A Comprehensive Literature Review on Enhancing Stock Price Prediction with Sentiment Analysis: The Case of Bank Nifty Index

Dr. Ravikant Zirmite
Associate Professor, MCA Department,
MES's Institute of Management and Career Courses,
Pune, India
rsz.imcc@mespune.in

Prakash Ukhalkar
PhD Scholar, Assistant Professor, MCA Department,
Pimpri Chinchwad College of Engineering,
Pune, India
prakashu.mca@gmail.com

Dr. Manasi Bhate
Professor, MCA Department,
MES's Institute of Management and Career Courses,
Pune, India
msb.imcc@mespune

Dr. Venugopal Narsingoju
Associate Professor, MCA Department,
MES's Institute of Management and Career Courses,
Pune, India
vsn.imcc@mespune.in

Abstract—
Sentiment analysis has emerged as a powerful tool in financial markets, offering the ability to harness market participants' collective wisdom and emotions to make more informed investment decisions. This research paper presents a comprehensive literature review on sentiment analysis models in predicting stock prices, focusing on the Bank Nifty Index, a critical benchmark in the Indian financial sector. The review begins by elucidating the significance of sentiment analysis in stock price prediction, highlighting its pivotal role in capturing market sentiment, investor emotions, and their impact on asset valuations. A detailed overview of sentiment sources, encompassing financial news, social media, reports, and other relevant data streams, provides insight into the rich tapestry of information that sentiment analysis can leverage. The paper delves into various sentiment analysis approaches, from traditional lexicon-based and rule-based methods to cutting-edge machine learning and deep learning techniques. It elucidates the intricate process of integrating sentiment analysis with stock price prediction models, exploring feature concatenation, time series integration, and weighted data to incorporate sentiment insights effectively. Evaluation metrics crucial for assessing the performance of both sentiment analysis models and stock price prediction models are thoroughly discussed. It concludes by highlighting the substantial potential of sentiment analysis in augmenting stock price prediction, offering investors and traders valuable tools to navigate the complexities of financial markets.
Keywords—Sentiment Analysis, Stock Price Prediction, Bank Nifty Index, Financial Markets, Sentiment Sources

I. INTRODUCTION

The modern financial landscape is driven not only by traditional factors such as economic indicators and company financials but also by sentiment and public perception. Sentiment analysis, a subset of natural language processing (NLP), has emerged as a valuable tool for gauging market sentiment from textual data sources like financial news articles, social media, and analyst reports[1]. Understanding sentiment is crucial for investors and traders as it can provide insights into market trends, influence market decisions, and impact stock prices. Sentiment analysis can capture explicit opinions and subtle nuances, making it a valuable component of predictive models for financial markets[2]. The Bank Nifty Index, a subsidiary of the Nifty 50 Index, comprises the most liquid and actively traded banking stocks listed on the National Stock Exchange of India (NSE). It is a

DOI:10.48047/intjecse/v14i5.1188 ISSN: 1308-5581 Vol 14, Issue 05 2022

vital barometer of the Indian banking sector's health and performance[3]. The banking sector plays a pivotal role in India's economy, facilitating credit creation, economic growth, and financial stability. As such, the Bank Nifty Index serves as a bellwether for broader economic trends and investor sentiment in the country[4].

This literature review aims to provide a comprehensive overview of sentiment analysis methods and their applications in predicting stock prices, with a specific focus on the Bank Nifty Index. By synthesizing existing research, this review aims to offer insights into the efficacy, challenges, and emerging trends in utilizing sentiment analysis for stock price prediction in the Indian financial context.

This literature review encompasses extensive sources, including academic papers, industry reports, and empirical studies, to offer a holistic understanding of the subject. It will explore various sentiment analysis techniques, data sources, feature engineering methods, and their integration with stock price prediction models.

By addressing these aspects, this literature review seeks to shed light on the significance of sentiment analysis in stock price prediction and its applicability to the unique dynamics of the Bank Nifty Index.

