IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Forecasting of Marketing News in Financial Sector for Sentiment Analysis Using LLMs (Large Language Models)

Sandeep Gupta
SATI, Vidisha
sandeepguptabashu@gmail.com

Abstract: *Sentiment Analysis is a process where a text is assigned whether it has a positive tone, a negative tone, or a neutral tone. In finance, sentiment analysis is used to derive a numerical signal out of financial text. In order to do financial textual analysis on a collection of financial texts that are a component of a larger dataset called the Financial Phrase Bank, this study presents a novel strong sentiment classification architecture based on an improved model of Bidirectional Encoder Representations from Transformers (BERT). The methodology involves all preprocessing requirements, which are text cleaning, BERT Word Piece tokenization, label encoding, and down-sampling to balance the classes. Common measures, such as accuracy of 95.29%, precision of 95.37%, recall of 95.24%, and f1-score of 95.32%, were used to train and evaluate the model. These metrics show that the model performs better than more traditional machine learning models like Random Forest. together with language models like GPT-4o. Visualization was used to explain the model behavior and generalization power; e.g., the visualization of a confounding array, training-validation curve plots, and a token distribution plot. The experimental findings justify the utility of the proposed BERT-based approach in retrieving subtle sentiment expression in elaborate financial narrative texts, thus providing scalable and effective accuracy to sentiment-based financial analytics*

Keywords: Large Language Models (LLMs), BERT Text Classification, Market News, Financial Phrase Bank Dataset, Sentiment Analysis

I. INTRODUCTION

Sentiment Analysis (SA) in the financial market has proved to be an important field of study because it is increasingly being used in fields like stock market prediction, investor behaviour, and policy impacts [1]. Financial sentiment analysis, or FSA, is the computer task of determining and assessing the subjective content of financial writings, such as news stories, reports, or market comments [2][3][4]. The process will help to comprehend market perceptions, investor sentiment, and how macroeconomic news may affect the market on asset prices and financial decisions.

The financial inclusion as a significant source of economic progress and social wellbeing is benefiting more and more in connection with the data-driven innovations, such as sentiment-aware finance [5]. It involves using and having access to appropriate financial products and services, particularly through digital channels, like credit, insurance, savings accounts, and remittance services [6] [7]. The development of financial inclusion as a fundamental developmental issue has only added speed to the number of people seeking to develop something meaningful out of financial communications [8].

Classical sentiment analysis solutions were based on statistical language models, i.e., n-grams, where a combination of word frequencies [9][10]is calculated as probabilities thereof. They did perform well in short contexts, but could not cope with sparse data and the curse of dimensionality, hence could not model complex financial narratives or model long-term dependencies [11][12]. Besides, domain specificity is inherent to sentiment analysis, where words that seem positive in general texts may be neutral or, worse, negative in financial documents, due to the presence of terminology and the lack of affective words.

DOI: 10.48175/IJARSCT-28347

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Understanding the polarity detection process has been made easier with the use of Financial Sentiment Analysis (FSA), identifying the positive and negative opinions spoken in the financial documents [13][14][15][16].In contrast to the conventional opinion analysis (BY binary or +/3 based), that is, aspect-based sentiment analysis offers a higher degree of granularity since it identifies an opinion on a specific piece of information, e.g., opinion on companies, opinions on sectors, opinions on instruments.

Large Language Models (LLMs) have revolutionized sentiment analysis by addressing the shortcomings of previous models. LLMs include Generative Pre-trained Transformers (GPT) and Bidirectional Encoder Representations from Transformers (BERT) [17][18][19], etc. have been proved to be more successful across different When compared to other deep pre-trained models, NLP tasks like sentiment analysis are able to extract extensive and deeper contextual and semantic information by unsupervised pretraining on large-scale organizations [20][21].Their capacity to generalize knowledge and understand context-sensitive patterns has significantly enhanced sentiment interpretation, especially in ambiguous or sarcasm-prone texts. If LLMs aren't adjusted with domain-specific data, they could still have trouble with numerical reasoning and context-specific financial semantics.

This paper examines the use of LLMs for financial sentiment analysis in light of recent advancements, focusing on news stories that affect market sentiment [22][23][24]. The emphasis is on using pretrained models that have been adjusted for the financial industry to improve sentiment detection accuracy, which will aid in market forecasting and investment decision-making.

A. Significance and Contribution

In order to influence market behaviour, direct investment strategies, and assist in risk management decisions, it is imperative that financial sentiment be accurately interpreted. Given the inherently complex and domain-specific nature of financial language, traditional sentiment analysis approaches often fall short in capturing nuanced expressions and contextual cues. This research addresses these challenges by leveraging the powerful language modelling capabilities of BERT, enabling a more refined understanding of sentiment in financial market news. By improving prediction accuracy and dependability, the suggested method greatly pushes the boundaries of financial sentiment research and is therefore extremely useful for algorithmic trading, financial journalism, and decision support systems. The following are the study's main contributions:

- Robust sentiment analysis framework is developed specifically for financial market news, utilizing the Financial Phrase Bank dataset.
- Comprehensive pre-processing techniques are applied, including HTML noise removal, BERT Word Piece tokenization, padding, truncation, label encoding, and down-sampling to ensure high-quality model input and balanced class distribution.
- A pre-trained BERT model tuned for financial sentiment classification successfully captures context-aware representations of financial content.
- A comprehensive performance analysis demonstrates that the proposed BERT-based model outperforms the Random Forest and GPT-4o baselines by a significant margin using standard metrics (accuracy, precision, recall, and F1-score).
- Visualization tools such as training/validation curves, confusion matrix, and token distributions are employed to provide deeper insights into model behaviour and generalization capability.

