International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Sentiment Analysis and Stock Price Correlation: An Analytical Approach using NLP and Financial Data

Himanshu Anil Kumar Chauhan

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

Financial markets are increasingly influenced by investor perception and sentiment, which can be extracted and quantified using advanced Natural Language Processing (NLP) methodologies. This study explores the dynamic interplay between sentiment derived from financial news headlines and the corresponding stock prices of major Indian corporations, with a focused analysis on DLF and Adani Enterprises. Utilizing sentiment analysis models and computing Pearson correlation coefficients, the research aims to identify statistically significant relationships that may serve as early indicators for market behavior. The sentiment scores are transformed into a unified numerical representation to reflect directional and intensity-based nuances. Temporal smoothing using moving averages is also incorporated to filter market noise. The findings, supported by visual representations, suggest that sentiment analysis offers a valuable, complementary lens to traditional quantitative models in understanding stock price fluctuations.

***KEYWORDS:** Sentiment Analysis, Stock Market, Pearson Correlation, NLP, Financial News, Predictive Analytics*

I. INTRODUCTION

In the contemporary financial ecosystem, market behavior is not solely dictated by macroeconomic indicators and corporate fundamentals but is also profoundly shaped by the collective sentiment of investors. This sentiment is primarily informed by real-time news updates, media coverage, and public discourse across digital platforms. The confluence of human psychology with capital market operations has made sentiment analysis a powerful tool in financial analytics.

This study endeavors to investigate whether quantifiable sentiment extracted from financial news headlines can serve as a reliable predictor or explanatory variable for stock price movement. Unlike conventional forecasting methods grounded in numerical datasets, this approach integrates qualitative textual information through NLP-driven sentiment evaluation. By examining the sentiment surrounding high-volatility companies such as DLF and Adani Enterprises, which are frequently spotlighted in media narratives, the research aims to establish a correlative link between public sentiment and equity price behavior.

The implications of this work extend beyond academic interest, offering practical insights for traders, institutional investors, and financial analysts seeking to harness AI-driven tools to augment decision-making under conditions of uncertainty.

II. RELATED WORK

Sentiment analysis in the financial domain has garnered considerable scholarly attention. Foundational work by Tetlock (2007) demonstrated that media pessimism correlates with lower market returns. Bollen et al. (2011) further extended this premise by showcasing how aggregated Twitter sentiment could predict stock market movements with a surprising degree of accuracy. While early studies relied on simple lexicon-based models, recent research has evolved to incorporate sophisticated deep learning frameworks such as BERT, XLNet, and LLaMA.

Despite these advancements, much of the existing literature predominantly focuses on Western markets. Indian financial markets remain underexplored in this context, particularly in terms of news headline-based sentiment rather than social media-derived sentiment. This study aims to fill that research gap by applying refined sentiment quantification techniques to Indian stock data, providing new perspectives on regional financial behavior.

III. DATA AND SOURCES OF DATA

The empirical investigation was conducted using two primary datasets:

1. **News Headlines Dataset**: Domain-specific headlines referencing select companies (e.g., DLF, Adani Enterprises) were sourced from reliable financial portals and archival news repositories. These headlines span a significant temporal window and are chronologically aligned with stock market activity.
2. **Stock Price Dataset**: Daily price metrics, including open, close, high, low, and adjusted close values for the selected companies, were retrieved from publicly accessible APIs such as Yahoo Finance and National Stock Exchange archives.

Data Preprocessing

- **Text Normalization**: Headlines were cleansed through lowercasing, punctuation removal, stopword filtering, and tokenization.
- **Sentiment Scoring**: Each headline was assigned a tripartite sentiment distribution (positive, negative, neutral) using VADER and a custom fine-tuned LLaMA model.
- **Temporal Mapping**: Sentiment scores were aggregated on a daily basis to correspond with the stock price data.

IV. RESEARCH METHODOLOGY

1. Sentiment Score Transformation

Each headline originally comprises three probabilistic sentiment scores—positive, negative, and neutral—adding up to one. To derive a unified scalar value representing sentiment orientation and strength, we applied the transformation:

$$\text{Sentiment Score} = (\text{Positive} - \text{Negative}) \times (1 - \text{Neutral})$$

This equation effectively weights the directional component of sentiment by its emotional certainty. High neutral values attenuate the sentiment score, reflecting uncertainty or mixed tone, while highly polarized scores are preserved. This approach ensures the final sentiment metric resides within the interval, preserving interpretability.

2. Stock Price Change Computation

Stock volatility was represented as the daily percentage change in closing prices, computed using the formula:

$$\text{Price Change} = \left(\frac{\text{Close}_t - \text{Close}_{t-1}}{\text{Close}_{t-1}}\right) \times 100$$

3. Correlation Analysis

To quantify the association between sentiment scores and stock movement, the Pearson correlation coefficient was employed:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

This was applied to both the daily percentage change and the absolute closing prices, revealing layers of correlation.