II. Sentiment Analysis in Financial Markets

Sentiment analysis, also known as opinion mining, is a branch of natural language processing (NLP) that focuses on extracting and analyzing subjective information from textual data. It aims to determine a text's sentiment or emotional tone, such as positive, negative, or neutral feelings[5]. Sentiment analysis is paramount in financial markets because it can provide valuable insights into investor sentiment, market dynamics, and price movements. Understanding market sentiment can help traders, investors, and financial institutions make more informed decisions, mitigate risks, and capitalize on market opportunities[6].

A. Role of Sentiment Analysis in Predicting Stock Prices

Sentiment analysis plays a crucial role in predicting stock prices by offering a complementary perspective to traditional financial analysis. Here are some key aspects of its role:

- Early Warning Indicator: Sentiment analysis can serve as an early warning system, detecting shifts in market sentiment before they manifest in stock price movements. For example, a surge in negative sentiment in news articles and social media posts may precede a stock market downturn [7].
- Enhancing Predictive Models: Sentiment scores extracted from textual data can be integrated into predictive models, enhancing their accuracy. By considering both quantitative and qualitative factors, models can capture the influence of sentiment on stock prices[8].
- Risk Management: Sentiment analysis helps identify potential market risks and sentiment-driven market events. Investors and traders can adjust their portfolios or trading strategies in response to changing sentiment patterns[9].

B. Overview of Sentiment Sources

- News Sources: Financial news outlets, both traditional and digital, are primary sources of sentiment data. These sources include news articles, press releases, and financial reports. News sentiment can be particularly impactful on stock prices, especially when significant events or earnings reports are announced [10].
- Social Media: Social media platforms like Twitter, Facebook, and Reddit are rich sources of real-time sentiment data. Traders and investors monitor social media sentiment to gauge public sentiment and its potential impact on stock prices [11].
- Reports and Analyst Insights: Analyst reports, corporate earnings calls, and financial research publications also contain sentiment-rich information. Analyst sentiment can influence market sentiment, making these sources valuable for stock price prediction[12].
- Blogs and Forums: Online investment blogs and forums like Seeking Alpha and StockTwits serve as informal sources of sentiment data. Investors and traders share opinions, analyses, and sentiments, which can provide additional insights [13].

III. Sentiment Analysis Approaches

The sentiment analysis approaches from the literature review identified are from traditional lexicon-based methods to advanced deep learning models which are suitable to stock price predictions using sentiment analysis.

A. Traditional Methods for Sentiment Analysis

- Lexicon-Based Approaches:Lexicon-based sentiment analysis relies on predefined sentiment dictionaries that associate words with positive or negative sentiments. Words in the text are assigned scores, and the overall sentiment is determined based on the sum or average of these scores. For example, words like "good" may have a positive score, while words like "bad" may have a negative score [14].

DOI:10.48047/intjecse/v14i5.1188 ISSN: 1308-5581 Vol 14, Issue 05 2022

- Rule-Based Systems: Rule-based sentiment analysis involves using manually crafted linguistic rules to analyse text sentiment. These rules can account for negations (e.g., "not good"), intensifiers (e.g., "very good"), and other linguistic nuances that affect sentiment. Rule-based systems are highly customizable but may require continuous updates to remain effective [15].

B. *Machine Learning Techniques*

- Supervised Learning (SVM, Random Forest): Supervised learning techniques for sentiment analysis require labelled training data, where each text example is associated with a sentiment label (positive, negative, neutral). Support Vector Machines (SVM) and Random Forest classifiers are commonly used for this task. They learn to classify text based on features derived from the data [11].
- Unsupervised Learning (Clustering, LDA): Unsupervised learning methods aim to identify patterns and clusters within text data without relying on labelled examples. Clustering techniques like K-Means can group similar texts together based on their content. Latent Dirichlet Allocation (LDA) is a probabilistic model that uncovers topics in a collection of texts, which can indirectly reveal sentiment patterns [16].

