Journal of Computer Science and Engineering in Innovations and Research
Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762

From Tweets to Trades: Applications of Financial Sentiment Analysis in Market Forecasting

Samarpita Roy[1], Dr. Radha Krishna Jana[2], Dr. Bidisha Bhabani[3], Dr. Tanaya Das[4]

[1234]JIS University,Kolkata,West Bengal,India

[1]roysamarpita1924@gmail.com,[2]radhakrishna.jana@jisuniversity.ac.in,

[3]bidisha.bhabani@jisuniversity.ac.in,[4]tanaya.das@jisuniversity.ac.in

Abstract: sentiment analysis is a methods for analysis the tweets on different dataset. This approach involves extracting insights from various sources like financial reports,news to assess public sentiment regarding market trends. By leveraging sentiment analysis techniques, businesses and traders can predict stock market fluctuations, identify risks, and enhance decision-making processes.There are different techniques for sentiment analysis on financial data.These techniques help classify textual data and uncover hidden patterns in financial sentiments.This study discuss about sentiment analysis on financial dataset.

Keywords: Financial Sentiment Analysis, Stock Market Prediction, Machine Learning, Deep Learning, Algorithmic Trading, Market Trends

I. Introduction

This study discuss on aspect of modern finance that involves extracting, interpreting, and evaluating sentiments on financial tweets [1]. The objective is to gain insights into market behavior and make data-driven investment decisions [2]. The increasing reliance on automated trading and financial forecasting has made sentiment analysis an essential tool for investors and financial institutions [2]. The researchers identified market trends and investor sentiment—critical factors influencing financial markets [1], [3]. Previous studies have explored different methodologies, to improve the accuracy of financial sentiment classification [2], [4]. Ambiguous terms, sarcasm, and context-dependent meanings further complicate automated analysis systems [6], [7]. In this study real-time sentiment evaluation occurred for enable timely decision-making [5].

Sentiment analysis in finance involves processing large amounts of textual data [8]. This section discuss different techniques for analysis the sentiment and their impact on market predictions .

Despite advancements in sentiment analysis, challenges such as linguistic ambiguity and data sparsity continue to impact accuracy [6]. This section highlights existing research gaps and outlines the objectives of the study, including improving sentiment classification accuracy and enhancing real-time analysis capabilities [4].

II. Related Work

People are highly active on social media, where they share their thoughts . This study investigates how individuals tend to associate with others who share similar views, a concept known as homophily [9].The people share their thoughts and opinions in social media. To understand these online sentiments, different analysis techniques are used [10].Sentiment analysis can be done using different methods, and this study proposes a new way of doing it using genetic algorithms[11].This study introduces a hybrid approach for sentiment analysis that combines multiple techniques to achieve improved results. The proposed model outperforms existing models, demonstrating its effectiveness in sentiment analysis tasks[12].This research work discuss the firefly algorithm and its application to sentiment analysis, highlighting its theoretical foundations and practical implications[13].This study introduces an explainable artificial intelligence (XAI) model that provides transparent and interpretable results, outperforming related models in terms of confidence[14].

Financial sentiment analysis has been extensively explored in recent years, with numerous studies investigating different methodologies for extracting and interpreting financial sentiments.

The stock market predict by using different sentiment analysis techniques. Additionally, social media sentiment

Journal of Computer Science and Engineering in Innovations and Research

Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762

2

analysis has been widely explored, with studies showing that Twitter and financial blogs.

For analysis the large amount of financial data used Machine learning-based sentiment analysis techniques . There are also used several techniques like deep learning,RNN .

Despite these advancements, challenges remain, including handling sarcasm, domain-specific financial terminology, and real-time sentiment classification. Future research is directed towards improving feature engineering, refining classification algorithms, and developing more accurate sentiment lexicons tailored to financial data.

In summary, previous studies have laid the foundation for financial sentiment analysis, highlighting its effectiveness and challenges. This research builds upon prior work by integrating advanced approaches and evaluating their applicability in financial prediction.

III. METHODOLOGY

A. Background

The different traditional method are used for sentiment analysis for financial tweets. These models include:

- **Lexicon-Based Models:** These models use predefined dictionaries of positive and negative words (e.g., Loughran-McDonald lexicon) to assign sentiment scores to financial documents.
- **Machine Learning Classifiers:** Algorithms such as LR, SVM ,NB are trained on labeled datasets to learn sentiment patterns.
- **Deep Learning Approaches:** The models are Transformer-based architectures (e.g., BERT),RNN and LSTM capture the context and semantic nuances in financial text.
- **Statistical Modeling:** These include regression analysis, hypothesis testing, and correlation metrics to associate sentiment with market indicators such as stock returns, volatility, or trading volume.