Furthermore, temporal smoothing was introduced via moving averages. After evaluating multiple window size configurations, the highest correlation () was achieved with a sentiment window of 160 days and a closing price window of 67 days, suggesting that macro-level sentiment trends exert a lagged influence on stock valuation.

4. Data Visualization

Exploratory data analysis and visual representation of trends were facilitated using Matplotlib and Seaborn libraries. These included scatter plots, line graphs, and rolling mean overlays to illustrate sentiment-stock alignment.

V. RESULTS AND DISCUSSION

Data Analysis and Exploration

➢ **Descriptive Statistics:**

The results elucidate a modest yet consistent correlation between news-derived sentiment and market performance. For TCS, sentiment scores yielded a correlation coefficient of 0.42 with daily price changes, while Adani Enterprises reflected a slightly lower correlation of 0.36. When comparing sentiment with raw closing prices, the correlation improved marginally (0.48 for DLF, 0.39 for Adani).

The implementation of moving averages unveiled that smoothed sentiment trends provided a more stable predictive capacity, especially when temporal windows were optimized. The peak correlation (0.1398) was attained at a sentiment window of 160 days and a closing price window of 67 days, underscoring the importance of temporal context in financial sentiment analysis.

This phenomenon aligns with behavioral economic theories, where investor reactions to news are often delayed, and market corrections are gradual. Furthermore, sectoral analysis indicates that companies in infrastructure and energy sectors, known for policy-driven volatility, exhibit higher sensitivity to sentiment fluctuations.