C. Deep Learning Methods

- Recurrent Neural Networks (RNNs): Recurrent Neural Networks (RNNs) are designed to handle sequential data, making them suitable for sentiment analysis tasks. They can capture dependencies and context in text by considering previous words in the sequence. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular RNN variants used in sentiment analysis[17].
- Transformers: Transformers, particularly models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformers), have revolutionized sentiment analysis. BERT, for example, can capture contextual information by considering both preceding and following words in a text, making it highly effective in understanding sentiment nuances[18].

IV. Sentiment Analysis for Bank Nifty

The following techniques provide comprehensive insights into the data sources, data preprocessing techniques, and sentiment lexicons, and resources specific to sentiment analysis for Bank Nifty, enabling a more accurate analysis of sentiment in the financial domain.

A. *Data Sources for Bank Nifty Sentiment Analysis*

- Financial News Articles: Financial news articles from reputable sources such as Bloomberg, Reuters, and Financial Times provide valuable sentiment data for Bank Nifty analysis. News articles often contain insights, opinions, and expert views on the banking sector, influencing sentiment.
- Social Media:Social media platforms like Twitter, LinkedIn, and specialized finance forums are rich sources of real-time sentiment data. Users discuss banking sector news, investment strategies, and opinions, making social media an essential sentiment source.
- Earnings Reports:Quarterly and annual earnings reports of major banks included in the Bank Nifty Index can provide structured sentiment data. These reports often include management commentary, which can influence market sentiment[19].

B. *Data Preprocessing Techniques for Financial Data*

- Text Cleaning:Financial text data often contains noise, including punctuation, special characters, and irrelevant information. Preprocessing involves removing these elements and ensuring text data is in a clean, usable format[20].
- Tokenization:Tokenization splits text into individual words or tokens, which facilitates analysis. It's crucial for creating word-based sentiment features for models[21].
- Stopword Removal:Stopwords (common words like "the" and "and") are often removed from financial text as they typically don't carry sentiment information[22].
- Lemmatization and Stemming:Lemmatization reduces words to their base or dictionary form (e.g., "running" to "run"), while stemming reduces words to their root form (e.g., "running" to "run"). Both techniques aid in feature extraction[20].

C. Sentiment Lexicons and Resources Specific to the Banking and Financial Domain

- Financial Sentiment Lexicons:Specialized sentiment lexicons or dictionaries tailored to the financial domain contain sentiment scores associated with financial terms. These lexicons help sentiment analysis models understand financial language[23].
- Word Embeddings:Word embeddings like Word2Vec and GloVe trained on financial text can capture context-specific relationships between words in the banking and financial domain. These embeddings enhance the understanding of financial sentiment[24][25].

DOI:10.48047/intjecse/v14i5.1188 ISSN: 1308-5581 Vol 14, Issue 05 2022

- Financial Ontologies:Financial ontologies provide structured knowledge about financial concepts, relationships, and entities. These ontologies can be valuable for sentiment analysis by providing context[26].

V. Stock Price Prediction Models

The various stock price prediction models, including regression-based models, time series forecasting techniques, and machine learning algorithms, that can be applied to Bank Nifty data. The literature review help to explore these models to experiment with the stock market data for forecasting the stock prices and perform the sentiment analysis.

A. Regression-Based Models

- Linear Regression:Linear regression models attempt to establish a linear relationship between independent variables (features) and the target variable (stock price). In stock price prediction, features may include historical prices, trading volumes, and sentiment scores. The model aims to find coefficients that minimize the error between predicted and actual prices[27].
- Polynomial Regression:Polynomial regression extends linear regression by allowing for non-linear relationships between features and stock prices. It can capture more complex patterns in the data[28].

