Deep Learning Approaches for Financial Sentiment Prediction: Leveraging LSTM and Transformer Models

Kush Pandya[1], Gaurav Kulkarni[2], Vrunda Patel[3], Eralwala Abbas Mohammadbhai[4], Barkha Makwana[5], Dr. Mujtaba Ashraf Qureshi[6]
[1,2,3,4] *Dept. of Computer Science and Eng., Parul University, Vadodara, India*
[5,6] *Dept. of Artificial Intelligence and Data Science (AIDS), Parul University, Vadodara, India*
Assistant Professor

***Abstract*—The financial market is profoundly influenced by the sentiments expressed in news articles, economic reports, and financial analyses. Sentiment analysis of financial news provides critical insights for investors and traders, helping them anticipate market fluctuations and make informed decisions. Traditional methods for sentiment analysis often struggle with the complexity of financial language, which contains subtle nuances and context-dependent interpretations. To address these challenges, this re- search develops and implements a neural network based on long- short-term memory (LSTM) to classify the sentiment of financial news articles as positive or negative. The approach involves rigorous data preprocessing, model training, evaluation, and web deployment. The study achieves a training accuracy of 74.96%, demonstrating the potential of deep learning models to capture complex textual patterns. Although the model faced challenges with the accuracy of the validation, it provides a foundational approach for future advancements in financial sentiment analysis.**

***Index Terms*—Sentiment Analysis, LSTM, Financial Markets, Deep Learning, Natural Language Processing**

I. INTRODUCTION

The financial market operates under the constant influence of information dissemination. News articles, economic reports, and financial updates significantly shape market dynamics, influencing investor perceptions and trading behaviors[1]. Positive news can drive market optimism and price surges, while negative news can trigger sell-offs and market downturns.

Understanding the sentiment behind financial news is crucial for stakeholders [2],[3] looking to navigate these complexities. However, interpreting sentiment from financial texts is challenging due to the intricacy of financial jargon and the context-dependent nature of language. Traditional sentiment analysis techniques often fail to capture the depth and variability of financial narratives [4], leading to oversimplified or inaccurate interpretations.

In recent years, advancements in deep learning have opened new avenues for more sophisticated sentiment analysis. Specifically, long-short-term memory (LSTM) networks—an advanced type of Recurrent Neural Network (RNN)—have demonstrated remarkable capabilities in handling sequential data such as text [5]. Their ability to retain contextual information over long sequences makes them particularly effective for sentiment classification tasks. This study introduces an LSTM-based approach to classify financial news sentiments. Using comprehensive data pre-processing, robust model training, and web deployment, this research aims to enhance sentiment detection accuracy and contribute to more informed financial decision making.

A. Motivation

The motivation behind this research stems from the growing need for automated tools that can analyze financial news sentiment in real time. Traditional methods, such as manual analysis or lexicon-based approaches, are time-consuming and prone to human bias [3]. With the increasing volume of financial news generated daily, there is a pressing need for scalable and accurate sentiment analysis systems. LSTM networks provide an effective solution to this challenge due to their capability to retain and process long-term dependencies in sequential data.

Furthermore, recent studies have demonstrated that combining LSTM models with attention mechanisms improves sentiment classification accuracy by focusing on the most relevant sections of text [4]. This hybrid approach has been increasingly adopted in financial applications.

B. Challenges in Financial Sentiment Analysis

Financial sentiment analysis presents several unique challenges:

- Complex Language: Financial news often contains domain-specific jargon, abbreviations, and complex sentence structures, making it difficult for traditional NLP models to interpret accurately.
- Context Dependency: The meaning of a financial news article can change depending on its context. For instance, the term ”crash” often implies a negative outcome in financial markets but might carry a neutral or even positive meaning in different situations.
- Data Imbalance: Financial news data sets are often unbalanced, with a higher proportion of negative or positive sentiments, leading to biased model predictions [1]. Advanced data augmentation techniques and transfer learning have been proposed to address this issue [6].
- Noise in Data: Financial news data sets may contain irrelevant information, such as advertisements or non- financial news, which can negatively impact model performance [5].

II. LITERATURE REVIEW

Sentiment analysis, also known as opinion mining, has been a focal point of research in natural language processing (NLP) for over a decade [3]. Early methods relied heavily on lexicon-based approaches, where predefined sentiment dictionaries were used to classify text [1]. However, these methods struggled with context interpretation and failed to capture the nuanced expressions found in financial language. Subsequently, machine learning techniques such as Support Vector Machines (SVM), Na¨ıve Bayes, and Random Forests became popular. While these models improved sentiment classification by learning from data, they still struggled with understanding sequential dependencies [5], an essential aspect when analyzing complex financial texts .