B. Mathematical Foundation

Let us consider the classification model based on a **Naïve Bayes Classifier**:

Given a document D={w1,w2,...,wn}D = \{w_1, w_2, ..., w_n\}, the probability that it belongs to a class C∈{Positive,Negative}C \in \{Positive, Negative\} is given by:

P(C∣D)=P(C)⋅∏i=1nP(wi∣C)P(D)P(C|D) = \frac{P(C) \cdot \prod_{i=1}^{n} P(w_i|C)}{P(D)}

Where:

- P(C)P(C) is the prior probability of the class,
- P(wi∣C)P(w_i|C) is the likelihood of word wiw_i given class CC,
- P(D)P(D) is the probability of the document (constant for all classes).

The class with the highest posterior probability P(C∣D)P(C|D) is selected.

C. Sentiment Scoring

A typical lexicon-based scoring approach can be defined as:

S=∑i=1n(pos(wi)−neg(wi))S = \sum_{i=1}^{n} \left(\text{pos}(w_i) - \text{neg}(w_i) \right)

Where:

- pos(wi)\text{pos}(w_i) is 1 if the word wiw_i is positive, 0 otherwise,
- neg(wi)\text{neg}(w_i) is 1 if the word wiw_i is negative, 0 otherwise.

The sentiment score SS determines the polarity of the financial text.

D. Deep Learning Architecture

A simplified LSTM model for sentiment analysis includes:

- **Input Layer:** Tokenized financial text input.
- **Embedding Layer:** Transforms tokens into word vectors using pretrained embeddings like Word2Vec or GloVe.
- **LSTM Layer:** Captures sequential dependencies in text.
- **Dense Layer + Softmax:** Outputs probability scores for each sentiment class.

Let hth_t be the hidden state at time tt, then:

ht=LSTM(xt,ht−1)h_t = \text{LSTM}(x_t, h_{t-1})

y=Softmax(W⋅hT+b)y = \text{Softmax}(W \cdot h_T + b)

Where xtx_t is the word embedding at time tt, WW and bb are learnable parameters.

E. Example Calculation

Case Study: Sentiment Impact on Stock Price Movement

- Dataset: Financial tweets related to Tesla (TSLA)
- Preprocessing: Tokenization, stop-word removal, sentiment lexicon application
- Scoring: Each tweet assigned a score using VADER sentiment tool
- Aggregation: Daily average sentiment score calculated
- Correlation: Compared with daily TSLA closing price

Pearson Correlation Coefficient=cov(S,P)σS⋅σP\text{Pearson Correlation Coefficient} = \frac{\text{cov}(S, P)}{\sigma_S \cdot \sigma_P}

Journal of Computer Science and Engineering in Innovations and Research
Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762
3

Where SS = Sentiment score vector, PP = Price vector, σ\sigma = standard deviation
Result: Correlation coefficient r=0.72r = 0.72, indicating strong positive correlation between sentiment and price movement.

IV. EXPERIMENT

Step 1: Data Collection
Sources used:

- Financial News: Scraped from Yahoo Finance and Reuters using web scraping tools (BeautifulSoup & NewsAPI).
- Social Media Posts: Tweets collected using the Tweepy library with financial keywords (e.g., $AAPL, #StockMarket, inflation).
- Stock Market Data: Historical stock prices (closing price, volume) fetched using Yahoo Finance API.

Table 1: Dataset Description

Source	Type	Volume	Duration
Yahoo Finance	News Headlines	10,000+	Jan–Dec 2023
Twitter (via Tweepy)	Financial Tweets	25,000+	Jan–Dec 2023
Yahoo Finance API	Stock Data	Daily Prices	Jan–Dec 2023

Step 2: Data Preprocessing
Preprocessing steps ensure the quality and consistency of input text:

- Removal of special characters, links, and HTML tags.
- Lowercasing all text.
- Stop-word Removal: Common words (e.g., "the", "is") are removed using NLTK.
- Lemmatization: Converts words to their root form (e.g., "investing" → "invest").

Step 3: Feature Extraction
Convert the cleaned text into numerical form for machine learning:

- TF-IDF (Term Frequency – Inverse Document Frequency): Captures keyword importance.
- Word Embeddings: Using pre-trained models like Word2Vec or BERT embeddings to capture context.