B. Time Series Forecasting Techniques

- ARIMA (AutoRegressive Integrated Moving Average): ARIMA models are designed specifically for time series data like stock prices. They consider auto-regressive and moving average components and account for seasonality and trends. ARIMA models are widely used for short-term stock price forecasting[29].
- Exponential Smoothing (Holt-Winters): Exponential smoothing methods forecast stock prices by giving more weight to recent observations. Holt-Winters is an extension that considers trend and seasonality, making it suitable for capturing complex stock price patterns[30].

C. Machine Learning Models for Stock Price Prediction

- Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It's robust, handles non-linear relationships, and can capture complex interactions between features[27].
- Support Vector Machines (SVM): SVMs aim to find a hyperplane that best separates data points into different classes. In stock price prediction, SVMs can be used for regression tasks, making them versatile models for price forecasting[31].
- Recurrent Neural Networks (RNNs): RNNs are neural networks designed for sequential data, making them suitable for time series-based stock price prediction. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular RNN variants for this task[32].
- Gradient Boosting (XGBoost, LightGBM): Gradient boosting algorithms like XGBoost and LightGBM are powerful for regression tasks. They construct a strong predictive model by iteratively adding weak learners[33].

VI. Integrating Sentiment Analysis and Stock Price Prediction

The research literature review provides comprehensive insights into methods for integrating sentiment analysis with stock price prediction models, feature engineering strategies, and the challenges and limitations associated with such integration.

A. Methods for Combining Sentiment Analysis with Stock Price Prediction Models

- Feature Concatenation: One straightforward method is to concatenate sentiment scores or features derived from sentiment analysis with traditional financial features used in stock price prediction. For example, researcher can include sentiment polarity scores, sentiment volatility, or sentiment divergence as additional features alongside historical stock prices, trading volumes, and other relevant financial variables[34].
- Time Series Integration: In this approach, sentiment data is converted into a time series, capturing daily, weekly, or even intraday sentiment trends. This sentiment time series is then combined with historical stock price time series. Time series forecasting models, such as ARIMA or LSTM, can be used to predict future stock prices based on this integrated time series data[35].
- Weighted Data: Assigning weights to historical data points based on sentiment scores is another method. Recent sentiment scores might carry more weight than older ones. This weighted data is then used to train stock price prediction models. Weighted regression or machine learning algorithms can be employed for this purpose[36].

DOI:10.48047/intjecse/v14i5.1188 ISSN: 1308-5581 Vol 14, Issue 05 2022

B. Feature Engineering Strategies

- Sentiment Aggregation: Aggregating sentiment scores over different time intervals (e.g., daily, weekly) can help capture longer-term sentiment trends. For instance, researcher can calculate weekly sentiment averages or monthly sentiment volatility. These aggregated sentiment features can be used alongside other financial features for prediction[13].
- Sentiment Volatility: Measuring the volatility of sentiment scores can provide insights into how rapidly market sentiment is changing. High sentiment volatility might indicate uncertainty or abrupt shifts in investor sentiment, which could impact stock prices[37].
- Sentiment Divergence: Analysing the divergence between sentiment and stock prices is a valuable strategy. For example, if sentiment is strongly positive while stock prices are declining, it may signal a potential market anomaly. Creating divergence-based features can help capture such situations[38].

C. Challenges and Limitations in Integration

- Data Quality and Noise: Sentiment data, particularly from social media, can be noisy and unreliable. Distinguishing between genuine sentiment and noise or spam is a significant challenge in sentiment analysis. Low-quality sentiment data can lead to inaccurate predictions[31].
- Model Overfitting: Integrating sentiment data can lead to overfitting if not handled carefully. Overfit models may perform well on training data but poorly on unseen data. Regularization techniques and feature selection are essential to mitigate this issue[39].
- Causality vs. Correlation: Determining causality between sentiment and stock prices is challenging. Sentiment analysis often identifies correlations, but proving causation requires rigorous analysis and may still be elusive[40].

VII. EVALUATION METRICS

The literature reviews help to find the various evaluation metrics for stock price prediction models, sentiment analysis accuracy, and domain-specific evaluation criteria, enabling a thorough assessment of integrated sentiment analysis models in financial contexts.