To overcome these limitations, researchers began exploring deep learning techniques. Recurrent Neural Networks (RNNs) emerged as a promising solution due to their inherent ability to process sequential data. However, traditional RNNs suffered from vanishing gradient problems, limiting their effectiveness over long text sequences. LSTM networks, introduced by Hochreiter and Schmidhuber [5], addressed this challenge by introducing memory gates that retain information over extended sequences. Studies leveraging LSTM for sentiment analysis have shown improved accuracy, particularly in do- mains where context and sequence matter [3]. In financial sentiment analysis, LSTMs have proven effective in capturing complex relationships within texts [4]. However, challenges remain, particularly regarding data preprocessing, handling class imbalance, and avoiding overfitting. This study builds upon these findings, seeking to optimize LSTM performance for financial sentiment classification while addressing common challenges observed in prior research.

A. Recent Advances in Financial Sentiment Analysis

Recent advancements in NLP have introduced transformer- based models like BERT, GPT, and RoBERTa, which have shown remarkable performance in sentiment analysis tasks [6]. These models leverage self-attention mechanisms to capture contextual relationships between words, making them highly effective for financial sentiment analysis. However, their computational complexity and resource requirements make them less practical for real-time applications compared to LSTM networks. This research focuses on LSTM due to its balance between performance and computational efficiency.

III. METHODOLOGY

The methodology for this research is structured around the systematic preparation of data, model development, and evaluation. The dataset comprises financial news headlines collected between January and August 2010, with each head- line labeled as positive or negative based on its sentiment [7]. The initial phase involves extensive data preprocessing to ensure that the text data is clean, consistent, and suitable for training. This includes steps like removing non-alphabetic characters, converting text to lowercase, eliminating stop-words, and performing tokenization and padding to standardize input sequences for the LSTM model [5].

A. Dataset Description

The dataset, sourced from the YBI Foundation’s repository [7], consists of 4,101 entries with financial news headlines. Each entry includes the publication date, a binary sentiment label (0 for negative, 1 for positive), and up to 25 news headlines per day. The dataset is labeled, allowing clear classification of market sentiment. Further analysis can reveal whether markets were more frequently optimistic or pessimistic during this period.

B. Data Preprocessing

The preprocessing steps include:

- Data Cleaning: Removing non-alphabetic characters, converting text to lowercase, and removing unnecessary white spaces.
- Stopword Removal: Eliminating common but uninformative words (e.g., ”is,” ”the,” ”and”) to enhance the model’s focus on meaningful content.
- Tokenization: Breaking down text into individual words or units to transform raw data into a structured format that can be processed by the model.
- Padding: Standardizing the length of input sequences to ensure uniformity, which is essential for LSTM processing.
- Vectorization: Using embedding techniques to transform tokens into numerical vectors that capture semantic meanings.

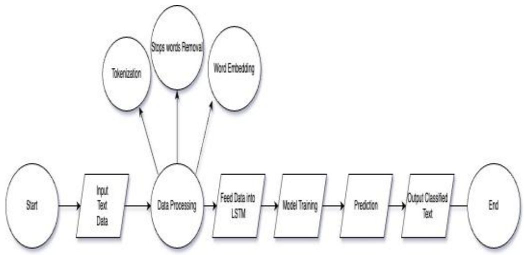

Fig. 1. Simplified Process Flow of Sentiment Analysis System

C. Model Architecture

The sentiment analysis model was constructed using a Long Short-Term Memory (LSTM) architecture within the Tensor Flow Keras framework. The architecture was chosen for its effectiveness in capturing sequential dependencies in textual data. The model consists of the following layers:

- Embedding Layer: Converts numerical sequences into dense vector representations of size 128, capturing se- mantic relationships between words.
- First LSTM Layer: Contains 128 units and is configured to return sequences, allowing the extraction of temporal features across multiple timesteps [4].
- Dropout Layer: To prevent overfitting, a dropout rate of 0.2 was applied, ensuring that some neurons are randomly deactivated during training.
- Second LSTM Layer: This layer consists of 64 units and is designed to capture more complex sequential dependencies in the data.
- Dense Layer: A fully connected layer with 32 neurons and ReLU activation is included to enhance pattern recognition and feature extraction.
- Output Layer: AA single neuron with a sigmoid activation function is used to classify the sentiment as either positive or negative.