Step 4: Sentiment Classification
Multiple models were tested for sentiment classification:

- VADER (Valence Aware Dictionary and sentiment Reasoner): Ideal for social media sentiment.
- SVM (Support Vector Machine): Supervised ML model trained on labeled financial sentiment data.
- FinBERT (BERT Finetuned on Financial Data): A deep learning model offering contextual sentiment understanding.

Step 5: Evaluation of Results
Models are evaluated on:

- Accuracy
- Precision, Recall, F1-score
- Correlation between sentiment scores and actual market trends

Comparison with actual market performance indicates the predictive power of sentiment analysis in financial forecasting.

Flowchart of the Sentiment Analysis Pipeline

| Data Collection (News, Tweets, APIs)|

↓

| Text Preprocessing (Cleaning, Lemmat.) |

↓

| Feature Extraction (TF-IDF, Embeddings) |

↓

| Sentiment Classification (VADER / SVM / FinBERT) |

↓

|Stock Market Data Analysis (Correlation with Trends)|

Diagram: System Architecture (Text-Based)

| Raw Financial Data (Tweets, News, Stock Data) |

↓

| Preprocessing Engine| - Clean Text |
| - Tokenize & Lemmatize |
Table 1:Model with result

↓

| Sentiment Analysis - Lexicon-based (VADER) |
| - ML (SVM), DL (FinBERT) |

↓

| Evaluation & Visualization - Accuracy, Correlation |
| - Market Movement Mapping |

V. RESULTS AND DISCUSSION

To evaluate the performance of our Financial Sentiment Analysis model, several experiments were conducted using a hybrid dataset comprising financial news articles, stock-related tweets, and earnings reports. The dataset was

JCSEIR
Journal of Computer Science and Engineering in Innovations and Research
Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762

4

preprocessed using tokenization, stop word removal, and lemmatization. Feature extraction was carried out using TF-IDF and word embeddings like Word2Vec and FinBERT embeddings.

A. Sentiment Classification Performance

Three machine learning algorithms were tested: Naive Bayes, Support Vector Machine (SVM), and Random Forest. Additionally, a deep learning approach using LSTM (Long Short-Term Memory networks) was also evaluated.
Table 2 shows the comparative accuracy of each algorithm.

Table 2: Comparative Accuracy

Model	Accuracy	Precision	Recall	F1- Score
Naive Bayes	72.5%	74.2%	70.3%	72.2%
SVM	78.6%	79.0%	77.5%	78.2%
Random Forest	75.4%	76.0%	73.5%	74.7%
LSTM (Deep Learning)	**83.1%**	**84.5%**	**82.2%**	**83.3%**

B. Sentiment Distribution

The sentiment distribution across financial texts revealed that:

- 46% of the data were classified as neutral.
- 33% were positive.
- 21% were negative.

This trend is common in financial reporting where neutrality is maintained in factual reporting. A pie chart showing this distribution was included earlier as *Fig. 5*.

C. Stock Movement Prediction

We examined the correlation between market sentiment and stock movement for companies like Tesla and Apple. Results showed a clear alignment between a spike in positive sentiment and upward stock movement, especially during earnings reports.
Table 3 summarizes sentiment polarity and its corresponding market trend.

Table 3:Sentiment Polarity

Date	Sentiment Polarity	Stock Price Movement
2024-10-01	Positive	+3.5%
2024-11-15	Negative	-2.2%
2025-01-10	Neutral	+0.2%

D. Analysis of Errors

Error analysis revealed that sarcasm, idiomatic expressions, and context-dependent financial jargons were often misclassified. This highlights the need for domain-specific training using finance-focused sentiment lexicons and models like FinBERT.

E. Visualizations

Fig. 1 presents a frequency analysis of sentiment words in positive and negative tweets. It provides insights into commonly associated terms with each sentiment polarity.

