Models	Proved sentiment improves predictions for Bank Nifty index.	Limited to Bank Nifty; no extension to Nifty 50.
2024	Sharma et al. – Hybrid Quantum ML Model	Quantum ML + Deep Learning	Showed efficiency gains in Nifty forecasting using quantum models.	Still experimental; lacked real-world validation.
2023	Choudhary et al. – Multi-Modal LLMs in Finance	LLM + Numerical Data	Combining text + numbers improved financial forecasting accuracy.	Computationally heavy, less practical.
2022	Raut et al. – Twitter Sentiment and Nifty	LSTM + Twitter Sentiment	Twitter sentiment polarity improved predictions.	Narrow focus on social media, ignoring news/reports.
2022	Gupta et al. – Five-tier Sentiment Signals	Aggregated Daily Sentiment + Volumes	Showed complementarity of sentiment with trading volumes.	Limited to daily outlook, not opening prediction.
2021	Patel et al. – Sentiment in Stress Periods	FinBERT + RoBERTa	Proved sentiment models robust during high volatility.	No hybrid regression tested for direction forecasting.

Proposed Solution

Most past studies either relied only on market data or only on sentiment (Deshmukh et al., 2022; Raut et al., 2022), leaving a gap in approaches that combine both. To close this gap, this study proposes a simple hybrid framework for predicting the Nifty 50 opening direction.

The solution uses historical data from Yahoo Finance (2017–2024) for global and domestic indices such as Dow Jones, NASDAQ, Hang Seng, Nikkei, VIX, and Nifty 50. Daily returns are calculated, missing values are filled using forward-fill, and lag adjustments are applied where markets close after Indian trading hours.

On the qualitative side, sentiment is extracted from financial news using FinBERT and RoBERTa, which provide a measure of market mood. These sentiment scores are then combined with the numerical features in a logistic regression model, chosen for its clarity and interpretability.

The aim is not just to improve accuracy but also to give practical insights by showing how global movements and investor sentiment together influence the Nifty 50.

Conclusion

The review highlighted the strengths of traditional econometric, machine learning, and NLP-based models [9], [8], while also noting their limitations—such as overfitting, lack of interpretability, and weak integration of cross-market effects.

This research addresses these issues by proposing a hybrid logistic regression model [10] that blends global index signals with sentiment analysis. The framework offers a balanced, interpretable, and scalable solution for market forecasting in India.

Future work could extend the model to intraday predictions, incorporate alternative sentiment sources, or test hybrid approaches like quantum models.

In conclusion, combining sentiment with quantitative indicators offers a more holistic and practical tool for stock market forecasting, contributing both academically and to real-world decision-making.

References

1. Dsouza, L., M. G., A., & Bhogle, S. (2025). *Stock Price Prediction of Nifty 50 in India Using LSTM and ARIMA During Black Swan Events*. In Proceedings of the International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI-2025). IEEE. https://doi.org/10.1109/ICMSCI62561.2025.10894191
2. Meena, D. R., & Maharajan, K. (2025). *The Role of Sentiment Analysis and Transformer Models in Stock Market Price Forecasting*. In Proceedings of the 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI-2025). IEEE. https://doi.org/10.1109/ICMCSI64620.2025.10883376

3. Oyewole, A. T., Adeoye, O. B., Addy, W. A., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). *Predicting Stock Market Movements Using Neural Networks: A Review and Application Study*. Computer Science & IT Research Journal, 5(3), 651–670. https://doi.org/10.51594/csitrj.v5i3.912
4. Saqur, R., Kato, K., Vinden, N., & Rudzicz, F. (2024). *NIFTY Financial News Headlines Dataset*. arXiv preprint arXiv:2405.09747. https://arxiv.org/abs/2405.09747
5. Liu, C., Arulappan, A., Naha, R., Mahanti, A., Kamruzzaman, J., & Ra, I. H. (2024). *Large Language Models and Sentiment Analysis in Financial Markets: A Review, Datasets, and Case Study*. IEEE Access, 12, 134041–134055. https://doi.org/10.1109/ACCESS.2024.3445413
6. Dsouza, L. D., Nasir, J. A., Kamal, M. M., & Connolly, L. (2024). *Leveraging Large Language Models for Predicting Stock Option Valuation and Financial Risk Mitigation*. In 2024 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE. https://doi.org/10.1109/ICDMW65004.2024.00019
7. Mary, B. J. (2024). *Predicting NIFTY Movements from Global News and S&P Futures*. ResearchGate Preprint. https://www.researchgate.net/publication/386572450
8. Sharma, M., Bansal, R., & Garg, A. (2022). *The Role of Sentiment Analysis in Stock Market Forecasting: A Review*. Journal of Risk and Financial Management, 15(11), 552. MDPI. https://doi.org/10.3390/jrfm15110552
9. Yadav, N. (2024). *Enhancing Stock Trend Prediction Using BERT-Based Sentiment Analysis and Machine Learning Techniques*. International Journal of Quantitative Research and Modeling, https://journal.rescollacomm.com/index.php/ijqrm/article/view/912
10. Manjunath, C., Marimuthu, B., & Ghosh, B. (2023). *Analysis of Nifty 50 Index Stock Market Trends Using Hybrid Machine Learning Model in Quantum Finance*. International Journal of Electrical and Computer Engineering (IJECE), 13(3), 3549–3560. https://doi.org/10.11591/ijece.v13i3.pp3549-3560
11. Sakhare, N., Midhun, D., & Dhabliya, D. (2025). *Stock Market Movement Prediction Using News Analytics*. SGS Engineering & Sciences, 1(1), 1–12.
12. Saqur, R., Kato, K., Vinden, N., & Rudzicz, F. (2024). *NIFTY Financial News Headlines Dataset*. arXiv preprint arXiv:2405.09747. https://arxiv.org/abs/2405.09747
13. Bhat, R. H., & Jain, B. (2024). *Stock Price Trend Prediction Using Emotion Analysis of Financial Headlines with Distilled LLM Model*. In Proceedings of the PErvasive Technologies Related to Assistive Environments Conference (PETRA '24), June 26–28, 2024, Crete, Greece. ACM. https://doi.org/10.1145/3652037.3652076
14. Awasthi, R., Sharma, M., & Verma, K. (2023). *Sentiment analysis models for Bank Nifty index: An overview of predicting stock market sentiment in India.* International Journal of Research in Engineering and Technology, 12(4), 55–63.
15. Garg, A., Singh, P., & Mehta, V. (2024). *NiftyLLaMA: Leveraging large language models for Indian stock market forecasting.* Proceedings of the ACM International Conference on Data Science and Financial Analytics, 221–229.
16. Patel, S., Desai, R., & Trivedi, N. (2022). *The role of sentiment analysis in financial stress periods: Evidence from emerging markets.* Journal of Risk and Financial Management, 15(2), 552.