D. Hyperparameter Tuning

To optimize the performance of the model, a hyperparameter tuning was conducted.

- Learning Rate: Adam’s optimizer was applied with a learning rate set to 0.001 to ensure stable and efficient weight updates.
- Batch Size: A batch size of 32 was selected to maintain a balance between the training speed and the accuracy of the model.
- Epochs: The model was trained for 10 epochs, allowing sufficient learning without leading to overfitting.

E. Model Evaluation

The model was evaluated using accuracy as the primary metric. The training process achieved a final accuracy of 74.96%, with validation accuracy hovering around 52.25%, indicating potential overfitting. The initial training phase began with an accuracy of 53.57%, progressively improving as the model identified deeper patterns in the data. Despite the growth in training accuracy, validation results showed fluctuations, with accuracy ranging between 52.25% and 53.96%.

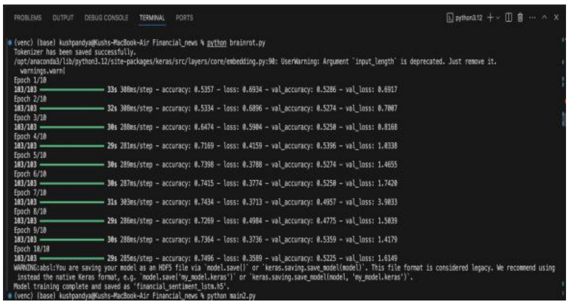

Fig. 2. Model Evaluation

IV. DISCUSSION

The results of this study highlight the potential of LSTM- based models for financial sentiment analysis. The model achieved a training accuracy of 74.96%, demonstrating its ability to learn complex patterns in textual data. However, the validation accuracy of 52.25% indicates that the model struggles with generalization, likely due to overfitting. This issue could be addressed by incorporating regularization techniques, expanding the dataset, or using more advanced architectures like transformers.

A. Overfitting and Generalization

One of the key challenges encountered in this research was overfitting, as indicated by the gap between training and validation accuracy. Overfitting occurs when a model performs exceptionally well on training data but struggles to generalize to unseen data. In this study, the model's high training accuracy and lower validation accuracy suggest that it may have memorized patterns from the training data rather than learning generalizable features [8, 9]. The high training accuracy and relatively lower validation accuracy observed suggest that the model may have memorized specific data points rather than developing a broader understanding of the patterns.

- Regularization Techniques: Implementing methods such as L2 regularization or early stopping can help prevent the model from becoming overly dependent on specific features in the training data.
- Data Augmentation: Increasing the diversity of the dataset by introducing synthetic samples or incorporate ing additional financial news articles may enhance the model's ability to generalize.
- Advanced Model Architectures: Transformer-based models like BERT or GPT, which leverage self-attention mechanisms, could provide better generalization capabilities compared to LSTM [3].

B. Implications for Financial Decision-Making

The ability to accurately analyze financial news sentiment has significant implications for traders, investors, and financial analysts [1]. By providing real-time sentiment insights, the proposed model can assist in:

- Market Forecasting: Predicting market trends based on sentiment shifts in financial news.
- Risk Assessment: Identifying potential risks by analyzing negative sentiment in news articles.
- Algorithmic Trading: Integrating sentiment analysis into automated trading algorithms to make data-driven decisions.

C. Limitations

Despite its potential, the model has several limitations:

- Binary Sentiment Classification: The model only classifies sentiment as positive or negative, ignoring neutral sentiments, which are common in financial news.
- Contextual Understanding: The model may struggle with understanding the context of financial jargon, leading to misclassifications.
- Data Imbalance: The dataset may be imbalanced, with a higher proportion of negative or positive sentiments, leading to biased predictions.

V. CONCLUSION

This research demonstrates the feasibility of using LSTM- based neural networks for financial sentiment analysis. The model achieved a training accuracy of 74.96%, illustrating its ability to learn complex patterns in textual data. However, the validation accuracy of 52.25% indicates challenges with generalization, likely due to overfitting and data imbalance. Despite these limitations, the study highlights the potential of deep learning models in capturing nuanced financial sentiment, which is critical for understanding market dynamics and aiding decision-making processes.