A. Metrics for Evaluating Stock Price Prediction Models[32][41]

- Mean Absolute Error (MAE): MAE measures the average absolute difference between predicted and actual stock prices. Lower MAE indicates better predictive accuracy.

  $$MAE = \Sigma|Actual - Predicted| / n \qquad (1)$$

- Mean Squared Error (MSE): MSE calculates the average squared difference between predicted and actual prices. It penalizes larger errors more heavily than MAE.

  $$MSE = \Sigma(Actual - Predicted)\hat{}2/n \qquad (2)$$

- Root Mean Squared Error (RMSE): RMSE is the square root of MSE, providing an interpretable measure in the same unit as the stock prices. It is sensitive to outliers.

  $$RMSE = \sqrt{}(MSE) \qquad (3)$$

- Mean Absolute Percentage Error (MAPE): MAPE expresses the prediction error as a percentage of the actual value. It is useful for understanding the relative size of errors.

  $$MAPE = (\Sigma|Actual - Predicted| / Actual) * 100 / n \qquad (4)$$

- R-squared (R^2) Score: R^2 measures the proportion of the variance in the dependent variable (stock price) that is predictable from the independent variables (features).

  $$R^2 = 1 - (SSR / SST) \qquad (5)$$

  SSR stands for "Sum of Squares Residual," and it represents the sum of the squared differences between the predicted values (usually from a regression model) and the actual observed values. In other words, SSR quantifies the variation in the dependent variable that is not explained by the regression model.

  SST stands for "Total Sum of Squares," and it represents the sum of the squared differences between the actual observed values and the mean of those values. SST quantifies the total variation in the dependent variable.

  Higher R^2 values indicate that the model does a better job of explaining the variation in the dependent variable, while lower R^2 values suggest that the model is less effective in explaining the variation.

- Directional Accuracy (DA):DA assesses whether the model correctly predicts the direction of stock price movements (e.g., up or down). It is valuable for binary classification tasks.

  $$Directional\ Accuracy\ (DA) = (Number\ of\ Correct\ Predictions) / (Total\ Predictions)$$

DOI:10.48047/intjecse/v14i5.1188 ISSN: 1308-5581 Vol 14, Issue 05 2022

(6)

A. Metrics for Evaluating Sentiment Analysis Accuracy[42]

- Accuracy: Accuracy measures the overall correctness of sentiment predictions. It is the ratio of correctly classified sentiment labels to the total number of samples.

$$Accuracy = (True\ Positives + True\ Negatives) / Total\ Samples \quad (7)$$

- Precision and Recall: Precision measures the proportion of true positive predictions among all positive predictions. Recall measures the proportion of true positives among all actual positives. These metrics are useful for imbalanced datasets.

$$Precision = True\ Positives / (True\ Positives + False\ Positives) \quad (8)$$

$$Recall = True\ Positives / (True\ Positives + False\ Negatives) \quad (9)$$

- F1-Score:The F1-score is the harmonic mean of precision and recall. It balances precision and recall and is particularly useful when the dataset is imbalanced.

$$F1 - score = 2 * (Precision * Recall) / (Precision + Recall) \quad (10)$$

B. Domain-Specific Evaluation Criteria[15][43][44]

- Sentiment Consistency with Market Movements:Evaluate how well sentiment analysis aligns with actual market movements. Calculate metrics like correlation coefficients or cross-correlation to assess the relationship between sentiment trends and stock price trends.
- Impact of Sentiment Features: Analyse the contribution of sentiment-related features to stock price prediction models. Feature importance scores or regression coefficients can indicate the influence of sentiment in the prediction.
- Evaluation Against Market Benchmarks:Compare the performance of sentiment-integrated stock price prediction models against established market benchmarks (e.g., S&P 500) to assess their effectiveness.
- Risk-Adjusted Metrics:Consider risk-adjusted metrics like the Sharpe ratio or Sortino ratio to evaluate the risk-adjusted performance of a trading strategy based on sentiment predictions.