B. Justification and Novelty

This study's rationale stems from the urgent demand for precise and context-aware sentiment analysis in the financial domain, where minor shifts in textual sentiment can have significant market implications. The complexity of linguistic patterns of financial texts and the domain-specific terms that occur in them are often neglected by traditional machine learning models. To resolve this issue, the proposed work takes advantage of the BERT model, which is praised for having profound means of bi-directional encoding and language comprehension based on context. The innovation of the current work lies in the combination of BERT and specifically domain-applied pre-proceeding approaches of pre-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, July 2025 Impact Factor: 7.67

composing financial domain news, applying down sampling to the data to mitigate the effect of imbalanced classes, and experimenting with the Financial Phrase Bank dataset, which makes the work relevant and authentic. Moreover, a comprehensive comparative analysis with 2 existing superior models, namely, GPT-4o and Random Forest, demonstrates superior performance and flexibility of the offered method. This version of domain-specific pre-processing, advanced natural language processing modelling, and thorough performance analysis sets a new benchmark for financial sentiment analysis performance.

C. Structure of the paper

The study is set up as follows: While Section III outlines the techniques and materials used, Section II presents pertinent studies on Financial Sentiment Analysis of Market News. Section IV displays the experimental findings of the proposed system. Section V provides a summary of the investigation's findings and its conclusion.

II. LITERATURE OF REVIEW

This segment reviews key studies on the financial sentiment of market news. Table 1 summarizes each work's methods and datasets Financial Phrase Bank, news feeds), findings (performance, accuracy gains), and limitations/future directions (scalability, domain adaptation, prompt tuning).

Dwiyanti, Aquarini and Gunawan, (2025) era results indicate that Scenario 4 consistently delivers the highest prediction accuracy across different evaluation metrics, with the log metric evaluation's Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) readings were 0.007298 and 0.009555, respectively, 47.616867 and 36.52605 in the absolute metric evaluation, and 85.146387 and 70.34775 in the stock closing price evaluation. While sentiment integration shows potential, its success is scenario-dependent and influenced by hyperparameter tuning. This study highlights the value of sentiment analysis in enhancing stock price projections and establishes the framework for further investigation into sentiment-based predictive models in financial markets [25].

Rizki Wiratama, Aquarini, and Harry Gunawan (2025) indicate that sentiment integration consistently improves the model's ability to anticipate outcomes in comparison to using historical data alone. Out all the scenarios that were investigated, Scenarios 2, 4, and 5 performed the best, with an RMSE value of 0.009163 and an MAE value of 0.007432 utilizing the logarithmic return type. These findings suggest that incorporating sentiment features into predictive models can enhance stock price forecasting and demonstrate the use of NLP and LLMs in stock market analysis [26].

Dmonte, Ko, and Zampieri (2024). Techniques have varied from lexicon-based approaches to the use of DNN and transformer-based models since machine learning was used in financial sentiment research. LLMs have been used for a variety of NLP applications due to recent advancements in the field; nevertheless, the financial industry has not yet given these models much attention [27].

Dsouza et al. (2024) focused on the new technique for valuing stock options using Large Language Models (LLMs) and integrating qualitative data, like news titles, full text content, and publication dates, with quantitative data, like prices lagging behind financial features, moving averages, and volatility indicators. More precisely, their method combines quantitative information on stock options, including the closing price of stock options, with sentiment analysis from LLMs applied to financial news from two reliable sources (i.e., Yahoo Finance India and Economic Times) [28].

Yadav and Kumar (2024) created a unique dataset that includes market prices, Wikipedia, and news articles. Additionally, the study refines the LLaMa-2 model using advanced approaches like PEFT and LoRA and employees Fin BERT for sentiment analysis. The methodology encompassed data collection, integration, and model optimization. The findings illustrate how well the approach works in summarizing information and addressing new questions, showing its potential for financial investment strategies. The paper concludes by suggesting future enhancements and positions the Nifty Llama model as a notable step in applying LLMs to financial analysis in emerging markets, indicating a direction for future data-driven decision-making in various sectors [29].

Chen (2023) helps Word2Vec transform the input text data into a word vector representation utilizing the financial direction and contributes to the development of the financial market sentiment lexicon. Bi-LSTM is then used to extract the financial news text data's context information and salient features. Lastly, a Multi-head Self-Attention technique is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

included to improve the model's capacity to extract important characteristics. In the SAFN and SACFN datasets, WMSA-Bi-LSTM performs better than RNN, LSTM, and GRU series models, with accuracy of 82.19%, 74.93%, and F1 of 82.22%, 75.05%, respectively [30].

Sandeep et al. (2023) gather viewpoints from a variety of online sources, including news websites and social media. Next, predictions on the correlation between public opinion and stock market outcomes are made using Naive Bayes and other machine learning approaches. The proposed findings show that mood and market trends have good relationships, suggesting that sentiment research might be a helpful tool for forecasting stock market movements. For traders, financial experts, and policymakers with an interest in the market, this findings might have significant implications [31].

Kohsasih et al. (2022) intend to demonstrate an RNN-LSTM sentiment analysis model. The model employed for sentiment analysis of financial news achieved an F1 score of 91.61%, 92.23% precision, 91.54% accuracy, and 90.99% recall, according to the findings. The methodology they created aids in identifying trends and viewpoints when making financial and other investing decisions [32]

Table 1: Review of current research on Financial Sentiment Analysis of Market News