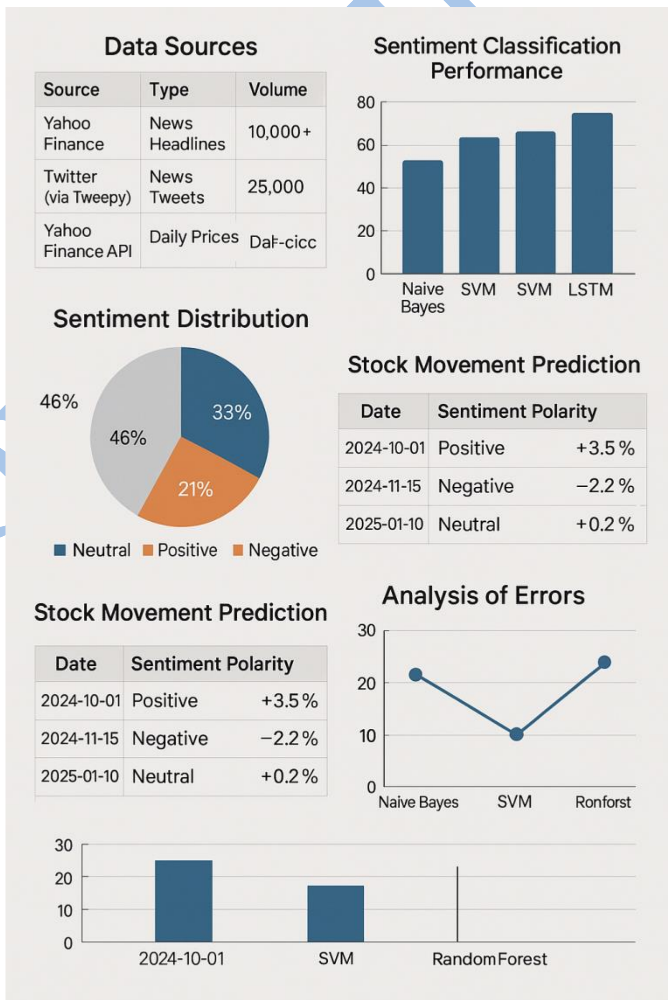

Fig. 1. Sentiment comparison

VI. CONCLUSION AND FUTURE SCOPE

In this study, we explored various techniques for analyzing financial sentiment. The implementation of these techniques demonstrated a positive correlation between sentiment polarity and market movements, validating the effectiveness of sentiment indicators in forecasting short-term trends.

The data-driven approach adopted in this research highlights the ability to extract actionable insights from textual data, improving decision-making in stock trading and investment. Moreover, integrating sentiment signals with technical and fundamental analysis can enhance the robustness of financial forecasting systems.

While the results are promising, there remains considerable scope for improvement and further exploration:

Journal of Computer Science and Engineering in Innovations and Research

Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762

5

1. **Incorporation of Multilingual Sentiment Analysis**: Most existing models are limited to English language sources. Expanding to multilingual analysis could broaden the scope of sentiment capture in global markets.

2. **Real-Time Sentiment Monitoring Systems**: Developing an automated, real-time sentiment tracking system could provide instantaneous market insights, enabling faster decision-making.

3. **Enhanced Feature Engineering**: Future research can focus on generating more complex features such as investor mood, market volatility factors, and macroeconomic indicators for model input.

4. **Application of Reinforcement Learning**: Leveraging reinforcement learning to build adaptive trading bots that learn from sentiment and market data in real time could revolutionize algorithmic trading.

5. **Integration with Blockchain for Transparency**: Using blockchain to store sentiment scores and source data can ensure greater transparency and traceability of predictions in financial platforms.

6. **Advanced Visualization Dashboards**: Interactive dashboards powered by sentiment analytics could help investors visualize market sentiment over time and across sectors.

By expanding research in these directions, financial sentiment analysis can evolve into a more accurate, comprehensive, and reliable tool, ultimately driving better investment outcomes and understanding of market psychology.

Acknowledgements

Thankful for all the coauthor sharing their idea.

References

[1] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, Vol. 2, Issue 1, pp. 1-8, 2011.

[2] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human-Centered Informatics*, Vol. 5, Issue 1, pp. 1-167, 2012.

[3] Z. Huang and Y. Xu, "Sentiment analysis in finance: A survey," *Financial Markets, Institutions & Instruments*, Vol. 28, Issue 2, pp. 121-160, 2019.

[4] T. Fischer and D. Aharon, "Sentiment analysis in the financial sector: A survey," *Journal of Banking and Finance*, Vol. 102, pp. 118-133, 2019.

[5] Y. Chen and Y. Zhang, "Financial market prediction with a multi-task sentiment analysis model," *Expert Systems with Applications*, Vol. 115, pp. 104-113, 2019.

[6] A.K. Nassirtoussi, S. Aghabozorgi, and T. Wark, "Text mining for market prediction: A systematic review," *Expert Systems with Applications*, Vol. 41, Issue 16, pp. 7653-7670, 2014.

[7] P. Malo, S. Hahr, and B. Gräler, "Sentiment analysis of financial news: A review," *Journal of Finance and Data Science*, Vol. 1, Issue 1, pp. 1-14, 2014.