VIII. Conclusion

The literature review has provided several valuable insights into the application of sentiment analysis in the context of stock price prediction, particularly concerning the Bank Nifty Index. Sentiment analysis plays a pivotal role in capturing market sentiment and investor emotions, which can significantly impact stock prices. Researchers and practitioners have extensively explored its potential. Sentiment analysis leverages diverse data sources, including financial news, social media, earnings reports, and analyst opinions. Each source contributes unique information that can aid in predicting stock price movements. Various sentiment analysis approaches exist, ranging from traditional lexicon-based and rule-based methods to sophisticated machine learning techniques. Deep learning models, such as RNNs and Transformers, have shown promise in capturing nuanced sentiment. While sentiment analysis has been applied broadly in the financial sector, there is a growing interest in applying it specifically to the Bank Nifty Index, given its significance in the Indian financial landscape. The integration of sentiment analysis with stock price prediction models is a complex task. Researchers have explored methods like feature concatenation, time series integration, and weighted data to incorporate sentiment information effectively. Proper evaluation of sentiment-integrated prediction models involves a range of metrics, including MAE, MSE, RMSE, MAPE for prediction models, and accuracy, precision, recall, and F1-score for sentiment analysis accuracy.

In conclusion, sentiment analysis holds significant promise in improving stock price prediction, including for the Bank Nifty Index. By leveraging diverse sentiment data sources, advanced analysis techniques, and appropriate evaluation metrics, practitioners can harness the potential of sentiment analysis to make more informed investment decisions and achieve better predictive accuracy in the dynamic world of finance.

References

[1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008, doi: 10.1561/1500000011.

[2] B. Yilmaz, "Stock Market Sentiment Analysis in 2023," *AIMultiple*, 2023. https://research.aimultiple.com/sentiment-analysis-stock-market/ (accessed Sep. 02, 2023).

[3] T. Reuters -NSEBANK, “August 31, 2023 Bloomberg-NSEBANK Index”, Accessed: Sep. 02, 2023. [Online]. Available: www.niftyindices.com

[4] "R. T. Ranganathan Badrinarayanan, Ashwaryaa Bhatia, Peeyush Dalmia, Satviek Goel, Siddhartha Gupta, Piyushi Jain, Siddhartha Jogidasani, Madhur Maheshwari", “Indian banks: Building resilient leadership | McKinsey,” *McKinsey*, 2023. https://www.mckinsey.com/featured-insights/india/indian-banks-building-resilient-leadership (accessed Sep. 02, 2023).

[5] B. Liu, “Sentiment analysis and opinion mining,” *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–184, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[6] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, doi: 10.1016/j.jocs.2010.12.007.

[7] P. Koukaras, C. Nousi, and C. Tjortjis, “Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning,” *Telecom*, vol. 3, no. 2, pp. 358–378, Jun. 2022, doi: 10.3390/TELECOM3020019.

[8] T. Preis, H. S. Moat, and H. Eugene Stanley, “Quantifying trading behavior in financial markets using google trends,” *Sci. Rep.*, vol. 3, no. 1, pp. 1–6, Apr. 2013, doi: 10.1038/srep01684.

[9] H. P H and A. Rishad, “An empirical examination of investor sentiment and stock market volatility: evidence from India,” *Financ. Innov.*, vol. 6, no. 1, pp. 1–15, Dec. 2020, doi: 10.1186/s40854-020-00198-x.

[10] Y. Dong, D. Yan, A. I. Almudaifer, S. Yan, Z. Jiang, and Y. Zhou, “BELT: A Pipeline for Stock Price Prediction Using News,” *Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020*, pp. 1137–1146, 2020, doi: 10.1109/BigData50022.2020.9378345.

