ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

Fine-Tuning of Bidirectional Encoder Representations from Transformers (BERT) for Sentiment Analysis with Reference to Financial News

Jyotirmoy Roy [1], Dr Bimal Debnath [2]

[1] Research Scholar, Department of Management, NEHU, Meghalaya, India.

[2] Assistant Professor, Department of Management, NEHU, Meghalaya, India.

Email- [1] jyotirmoyroy88@gmail.com, [2] bimal.dn@gmail.com

ABSTRACT:

In an era where financial markets are increasingly influenced by the rapid dissemination of information through various media, understanding market sentiment has become crucial for investors and financial analysts. Sentiment analysis, a field at the intersection of finance and natural language processing (NLP), offers valuable insights into market trends and investor behaviour. Recognizing the potential of advanced NLP in the finance domain, this study presents an adaptation of BERT for sentiment analysis within the stock market domain, achieving an overall accuracy of 77 percent. Through meticulous preprocessing—including tokenization, lemmatization, and standardization of text—combined with the advanced capabilities of BERT's pre-trained models, the research aimed to capture and classify market sentiment from textual data accurately. The fine-tuned model demonstrated a precision of 72 percent for negative sentiment detection and 80 percent for positive, with recall rates at 61 percent and 87 percent, respectively. The resulting F1-scores were 66 percent for negative and a robust 83 percent for positive sentiments, indicating a more reliable identification of positive over negative sentiment. These metrics affirm the model's effectiveness in discerning sentiment polarity and point to potential areas for enhancement, particularly in the Recall of negative sentiment. The findings underscore the significant promise of employing sophisticated NLP methodologies like BERT in financial sentiment analysis, which could be transformative for fields such as algorithmic trading and economic prediction by leveraging the subtle nuances captured through sentiment classification.

Keywords: Sentiment Analysis, BERT Model, Stock Market Text, Natural Language Processing, Data Preprocessing, Transformer-based Learning.

INTRODUCTION:

Sentiment analysis in finance has become an indispensable tool for interpreting market trends and guiding investment decisions. The affective states of market-related documents, such as news articles, reports, and social media posts, can provide insights into the collective mood of investors, potentially influencing stock prices (Bollen et al., 2011; Tetlock, 2007). NLP techniques are pivotal in extracting sentiment from such unstructured data, transforming

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022**

qualitative information into quantitative analysis (Loughran & McDonald, 2016). The advent of advanced NLP methods has significantly enhanced the precision of sentiment assessment, paving the way for more accurate market predictions.

The introduction of BERT by Devlin et al. (2018) marked a turning point in NLP. BERT's contextual analysis capabilities, stemming from its deep bidirectional training, make it uniquely suited for sentiment analysis. Unlike previous models that processed text in one direction, BERT considers the full context of a word by looking at the words that come before and after it, resulting in a nuanced understanding of language (Devlin et al., 2018).

Given the complex and often ambiguous nature of financial texts, the application of BERT to stock market sentiment analysis is particularly promising. This research aims to capitalize on BERT's advanced features to fine-tune a model capable of discerning sentiment with high accuracy in the financial domain. We hypothesize that BERT will outperform traditional models, providing deeper insights into market sentiment.

Our objectives are manifold: to implement BERT effectively for sentiment classification on stock market texts, to refine preprocessing techniques that cater to the specialized language of finance, and to evaluate the model's performance comprehensively. Through this research, we seek to contribute to the body of knowledge on applying cutting-edge NLP techniques in finance, demonstrating BERT's potential in enhancing the field of sentiment analysis.

The structure of this study is methodically designed to build a comprehensive understanding of sentiment analysis in the financial domain using advanced NLP techniques. We begin with a detailed introduction, setting the scene for the uninitiated reader and providing a succinct overview of the existing literature. This is followed by a granular look at our research methodology, where we delve into the dataset specifics, the preprocessing steps taken, and the intricacies of our BERT-based model selection and training. Subsequently, the system design is discussed, illustrating the architecture of our model and the computational considerations involved. We meticulously dissect the results to evaluate the effectiveness of our approach and conclude with a discussion that frames the significance of our findings within the broader context of financial sentiment analysis.