Author	Methods	Dataset	Key Findings	Limitations & Future Work
Dwiyanti, Aquarini, and Gunawan (2025)	Sentiment integration with predictive models; evaluation using RMSE & MAE	Log metric, absolute metric, and stock closing price data	Scenario 4 achieved highest accuracy (e.g., RMSE = 0.009555, MAE = 0.007298); sentiment analysis improves stock prediction	Sentiment effectiveness is scenario-dependent; it needs further exploration and tuning
Rizki Wiratama, Aquarini, and Gunawan (2025)	Historical data + Sentiment features; LLMs for stock prediction	Logarithmic return type	Scenarios 2, 4, and 5 performed best (e.g., RMSE = 0.009163); sentiment boosts accuracy.	Emphasizes role of NLP & LLMs; recommends more robust experimentation
Dmonte, Ko, and Zampieri (2024)	Machine Learning (from lexicon to transformers); LLMs for NLP	Kaggle	LLMs have potential but are underused in finance	Demands that LLMs be studied further in financial sentiment research.
Dsouza et al. (2024)	LLM combined with sentiment analysis of financial news and quantitative aspects	Economic Times, Yahoo Finance India + stock option data	Combining LLM-based sentiment with technical indicators enhances valuation models	Future work could expand data sources and improve fusion techniques
Yadav and Kumar (2024)	FinBERT, fine-tuned LLaMa-2 with PEFT & LoRA	Wikipedia, market prices, news articles	Effective at summarizing and answering queries; potential in financial strategies	Suggests future enhancements for the model and broader application in emerging markets
Chen (2023)	Word2Vec + Bi-LSTM + Multi-head Self-Attention	SAFN, SACFN datasets	Acc: 82.19%, 74.93%, F1: 82.22%, 75.05% were attained via WMSA-Bi-LSTM.	Outperforms RNN, LSTM, and GRU models; further improvements possible with attention tuning
Sandeep et al. (2023)	Naive Bayes and ML on social media/news	Online sources (social	Found a strong correlation between public mood and	Recommends refinement for use by traders, analysts,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

	sentiment	media, news)	market shifts	and policymakers
Kohsasih et al. (2022)	RNN-LSTM for financial sentiment analysis	News about finances	90.99% recall, 91.61% F1, 92.23% precision, and 91.54% accuracy	Validates usefulness in trend/opinion tracking; can aid investment decisions

III. METHODOLOGY

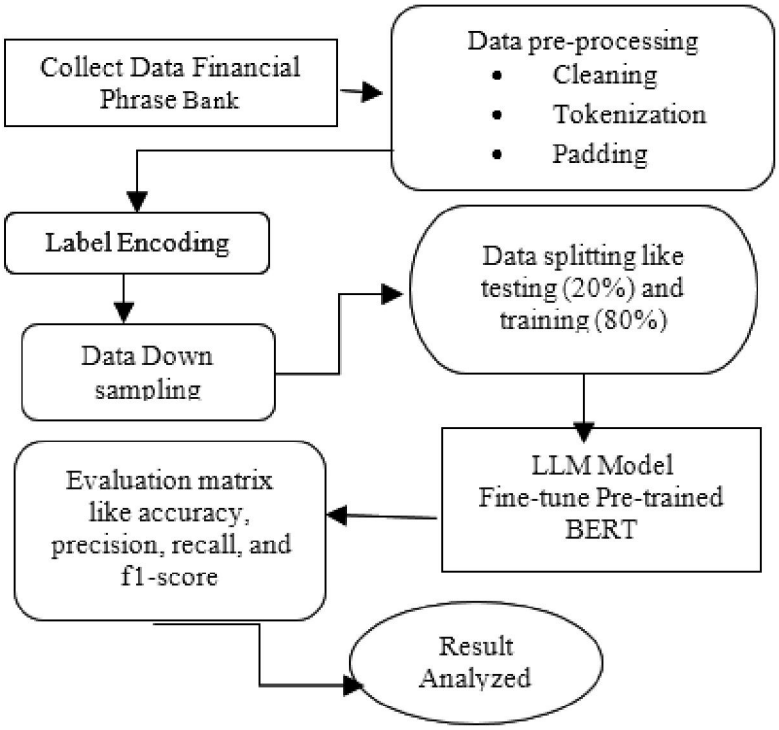

Fig. 1. Financial Flowchart Sentiment Analysis of Market News

A structured pipeline for conducting sentiment analysis in financial market news is demonstrated by the technique. Steps commence with gathering the data on benchmark sources that include the datasets of the Financial Phrase Bank. Raw text is then cleaned to get rid of noise, including special characters and HTML tags. Such sentiment labels (e.g., positive, neutral, negative) are then transformed into numbers, making use of label encodings. To address class imbalance, where there may be many instances of neutral or positive classes, down-sampling is applied to achieve equal representation of classes. In order to guarantee that the standard input to the model is fixed, the pre-processing step additionally entails tokenizing the text using BERT's Word-Piece tokenizer and either padding or truncating it. The data is then separated into training (80%) and testing (20%) categories to allow for reliable model testing. To understand contextual sentiment in financial writing, a pre-trained BERT model is improved on the training data using its bidirectional transformer architecture. The model is assessed using common measures like as accuracy, precision, recall, and F1-score to provide a complete view of performance. The flowchart for financial sentiment analysis of market news is shown in Figure 1.

The subsequent actions of the proposed methodology are briefly discussed below:

A. Data Collection

For sentiment analysis, the data were gathered using a financial dataset comprising 5,842 test samples from the Financial Phrase Bank. This dataset's sentiment labels are as follows: negative (860), positive (1852), and neutral (3130). Essential Python technologies such as Porch, scikit-learn, NumPy, TensorFlow, and Kera's were used to handle

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

this dataset effectively.

These samples originated from financial news, question-answering entries, and expert annotations, ensuring a diverse representation of sentiment-driven financial text to facilitate consistent and reliable modelling.