[11] L. Yang *et al.*, “Machine Learning Stock Market Prediction Studies : Review and Research Directions,” *IEEE Access*, vol. 2018-Janua, no. 1, pp. 1–3, 2020, doi: 10.1109/INFOMAN.2019.8714662.

[12] R. Tortoriello, “Their Sentiments Exactly : Sentiment Signal Diversity Creates Alpha Opportunity,” *S&P Glob. Mark. Intell.*, no. September, pp. 1–24, 2018.

[13] R. Gupta and M. Chen, “Sentiment Analysis for Stock Price Prediction,” *Proc. - 3rd Int. Conf. Multimed. Inf. Process. Retrieval, MIPR 2020*, pp. 213–218, 2020, doi: 10.1109/MIPR49039.2020.00051.

[14] D. Kumar, “Stock Forecasting Using Natural Language and Recurrent Network,” *Proc. 3rd Int. Conf. Emerg. Technol. Comput. Eng. Mach. Learn. Internet Things, ICETCE 2020*, no. February, pp. 36–40, 2020, doi: 10.1109/ICETCE48199.2020.9091732.

[15] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-basedmethods for sentiment analysis,” *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, Jun. 2011, doi: 10.1162/COLI_a_00049.

[16] A. Mahajan, L. Dey, and S. M. Haque, “Mining financial news for major events and their impacts on the market,” *Proc. - 2008 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2008*, pp. 423–426, 2008, doi: 10.1109/WIIAT.2008.309.

[17] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, “NSE Stock Market Prediction Using Deep-Learning Models,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1351–1362, 2018, doi: 10.1016/j.procs.2018.05.050.

[18] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, “A review on sentiment analysis from social media platforms,” *Expert Syst. Appl.*, vol. 223, no. August 2022, p. 119862, 2023, doi: 10.1016/j.eswa.2023.119862.

[19] NSE, “Nifty Bank F&O: Stocks, Lot Size, Quotes, Quantity Freeze - NSE India,” *Bank Nifty, NSE India*, 2023. https://www.nseindia.com/products-services/equity-derivatives-banknifty (accessed Sep. 02, 2023).

[20] M. Rashmi, “Introduction to Information Retrieval Systems,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 3, no. 4, pp. 2051–2054, 2015, doi: 10.17762/ijritcc2321-8169.150462.

[21] D. Jurafsky and J. Martin, “Speech and Language Processing,” in *Speech and Language Processing.*, 2014, pp. 441–458. Accessed: Sep. 02, 2023. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/

[22] R. Rani and D. K. Lobiyal, “Performance evaluation of text-mining models with Hindi stopwords lists,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2771–2786, Jun. 2022, doi: 10.1016/J.JKSUCI.2020.03.003.

[23] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,”

Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/J.ASEJ.2014.04.011.

[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, Jan. 2013, Accessed: Sep. 02, 2023. [Online]. Available: https://arxiv.org/abs/1301.3781v3

[25] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014, pp. 1532–1543. doi: 10.3115/v1/d14-1162.

[26] E. Bjoraa, "Ontology guided financial knowledge extraction from semi- structured information sources," Agder University College, 2003.

[27] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *Int. J. Financ. Stud. 2023, Vol. 11, Page 94*, vol. 11, no. 3, p. 94, Jul. 2023, doi: 10.3390/IJFS11030094.

[28] A. Sharma, D. Bhuriya, and U. Singh, "Survey of stock market prediction using machine learning approach," *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2017*, vol. 2017-Janua, pp. 506–509, 2017, doi: 10.1109/ICECA.2017.8212715.

[29] S. M. Idrees, M. A. Alam, and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," *IEEE Access*, vol. 7, pp. 17287–17298, 2019, doi: 10.1109/ACCESS.2019.2895252.

[30] B. S. Reddy, "Prediction of Stock Market indices - Using SAS," *Proc. - 2010 2nd IEEE Int. Conf. Inf. Financ. Eng. ICIFE 2010*, pp. 112–116, 2010, doi: 10.1109/ICIFE.2010.5609262.