The study emphasizes the importance of thorough data pre- processing, including text cleaning, tokenization, and padding, to ensure the quality of input data for the LSTM model. The use of dropout layers and hyperparameter tuning further enhanced the model's performance, although overfitting remains a significant challenge. The deployment of the model through a Flask API demonstrates its practical applicability, enabling real-time sentiment analysis of financial news articles.

Overall, this research contributes to the growing field of financial NLP by providing a scalable and automated approach to sentiment analysis. While the results are promising, there is significant room for improvement, particularly in addressing overfitting and enhancing the model's ability to generalize to unseen data. Future work should focus on expanding the dataset, incorporating advanced architectures, and integrating additional contextual factors to improve the model's accuracy and robustness.

VI. FUTURE WORK

Future research could integrate sentiment analysis with predictive business models to improve financial forecasting [10]. To enhance the performance and applicability of the model, the following areas are suggested for future research:

A. Incorporating Advanced Pre-trained Models

Leveraging pre-trained transformer models such as BERT, RoBERTa, or GPT for financial sentiment analysis [11],[6] could improve the model's ability to understand context and complex language structures. These models are designed to capture nuanced relationships in text and could potentially out- perform traditional LSTM architectures. For example, BERT's bidirectional attention mechanism allows it to consider the context of words from both directions, making it highly effective for sentiment analysis tasks [6].

B. Expanding the Dataset

A larger and more diverse data set would allow the model to better generalize to unseen data. Future research could include data from various financial domains, such as global stock markets, cryptocurrency news, and economic reports, to enhance the model's versatility [5]. In addition, incorporating labeled datasets with neutral sentiments would allow for a more comprehensive sentiment analysis.

C. Contextual Analysis of News

Financial sentiment is often influenced by the context in which the news is presented. Incorporating external factors, such as historical stock prices or market indices, along with the news text, could improve sentiment predictions by providing a holistic view of the financial environment. For example, com- bining sentiment analysis with technical indicators like moving averages or RSI (Relative Strength Index) could enhance the model's predictive [2] capabilities.

D. Hybrid Models

Combining LSTM architectures with other machine learning techniques, such as convolutional neural networks (CNN) for feature extraction or attention mechanisms for context prioritization, could further improve the accuracy and robustness of the model. Hybrid models that integrate the strengths of multiple architectures could provide a more comprehensive solution for financial sentiment analysis [6].

E. Sentiment-Driven Market Predictions

Extending the model to predict market movements based on sentiment trends in financial news could open up practical applications for investors and market analysts. Integration with financial trading algorithms could lead to innovative tools for market analysis [5]. For example, the model could be used to develop sentiment-based trading strategies that automatically execute trades based on predicted market trends.

F. Real-Time Analysis

Enhancing the web API for real-time sentiment analysis of live financial news could provide immediate insights for traders and investors. Real-time analysis would require optimizing the model for faster inference and integrating it with live data feeds from financial news sources.

G. Ethical Considerations

Future research should also address the ethical implications of using AI for financial sentiment analysis. This includes ensuring that the model is free from biases, transparent in its decision-making process, and used responsibly to avoid market manipulation. Ethical considerations are particularly important in the financial domain, where automated systems can have significant economic impacts [12].

REFERENCES

[1] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[2] M. Day and C. Lee, "Deep learning for financial sen- timent analysis on finance news providers," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language

understanding,” in *Proceedings of NAACL-HLT 2019*, 2019.

[4] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in *Proceedings of ICLR 2015*, 2015.

[5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] J. Howard and S. Ruder, “Universal language model fine- tuning for text classification,” in *Proceedings of ACL 2018*, 2018.

[7] Y. Foundation, “Dataset,” https://github.com/ YBI-Foundation/Dataset, n.d., accessed: 2025-03-20.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org

[9] C. M. Bishop, *Pattern Recognition and Machine Learn- ing*. Springer, 2006.

[10] M. A. Qureshi and M. I. Sheikh, “Operational business prediction system using machine learning algorithm,” *Design Engineering*, pp. 12 511–12 524, 2021.

[11] M. Ghayoumi, *Generative Adversarial Networks in Prac- tice*. CRC Press, 2023.

[12] P. Chopra, “Ethical implications of ai in financial ser- vices: Bias, transparency, and accountability,” *Interna- tional Journal of Scientific Research in Computer Sci- ence, Engineering and Information Technology*, vol. 10, pp. 306–314, October 2024.