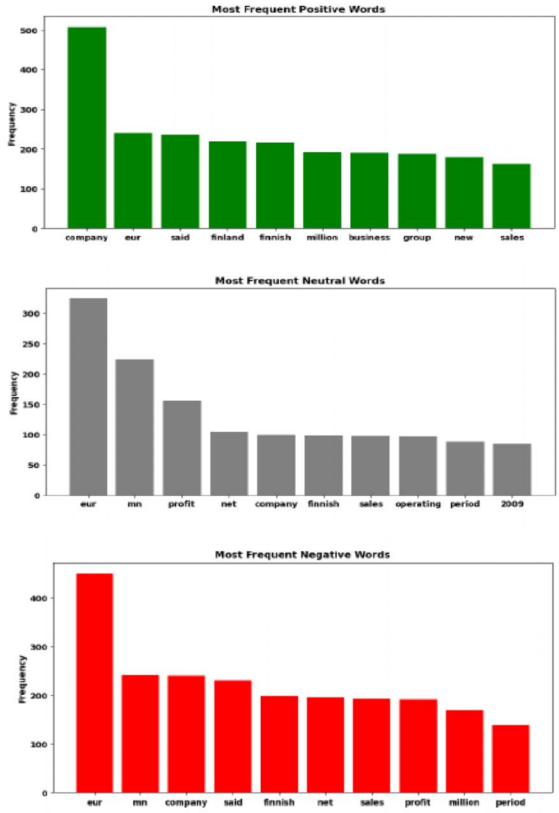

Fig. 2. Frequency Distribution of Positive, Neutral, Negative Words of Financial Phrase Bank dataset

A comparison of the most common terms by sentiment positive, neutral, and negative is shown in Figure 2. This comparison was probably taken from a textual dataset, such news stories or financial reports. The "Positive" category includes terms like *company*, *business*, and *new*, suggesting an optimistic tone. The "Neutral" and "Negative" categories share overlapping words (*our*, *company*, *sales*), but the latter may imply adverse contexts when paired with terms like *profit* and *million*, hinting at financial downturns. This visualization highlights how identical words (e.g., *finish*, *sales*) can appear across sentiments, highlighting how crucial context is to sentiment analysis.

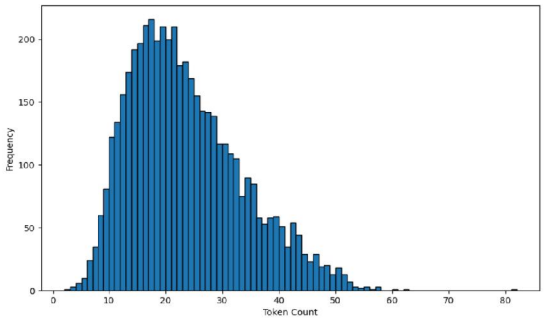

Fig. 3. Token Allocation of the Financial Sentiment Analysis

DOI: 10.48175/IJARSCT-28347

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

A histogram showing the frequency distribution of the token counts in the dataset is shown in Figure 3. The y-axis represents the frequency of occurrences, with values up to 200, while the x-axis shows the amount of tokens, which runs from 0 to 80. The bars indicate how often specific token counts appear, revealing the dataset's structural patterns. For instance, the highest frequency occurs at the lower end of the token count (near 0), suggesting that shorter texts (or tokens) are more common in the dataset. As the token count increases, the frequency sharply declines, with few or no instances beyond a certain point (e.g., 40–50 tokens). This distribution implies that the dataset is skewed toward shorter texts.

B. Data Preprocessing

The pre-processing of the financial sentiment dataset constructed from Financial Phrase Bank sources was crucial to ensure textual quality and modelling effectiveness. The initial step involved cleaning, which removed HTML tags, special characters, and extraneous formatting to reduce noise. Next, tokenization was performed using BERT's word piece tokenizer, breaking down sentences into sub word units compatible with the model's input requirements, followed by padding and truncation to maintain a uniform input length. The pre-processing steps are discussed below:

- **Clean text data:** Stripped HTML tags, URLs, punctuation, special characters, and excessive whitespace to reduce noise and ensure uniformity in text inputs.
- **Tokenization with BERT Word Piece Tokenizer:** BERT utilizes the Word Piece tokenizer to segment input text into tokens, effectively handling complex or unknown words by decomposing them into subworld units. Each input sequence, consisting of either a One or two sentences are transformed into a structured array of tokens, with a separator token (SEP) separating them and a specific categorization token (CLS) at the beginning.) [33]. A vocabulary of 30,000 terms is employed, and tokens are enriched with positional and segment embeddings to retain contextual and structural information. The combined embedding comprising token, position, and segment components is used to represent each input. This tokenized output is denoted as Y, which is subsequently passed to the feature extraction stage for downstream processing tasks.
- **Padding:** Standardized sequence lengths across the dataset by padding shorter texts and truncating longer ones to a fixed maximum length. This ensures efficient batch processing and maximizes GPU memory utilization.

C. Label Encoding

The kind of data used to feed the machine learning algorithms may be one of the numerous variables influencing the machine learning performance. Because it transforms data into categorical variables that machine learning models can understand, encoding categorical data is an essential procedure. The most popular method for handling categorical data is label encoding.

D. Down Sampling

It is employed to guarantee that classes are distributed fairly. lowering the number of descriptors for positive and neutral sentiment to 860 the same number as the test set's negative sentiment count ensures an equitable representation. This makes sentiment analysis more accurate by removing any bias towards a certain sentiment category. As a result, the model is trained on a large dataset, which improves its capacity to generate accurate predictions across a wide range of sentiment labels.

Figure 4 illustrates the sentiment classification distribution before and after applying down sampling. Initially, the dataset exhibited class imbalance, with 1852 positive, 860 negative, and 3130 neutral instances. After down sampling, all three sentiment classes, positive, negative, and neutral, were uniformly reduced to 860 samples each. This balancing technique ensures equal class representation, thereby mitigating model bias toward majority classes and improving the overall reliability and fairness of sentiment classification.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

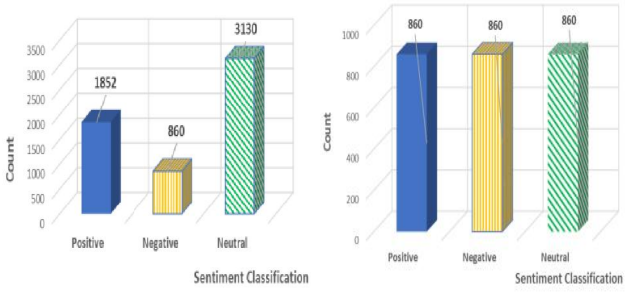

Fig. 4. Sentiment Classification Distribution (before and after) with Down sampling

E. Data Splitting

In the Financial Phrase Bank, split up into two subgroups, 20% goes toward testing, and 80% goes toward training, with one designated for each purpose.