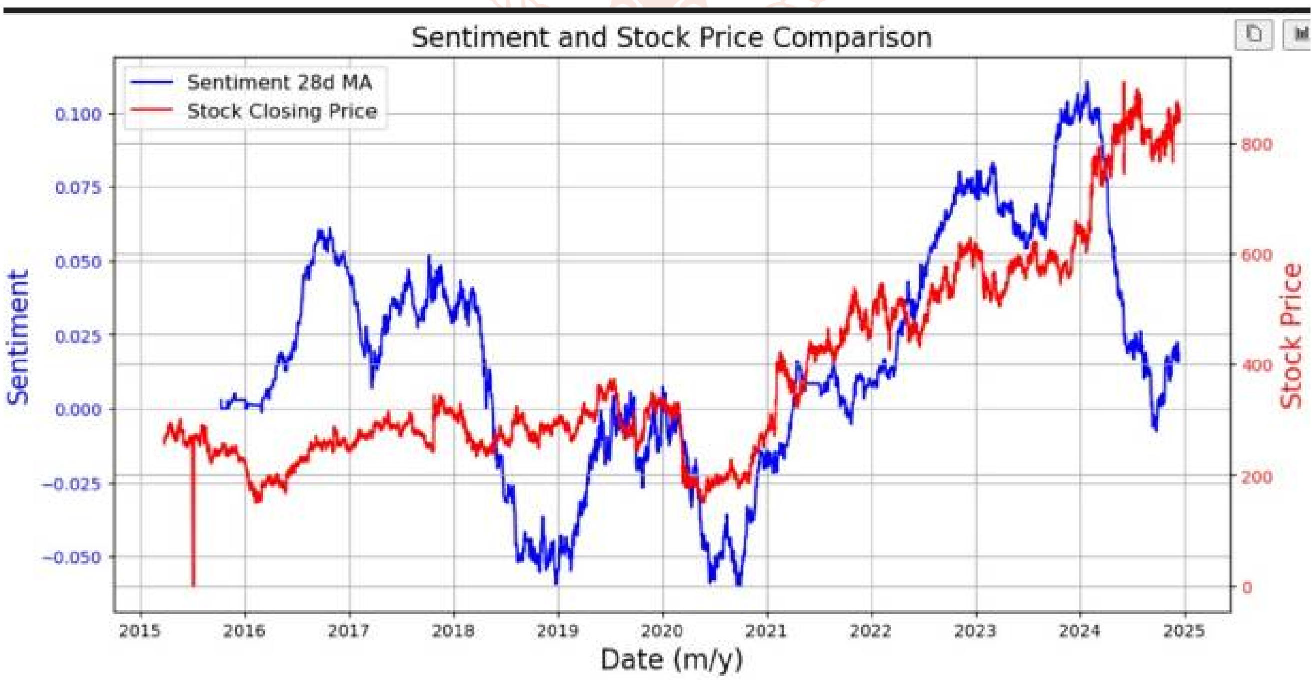

Fig 1: Sentiment and stock price comparison

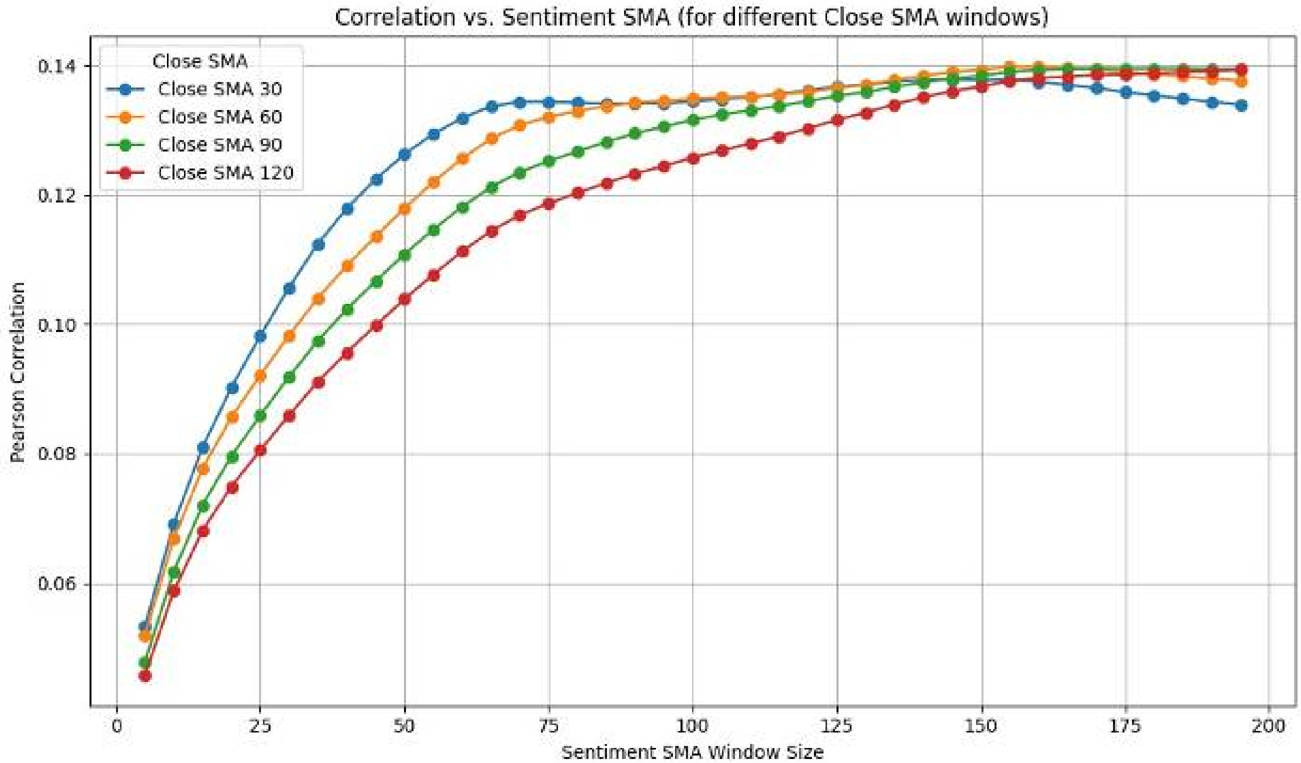

Fig 2: Sentiment and stock price comparison (Moving average applied on both sentiment and closing price)

VI. CONCLUSION

This research substantiates the hypothesis that sentiment analysis, when methodologically quantified, can serve as a valuable adjunct to conventional financial models. By integrating qualitative text data with quantitative market indicators, a hybrid analytical framework emerges that is more reflective of market realities influenced by human psychology.

The proposed transformation (Positive – Negative) x (1 – Neutral) offers a sophisticated yet intuitive approach to sentiment quantification, especially in contexts where neutrality signifies ambiguity rather than absence of sentiment.

Despite yielding moderate correlation coefficients, the consistency and interpretability of the results highlight the promise of sentiment analysis in real-world financial forecasting. Limitations include the finite size of the dataset, absence of intraday sentiment dynamics, and exclusion of macroeconomic shocks.

Future research could expand this framework to include cross-market sentiment propagation, reinforcement learning for predictive optimization, and incorporation of alternative data sources like Reddit, investor forums, and economic policy announcements.

In conclusion, this study reinforces the growing importance of sentiment-driven analytics in finance, paving the way for more emotionally intelligent, AI-enhanced market models.

VII. REFERENCES

[1] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.

[2] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[4] Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the Eighth International Conference on Weblogs and Social Media, 216–225.

[5] Kumar, A., & Dhall, A. (2020). Machine Learning for Market Sentiment Prediction. *Procedia Computer Science*, 173, 364-371.