[31] R. Y. Nivetha and C. Dhaya, "Developing a Prediction Model for Stock Analysis," *Proc. - 2017 Int. Conf. Tech. Adv. Comput. Commun. ICTACC 2017*, vol. 2017-Octob, pp. 1–3, 2017, doi: 10.1109/ICTACC.2017.11.

[32] K. R. Dahal *et al.*, "A comparative study on effect of news sentiment on stock price prediction with deep learning architecture," *PLoS One*, vol. 18, no. 4, p. e0284695, 2023, doi: 10.1371/journal.pone.0284695.

[33] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, "Predicting the direction of stock market prices using tree-based classifiers," *North Am. J. Econ. Financ.*, vol. 47, no. December 2017, pp. 552–567, 2019, doi: 10.1016/j.najef.2018.06.013.

[34] M. Faraz, H. Khaloozadeh, and M. Abbasi, "Stock Market Prediction-by-Prediction Based on Autoencoder Long Short-Term Memory Networks," *2020 28th Iran. Conf. Electr. Eng. ICEE 2020*, no. December, 2020, doi: 10.1109/ICEE50131.2020.9261055.

[35] A. Sharaff and M. Choudhary, "Comparative Analysis of Various Stock Prediction Techniques," *Proc. 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018*, no. Icoei, pp. 735–738, 2018, doi: 10.1109/ICOEI.2018.8553825.

[36] S. R. Dash and D. Maitra, "Does sentiment matter for stock returns? Evidence from Indian stock market using wavelet approach," *Financ. Res. Lett.*, vol. 26, pp. 32–39, 2018, doi: 10.1016/j.frl.2017.11.008.

[37] F. Audrino, F. Sigrist, and D. Ballinari, "The impact of sentiment and attention measures on stock market volatility," *Int. J. Forecast.*, vol. 36, no. 2, pp. 334–357, Apr. 2020, doi: 10.1016/J.IJFORECAST.2019.05.010.

[38] J. Ayala, M. García-Torres, J. L. V. Noguera, F. Gómez-Vela, and F. Divina, "Technical analysis strategy optimization using a machine learning approach in stock market indices[Formula presented]," *Knowledge-Based Syst.*, vol. 225, p. 107119, 2021, doi: 10.1016/j.knosys.2021.107119.

[39] D. Shah, W. Campbell, and F. H. Zulkernine, "A Comparative Study of LSTM and DNN for Stock Market Forecasting," *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, no. 1, pp. 4148–4155, 2019, doi: 10.1109/BigData.2018.8622462.

[40] M. Babu, C. Hariharan, and S. Srinivasan, "Inter Linkages of Asian Pacific Stock Markets : An Empirical Study," *AJF Amity J. Financ.*, vol. 1, no. 1, pp. 48–56, 2016.

[41] A. Atkins, M. Niranjan, and E. Gerding, "Financial news predicts stock market volatility better than close price," *J. Financ. Data Sci.*, vol. 4, no. 2, pp. 120–137, Jun. 2018, doi: 10.1016/J.JFDS.2018.02.002.

[42] Y. Li and Y. Pan, "A novel ensemble deep learning model for stock prediction based on stock prices and news," *Int. J. Data Sci. Anal.*, vol. 13, no. 2, pp. 139–149, 2022, doi: 10.1007/s41060-021-00279-

9.
[43] V. Dogra, F. S. Alharithi, R. M. Álvarez, A. Singh, and A. M. Qahtani, “NLP-Based Application for Analyzing Private and Public Banks Stocks Reaction to News Events in the Indian Stock Exchange,” *Systems*, vol. 10, no. 6, 2022, doi: 10.3390/systems10060233.
[44] P. Paranjape-Voditel and U. Deshpande, “A stock market portfolio recommender system based on association rule mining,” *Appl. Soft Comput. J.*, vol. 13, no. 2, pp. 1055–1063, 2013, doi: 10.1016/j.asoc.2012.09.012.