F. Classification with BERT Model

Regarding the BERT model, one powerful pre-trained language model is BERT [34]. Pre-training and fine-tuning are the two stages that make up its framework. Training the model on a sizable, unlabelled corpus is known as pre-training. The pre-trained parameters are used to initialize the model, then labelled data for particular tasks is used to modify every parameter.

The multi-layer bidirectional Transformer encoder model architecture of BERT's original implementation serves as the basis for A stack of N = 6 identical layers makes up this type of encoder. There are two sub-layers inside each of these tiers. A multi-head self-attention mechanism is the second, while the first is a straightforward position-wise entirely connected feedforward network. It implements a layer normalization after a residual connection around both sub-layers.

First, we need to clarify what multi-head self-attention is in terms of scaled dot product attention. It is defined as follows. Equation (1)

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \qquad (1)$$

where dk is the dimension of the Q and K matrices of values, Q is the matrix of queries, K is the matrix of keys, and V is the matrix. At the moment, multi-head attention is explained by Equation (2)

$$MultiHead(Q,K,V) = Concat(head_1, \ldots., head_h)W^o \ (2)$$

$head_i$ = Attention (QW_i^Q, KW_i^k, VW_i^v). Projecting the queries, keys, and values h times using various learned linear projections to dk, dk, and dv (dimension of the values matrix) is the foundation of multi-head attention. The attention function is then executed in parallel on each of these projected copies of the queries, keys, and values to generate dv-dimensional output values. The final values are then obtained by concatenating and projecting them. The keys, values, and questions all originate from the same location when self-attention is present.

The following traits are used by BERT to depict a As a series of tokens, one sentence, or two phrases (such as the question and answer pair) BERT uses Word Piece embeddings

Next Sentence Prediction (NSP) and Masked LM comprise pre-training. In the first, the "[MASK]" token is used to mask a random subset of the input tokens, and the masked tokens are then predicted. Given two sentences, A and Bathe second one is composed of 50% of the actual text that comes after A (designated as Is Next) and 50% of a sentence chosen at random from the corpus (marked as Not Next).

BERT may represent a range of downstream activities due to the Transformer's self-attention mechanism, and fine-tuning is simple. We just feed each task's exact inputs and outputs into BERT and modify each parameter.

G. Evaluation Metrics

Accuracy, F1 Score, Precision, and Recall are used to evaluate model performance in order to more accurately identify

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, July 2025 Impact Factor: 7.67

the precision of models. These measurements offer a thorough evaluation of how well every model does on tasks, for instance, classification or prediction, especially in large language processing tasks like sentiment analysis in Finance [35]. Below is an explanation of each metric and its significance, as well as how they are calculated and applied. The following is confusion matrix:

- **True Positive:** representing instances where the model correctly identifies a piece of market news as negative sentiment, i.e., the analysis flags negative sentiment and the true label is indeed negative.
- **False Positive,:** indicating the model labels a news item as adverse, even though it is actually neutral or positive sentiment. This represents a false.
- **False Negative (FN):** denoting the model fails to detect negative sentiment and instead classifies a truly negative article as neutral or positive, representing a missed prediction.
- **True Negative:** Scenarios where the model accurately classifies non-negative news (neutral or positive) as such, correctly identifying it as not negative. The performance metrics shown below:

Accuracy

Among all projections, the model's accuracy measures the proportion of accurate predictions. It's the easiest metric to understand. Calculated as Equation (3)

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \quad (3)$$

Accuracy is useful when the dataset is balanced, meaning that all categories (e.g., positive, neutral, negative sentiments) are equally represented.

Precision

The ratio of accurately predicted positive cases to all expected positive cases is known as precision. It measures how well the model predicts the positive results and is calculated as Equation (4)

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

When false positives come at a high cost, precision is especially beneficial. For instance, in Green Finance, when determining if an investment is environmentally sustainable (positive class), having high precision ensures that only truly sustainable investments are identified.

Recall

The recall of the model measures how well it can identify each positive case. It is calculated as Equation (5)

$$Recall = \frac{\text{TP}}{TP+FN} \quad (5)$$

In Finance, Recall makes sure the model records every possible positive instance (e.g., sustainable investments), even if it misclassifies some negative instances. High recall is crucial when a positive prediction (false negative) is more costly.

F1-score

The F1 Score balances memory and accuracy by taking the harmonic mean of the two criteria. It is especially helpful when you need to balance memory and accuracy. The formula is Equation (6).

$$F1 - Score = \frac{2(Precision*Recall)}{\text{Precision+Recall}} \quad (6)$$

When dealing with imbalanced datasets, The F1 score is very important in situations when a high accuracy or high recall by itself would not accurately represent the model's performance.

IV. RESULT ANALYSIS AND DISCUSSION

An Intel Core i5-equipped Windows 10 workstation was used for the experiments. processor operating at a base clock of 3.3 GHz, capable of boosting up to 3.7 GHz, and backed by 32 GB of RAM a configuration appropriate for fine-tuning transformer-based models. Evaluation of the BERT model's effectiveness in financial sentiment analysis using market data. A high accuracy of 95.29% obtained by the model specifies that the model has a considerable strength to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

classify sentiments successfully. It also showed a precision of 95.37 percent, which shows that it produces a low rate of false positives, whereas it had a recall of 95.24 percent, which serves as evidence of its efficacy in finding the instances of relevant sentiments. The model's balanced performance in terms of accuracy and recall is demonstrated by its 95.32 percent F1-score. The Table II shows the performance of BERT model for Sentiment analysis of market are given below:

Table 2: Performance of Bert model on the financial sentiment analysis of market

Performance Measures	BERT
Accuracy	95.29
Precision	95.37
Recall	95.24
F1-score	95.32