LITERATURE REVIEW:

The advent of computational linguistics has revolutionized how we process textual information, especially in sectors with high stakes, such as finance. In this section, we explore a body of work that establishes the bedrock of sentiment analysis, mainly focusing on how the sentiment conveyed in financial reports, news articles, and social media can be quantified and used to predict market movements. We shall traverse the landscape of prior studies, benchmarking their methodologies and outcomes against the capabilities of BERT to identify the gaps our research aims to fill.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022**

The role of sentiment analysis within financial domains has been extensively studied, with researchers acknowledging its significance in predicting market movements (Kearney & Liu, 2014). Prior literature has evidenced the predictive power of sentiment derived from financial news and social media on stock prices (Bollen et al., 2011; Tetlock, 2007). Sentiment analysis has been employed to forecast market volatility and trading volumes, emphasizing its practical implications in financial decision-making (Loughran & McDonald, 2016).

The advent of BERT and its subsequent applications across various fields have revolutionized how machines understand human language. In the healthcare sector, for instance, BERT has been utilized to extract medical information from unstructured text, showcasing significant improvements over previous models (Lee et al., 2020). Within the realm of law, BERT's application to legal document analysis has provided innovative approaches to information retrieval (Chalkidis & Kampas, 2019). These studies exemplify BERT's flexibility and (Bordoloi & Biswas, 2023)superior ability to comprehend context, which is essential for complex linguistic structures in specialized domains.

However, despite the successes of BERT in various disciplines, gaps remain in its application to financial sentiment analysis. Most existing research in financial NLP has relied on simpler models or earlier forms of deep learning architectures that do not fully exploit contextual language representation (Hu et al., 2018). Additionally, while BERT's potential has been noted, its implementation for sentiment analysis in finance is not yet widespread, with few studies investigating the fine-tuning of BERT on financial texts (Sousa et al., 2019). This study seeks to address these gaps by providing an in-depth analysis of BERT's capabilities in interpreting the nuanced language of financial texts, which often include jargon, idiomatic expressions, and complex syntactical structures that challenge traditional sentiment analysis models.

The current study is thus positioned at the intersection of advanced NLP techniques and financial sentiment analysis, aiming to contribute to the emerging body of research exploring transformer-based models' applicability in finance. By leveraging BERT's deep bidirectional approach, this research aspires to offer new insights and demonstrate the model's capacity to elevate sentiment analysis within the financial sector to a new level of precision and reliability.

METHODOLOGY:

This section provides a blueprint of the procedural approach undertaken to validate the effectiveness of BERT in financial sentiment analysis. We meticulously outline the stages of data collection, detailing the source and nature of our dataset. The subsequent preprocessing steps—tokenization, lemmatization, and noise reduction—are described, setting the stage for the model selection process. We will discuss our rationale for selecting BERT and the specifics of its configuration, training procedures, and the metrics used to evaluate its performance.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022**

3.1 Data Collection

The dataset for this study was retrieved from the publicly available "Stock Market Sentiment Dataset" on Kaggle (Chaudhary, 2020). The dataset comprises 5791 textual data derived from financial news headlines, tagged with binary sentiment labels indicating positive or negative sentiment. This dataset is representative of the real-world data financial analysts encounter, making it suitable for this research.

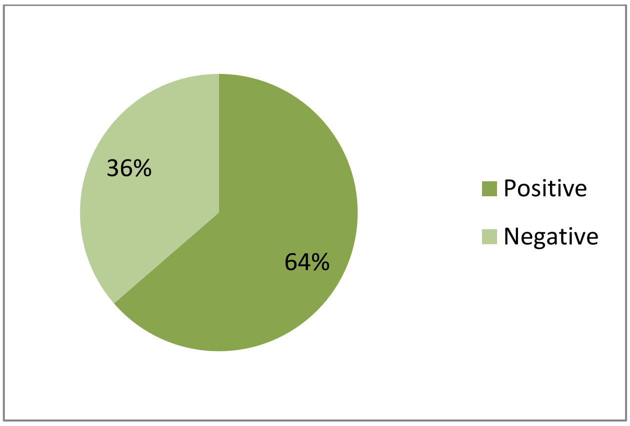

Figure 1: Composition of Sentiment in Dataset

Source: Author's own

Table 1: Sample of the News Dataset

Index	Text	Sentiment
1	AAP 50EMA continue to work as AAP is making lower highs in the MACO sense	0
2	sold a little goog ...first time in a while	1
3	Nice cup with handle breakout on MCA	1
4	FAF new highs.	1
5	SPY giving AAP the finger. We don't care about you anymore lol	0