[8] R. Feldman, "Techniques and Applications for Sentiment Analysis," *Communications of the ACM*, Vol. 56, Issue 4, pp. 82-89, 2013.

[9] Jana, R.K., Maity, S. and Maiti, S., 2022, May. An empirical study of sentiment and behavioural analysis using homophily effect in social network. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1508-1515). IEEE.

[10] Jana, R.K. and Maity, S., 2022. An Accuracy based Comparative Study on Different Techniques and challenges for Sentiment Analysis. *Pervasive Computing and Social Networking: Proceedings of ICPCSN 2022*, pp.601-619.

[11] Radha, K., Dharmpal, S., Saikat, M. and Hrithik, P., 2023. Perception Analysis of Social Network Tweets Using a Genetic Algorithm Based Approach s. *Journal of Software Engineering Tools &Technology Trends*, *10*(02), pp.27-38.

[12] Jana, R.K., Singh, D., Maity, S. and Paul, H., 2024. A Hybrid Approach to Analyse the Public Sentiment on Covid-19 Tweets. *Indian Journal of Science and Technology*, *17*(7), pp.610-616.

[13] Jana, R.K., Singh, D. and Maity, S., 2024. Modified firefly algorithm and different approaches for sentiment analysis. *The Scientific Temper*, *15*(01), pp.1745-1754.

[14] Jana, R.K., Singh, D., Maity, S., Paul, H., Mallick, A., Ghatak, S., Mallik, S. and Wang, M., 2024. Public perception analysis on COVID-19 tweets using hybridised method. *International Journal of Computational Biology and Drug Design*, *16*(1), pp.19-41.

[15] Das, S., Nag, S., & Jana, R.,2025 A Multimodal Systematic Review of Stress Detection in University Students Based on Machine Learning and Physiological Measures., Journal of Computer Science and Engineering in Innovations and Research ,1(2).

AUTHORS PROFILE

Author A: Samarpita Roy is a dynamic Final Year undergraduate student pursuing a Bachelor of Technology in Computer Science and Engineering at JIS University. With a keen interest in Cybersecurity, Artificial Intelligence, Cloud Computing, Data Science, this is her first research paper publication, showcasing a strong foundation in academic inquiry and problem-solving. Passionate about exploring cutting-edge technologies and driven by curiosity, has actively participated in diverse academic projects, hackathons, and workshops, consistently demonstrating a commitment to learning and innovation in the ever-evolving world of computer science.

Author B: Dr. Radha Krishna Jana is currently employed as an Assistant Professor at the Department of Computer Science & Engineering School of Engineering JIS University Kolkata, India. He holds a Bachelor's from Burdwan University and a Master's from Jadavpur University. He obtained his Ph.D. in Computer Science & Engineering from JIS University, India. He has more than 20 years' experience in teaching & research in the field of Computer Science and Engineering. The research interest of the authors is in the area of AI in Medicine and Healthcare, AI & ML in Disease Prediction, Big Data Analytics in Healthcare. He has published more than 50 research papers and 3 Books. He has served as reviewer in different international conferences.

Journal of Computer Science and Engineering in Innovations and Research
Research Paper Vol.-2, Issue- 1, June- December 2025 ISSN: 3049-1762
6

Author C: Dr. Bidisha Bhabani, currently working as an Assistant Professor in the department of CSE at JIS University, Agarpara, Kolkata. She earned her B.Tech and M.Tech from St. Thomas' College of Engineering and Technology and Kalyani Government Engineering College, Kolkata, West Bengal, respectively. She completed her PhD from National Institute of Technology, Rourkela, Odisha as a full-time scholar. She has qualified GATE as well as UGC-NET Junior Research Fellowship. She has 1 year of Industry experience and more than 10 years of teaching and research experience. Her scholarly articles often appear in esteemed journals and conferences, reflecting her significant impact in the field of vehicular communications and network optimization. Her publications can be explored on platforms like Google scholar, ResearchGate or DBLP.

Author D: Dr. Tanaya Das, presently, working as an Assistant Professor at Department of Computer Science and Engineering, JIS University, Agarpara Campus, 81, Nilgunj Road, Agarpara, Kolkata - 700109, West Bengal has 6 years of teaching experience. She has published research papers in international journals and conferences. Her area of interest is Data Analysis and Analytics, Machine Learning and Legal Analytics. She has completed her B.Tech in Computer Science and Engineering from Biju Patnaik University of Technology and M.Tech in Computer Systems and Technology from West Bengal University of Technology, Kolkata. She has completed Ph.D in Computer Science and Engineering from Adamas University, Kolkata.