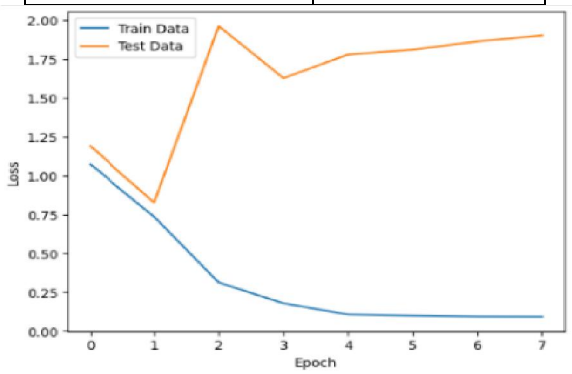

Fig. 5. BERT Model Training and Validation Loss

Figure 5 illustrates a typical loss curve where While validation loss first declines, training loss gradually declines, spikes at epoch 2, and then rises, indicating overfitting. The sharp fluctuation at epoch 2 suggests optimization instability, possibly due to a high learning rate or noisy validation data. The stable training loss versus erratic validation loss points to variance in a smaller validation set. This pattern implies the model is memorizing rather than learning. Mitigation strategies include early stopping, learning rate reduction, and regularization techniques like dropout or weight decay.

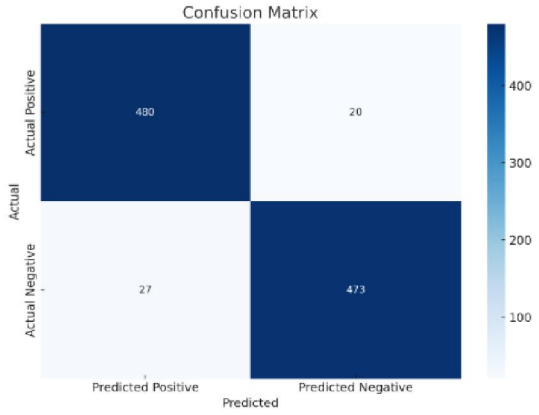

Fig. 6. Confusion Matrix of BERT Model

The efficacy of a classification model is graphically represented by the confusion matrix in Figure 6.. In this case, the model correctly predicted 480 actual positives as positive (True Positives) and 473 actual negatives as negative (True Negatives). However, it made some errors: 20 actual positives were incorrectly predicted as negative (False Negatives), and 27 actual negatives were wrongly predicted as positive (False Positives). The diagonal elements (480 and 473)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

represent correct predictions, indicating that the model performs very well on both classes.

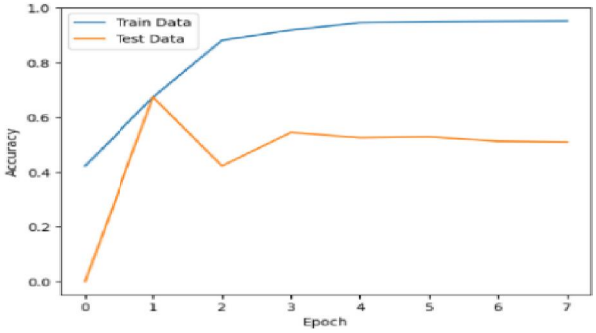

Fig. 7. BERT Model Accuracy Curve Training vs. Validation Over Epochs

Figure 7 shows the BERT model's accuracy trends on training and validation datasets over eight epochs. instruction accuracy (blue curve) shows a steady increase, reaching near-perfect accuracy after epoch 2 and remaining consistently high. Conversely, at first, the validation accuracy (orange curve) rises sharply, peaking at epoch 1, but then drops significantly at epoch 2, followed by a gradual decline and stagnation through subsequent periods. This difference between training and validation accuracy is an indication of overfitting when the model continues to improve on training data but fails to generalize to fresh validation data.

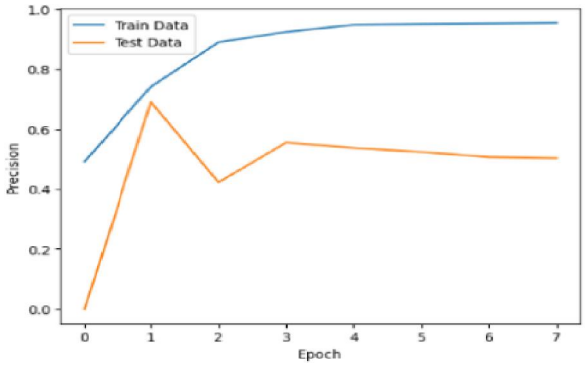

Fig. 8. BERT Model Precision curve of Training vs. Validation Across Epochs

Figure 8 illustrates the precision performance of the BERT model for both instruction and validation datasets across seven periods. The training precision begins at approximately 0.5 in epoch 0 and steadily increases, reaching around 0.95 by epoch 4, where it plateaus, indicating consistent improvement on the instruction data. Conversely the confirmation precision starts at 0, rises sharply to about 0.7 at epoch 1, but then declines to approximately 0.42 at epoch 2. Although it slightly recovers to 0.57 at epoch 3, it gradually decreases and stabilizes near 0.51 from epoch 5 onward.

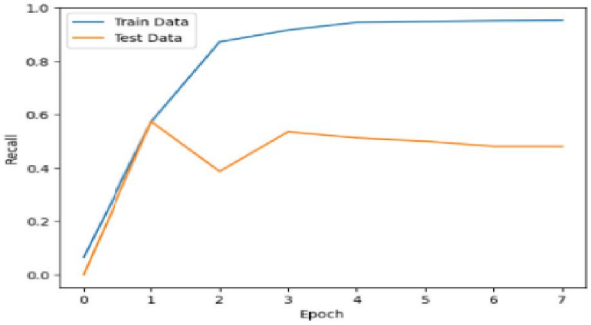

Fig. 9. BERT Model Recall Training vs. Validation Across Epochs

Figure 9 shows a simple line graph that plots the BERT models recall on training and test sets from epoch 0 through epoch 7. The vertical axis tracks recall values between 0.0 and 1.0, while the horizontal axis marks each training epoch.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

The training recall line climbs steadily, showing the model learns to spot positive examples more accurately as it sees more batches. By contrast, the test recall line wiggles and then flattens early, indicating that the model may begin overfitting and has trouble generalizing outside of the training data. The gap that opens between the two curves highlights the well-known balance between fitting the training set and performing reliably on unseen data.

Comparative Evaluations and Conversation

This section provides a side-by-side look at how well different models read the mood of market news using the Financial Phrase Bank dataset. Table III compares three approaches- Random Forest (RF), the large-language model GPT-4o, and the transformer BERT- on the task of reading market news and scoring its sentiment. Across every measure, BERT comes out on top, hitting 95.29% accuracy,95.37% precision, 95.24% recall and an F1 of 95.32%. GPT-4o lands in the middle, recording 86.00% for accuracy and precision, 84.00% for recall, and an F1 of 85.00%, which underscores its solid grasp of everyday language. RF, however, trails the pack, posting only 65.31% accuracy and showing that rule-based trees still struggle to untangle the nuance in financial headlines.

Table 3: ML and LLMs models comparison on the Financial Sentiment Analysis of Market News

Performance Measures	**BERT**	**GPT-4o**[36]	**RF**[37]
Accuracy	95.29	86.00	65.31
Precision	95.37	86.00	65.55
Recall	95.24	84.00	65.31
F1-score	95.32	85.00	65.31

The proposed BERT model achieves excellent results in financial sentiment analysis, with an accuracy of 95.29%. These results showcase BERT's skill in understanding and interpreting complex financial language thanks to its deep bidirectional transformer structure. The main benefit of the BERT model is its strong contextual understanding, allowing it to pick up subtle sentiment cues in market news. Because of this, it is very useful for financial applications, providing trustworthy sentiment categorization that is essential for data-driven investing and decision-making tactics.

V. CONCLUSION AND FUTURE WORK

Sentiment analysis has been a vital tool in recent years for understanding expert and public viewpoints in financial texts such as market news, reports, and commentary. The complex language of finance and the varied ways feelings are expressed create unique challenges that need effective natural language processing (NLP) techniques. Better techniques are required since traditional machine learning models frequently fail to capture these nuances. In this situation, transformer-based models like BERT present a helpful solution because of their capacity to comprehend context and their bidirectional encoding. This study shows how well a BERT-based framework works for the analysis of sentiment on monetary market news using the Phrase Bank dataset for monetary. With an impressive accuracy of 95.29%, the proposed model beat both traditional machine learning models like Random Forest and complex big language models like GPT-4o. A comparison of performance revealed BERT's excellent ability to understand context, which is especially important in financial news where small changes in language can greatly affect sentiment interpretation. Although overfitting was observed during training, the BERT model continued to show high F1-score, recall, and precision, confirming its strength. The findings support using transformer-based structures for financial sentiment analysis, showing great potential for improving data-driven investment decisions and financial forecasting systems.

Future work can investigate approaches including retrieval-augmented prompting, chain-of-thought reasoning, and multimodal integration to further improve accuracy, interpretability, and responsiveness in real-time financial sentiment systems.

REFERENCES

[1] C. Liu, A. Arulappan, R. Naha, A. Mahanti, J. Kamruzzaman, and I. H. Ra, "Large Language Models and Sentiment Analysis in Financial Markets: A Review, Datasets and Case Study," *IEEE Access*, pp. 1–21, 2024, doi: 10.1109/ACCESS.2024.3445413.

[2] D. Mishra, V. Kandpal, N. Agarwal, and B. Srivastava, "Financial Inclusion and Its Ripple Effects on Socio-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Economic Development: A Comprehensive Review," *J. Risk Financ. Manag.*, vol. 17, no. 3, 2024, doi: 10.3390/jrfm17030105.

[3] S. J. Wawge, "A Survey on the Identification of Credit Card Fraud Using Machine Learning with Precision , Performance , and Challenges," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, p. 8, 2025.

[4] G. Mantha, "Transforming the Insurance Industry with Salesforce: Enhancing Customer Engagement and Operational Efficiency," *North Am. J. Eng. Res.*, vol. 5, no. 3, 2024.

[5] Y. Mun and N. Kim, "Leveraging Large Language Models for Sentiment Analysis and Investment Strategy Development in Financial Markets," *J. Theor. Appl. Electron. Commer. Res.*, vol. 20, no. 2, p. 77, Apr. 2025, doi: 10.3390/jtaer20020077.

[6] P. Malo, A. Sinha, P. Takala, O. Ahlgren, and I. Lappalainen, "Learning the Roles of Directional Expressions and Domain Concepts in Financial News Analysis," in *2013 IEEE 13th International Conference on Data Mining Workshops*, IEEE, Dec. 2013. doi: 10.1109/ICDMW.2013.36.

[7] P. Chatterjee, "Proactive Infrastructure Reliability: AI-Powered Predictive Maintenance for Financial Ecosystem Resilience," *J. Artif. Intell. Gen. Sci. ISSN3006-4023*, vol. 7, no. 01, pp. 291–303, Dec. 2024, doi: 10.60087/jaigs.v7i01.358.

[8] J. Kumar Chaudhary, S. Tyagi, H. Prapan Sharma, S. Vaseem Akram, D. R. Sisodia, and D. Kapila, "Machine Learning Model-Based Financial Market Sentiment Prediction and Application," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2023, pp. 1456–1459. doi: 10.1109/ICACITE57410.2023.10183344.

[9] E. Memiş, H. Akarkamçı, M. Yeniad, J. Rahebi, and J. M. Lopez-Guede, "Comparative Study for Sentiment Analysis of Financial Tweets with Deep Learning Methods," *Appl. Sci.*, vol. 14, no. 2, 2024, doi: 10.3390/app14020588.

[10] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," pp. 101–110, 2025, doi: 10.48175/IJARSCT-25619.

[11] H. Reddy, B. Seshakagari, and A. Umashankar, "Dynamic Financial Sentiment Analysis and Market Forecasting through Large Language Models," pp. 397–410, 2025, doi: 10.5281/zenodo.15111609.

[12] N. Malali, "AI Ethics in Financial Services : A Global Perspective," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 2, 2025.

[13] B. Zhang, H. Yang, and X.-Y. Liu, "Instruct-FinGPT: Financial Sentiment Analysis by Instruction Tuning of General-Purpose Large Language Models," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4489831.

[14] S. Garg, "AI, Blockchain and Financial Services: Unlocking New Possibilities," *IJIRCT*, vol. 8, no. 1, 2022, doi: 10.5281/zenodo.15537568.

[15] B. Chaudhari, S. Chitraju, and G. Verma, "Achieving High-Speed Data Consistency in Financial Microservices Platforms Using NoSQL Using Nosql (Mongodb , Redis) Technologies," pp. 750–759, 2024, doi: 10.48175/IJARSCT-18890.

[16] S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, 2022, pp. 260–264. doi: 10.1109/IC3I56241.2022.10073077.

[17] M. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021. doi: 10.48550/arXiv.2103.11943.

[18] S. P. Kalava, "Revolutionizing Customer Experience: How CRM Digital Transformation Shapes Business," *Eur. J. Adv. Eng. Technol.*, p. 4, 2024.

[19] S. R. Singamsetty, "Deep Learning Ensemble Model for Fake News Detection"

[20] P. Chatterjee, "AI-Powered Payment Gateways : Accelerating Transactions and Fortifying Security in Real-Time Financial Systems," *Int. J. Sci. Res. Sci. Technol.*, 2023.

[21] B. Ramanujam, "Statistical in Sights in to Anti-Money Laundering : Analyzing Large-Scale Financial Transactions," vol. 14, no. 04, 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

[22] A. Upadhye, "Sentiment Analysis using Large Language Models: Methodologies, Applications, and Challenges," *Int. J. Comput. Appl.*, vol. 186, no. 20, pp. 30–34, 2024, doi: 10.5120/ijca2024923625.

[23] R. Kumar, "Evaluating and Enhancing Spatial Reasoning in Large Language Models," *ICIDA*, 2025.

[24] H. Kali, "Optimizing Credit Card Fraud Transactions identification and classification in banking industry Using Machine Learning Algorithms," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 85–96, 2024.

[25] W. Dwiyanti, N. Aquarini, and P. H. Gunawan, "The Effect of News Sentiment on Jakarta Composite Index Prediction Using Support Vector Regression Method," in *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, 2025, pp. 1–5. doi: 10.1109/ICADEIS65852.2025.10933431.

[26] M. Rizki Wiratama, N. Aquarini, and P. Harry Gunawan, "A Sentiment-Augmented Machine Learning Approach to Forecasting IHSG Prices Using XGBoost," in *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, 2025, pp. 1–6. doi: 10.1109/ICADEIS65852.2025.10933440.

[27] A. Dmonte, E. Ko, and M. Zampieri, "An Evaluation of Large Language Models in Financial Sentiment Analysis," in *2024 IEEE International Conference on Big Data (BigData)*, 2024, pp. 4869–4874. doi: 10.1109/BigData62323.2024.10825272.

[28] L. D. Dsouza, J. A. Nasir, M. M. Kamal, and L. Connolly, "Leveraging Large Language Models for Predicting Stock Option Valuation and Financial Risk Mitigation," in *2024 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2024, pp. 97–105. doi: 10.1109/ICDMW65004.2024.00019.

[29] V. Yadav and K. K. Kumar, "NiftyLLaMA: An open information extraction system integrated with sentiment analysis for Nifty-50 companies using LLaMa-2," in *2024 2nd International Conference on Foundation and Large Language Models (FLLM)*, 2024, pp. 280–283. doi: 10.1109/FLLM63129.2024.10852501.

[30] Y. Chen, "Financial News Sentiment Analysis Method Based on WMSA-Bi-LSTM," in *2023 4th International Conference on Intelligent Design（ICID）*, 2023, pp. 319–322. doi: 10.1109/ICID60307.2023.10396691.

[31] U. Sandeep, T. V. Vardhan, A. M. J, B. Yaswanth, and R. D. A, "Cracking the Code: Unleashing the Power of Sentiment Analysis & ML for Moroccan Stock Market Forecasting," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, 2023, pp. 1274–1278. doi: 10.1109/ICCES57224.2023.10192697.

[32] K. L. Kohsasih, B. H. Hayadi, Robet, C. Juliandy, O. Pribadi, and Andi, "Sentiment Analysis for Financial News Using RNN-LSTM Network," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 2022, pp. 1–6. doi: 10.1109/ICORIS56080.2022.10031595.

[33] S. Bandari, "BERT Tokenization and Hybrid-Optimized Deep Recurrent Neural Network for Hindi Document Summarization," *Int. J. Fuzzy Syst. Appl.*, vol. 11, no. 1, pp. 1–28, 2022, doi: 10.4018/IJFSA.313601.

[34] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, "Comparing BERT Against Traditional Machine Learning Models in Text Classification," *J. Comput. Cogn. Eng.*, vol. 2, no. 4, pp. 352–356, Apr. 2023, doi: 10.47852/bonviewJCCE3202838.

[35] T. Chen, "Sentiment Analysis in Green Finance with LLMs," vol. 0, pp. 1–9, 2024, doi: 10.54254/2754-1169/124/2024.MUR17868.

[36] Y. Shen and P. K. Zhang, "Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 2024, pp. 717–721. doi: 10.1109/ICPICS62053.2024.10796670.

[37] D. K. Nasiopoulos, K. I. Roumeliotis, D. P. Sakas, K. Toudas, and P. Reklitis, "Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models," *Int. J. Financ. Stud.*, vol. 13, no. 2, p. 75, May 2025, doi: 10.3390/ijfs13020075.