Source: Chaudhary, 2020

3.2 Data Preprocessing

Data preprocessing involved several steps to convert raw text into a format amenable to machine learning algorithms. First, tokenization was performed using a regular expression tokenizer to identify alphabetic strings, thus filtering out numbers and punctuations. All tokens were converted to lowercase and subjected to lemmatization using the WordNetLemmatizer to reduce words to their base or root form. A custom function was applied to convert British English to American English spellings to address variations in

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

English spelling. Additionally, stopwords—commonly occurring words that offer little value in the context of sentiment analysis—were removed from the text. This process was facilitated by the Natural Language Toolkit (NLTK) library, which provides a comprehensive list of stopwords.

3.3 Model Selection

BERT was chosen as the primary model for this research due to its state-of-the-art performance in various NLP tasks, including its ability to understand the context and nuance in text (Devlin et al., 2018). A pre-trained BERT model, 'Bert-base-uncased', was selected from the Hugging Face Transformers library, which provides a robust framework for transformer models. The 'uncased' variant was preferred to ensure that the model did not distinguish between uppercase and lowercase letters, which is unnecessary in sentiment analysis.

3.4 Training Process and Evaluation Metrics

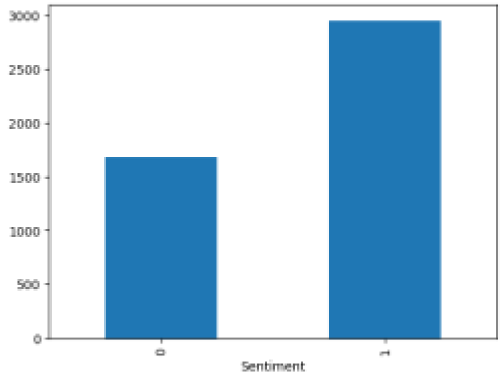

Figure 2: Training Dataset Sentiment Count

Source: Author's own

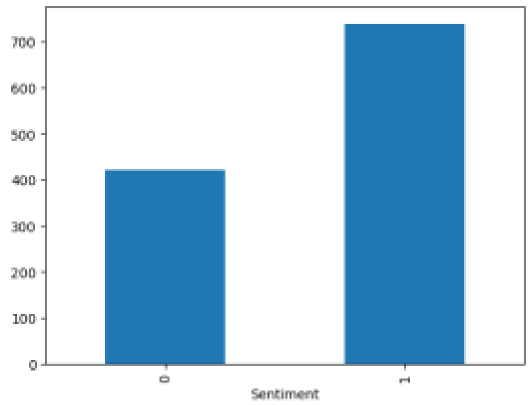

Figure 3: Test Dataset Sentiment Count

Source: Author's own

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

The dataset was split into training (Figure 2) and testing (Figure 3) sets using an 80-20 (4632, 1159) ratio, ensuring a stratified distribution of sentiments across both sets. This approach preserved the sentiment distribution from the original dataset, which is crucial for maintaining consistency in model training and evaluation. The training process utilized the Adam optimizer with a learning rate of 0.00005, as the original BERT authors recommended, along with a categorical cross-entropy loss function appropriate for binary classification tasks (Kingma & Ba, 2015).

Model performance was assessed using categorical and balanced accuracy metrics to account for any potential class imbalance within the dataset. Using both Accuracy-- Macro Average and Weighted Average Accuracy-- metrics provides a more comprehensive evaluation of the model's predictive capabilities. A classification report was generated post-training to elucidate model performance further, providing Precision, Recall, and F1-score for each sentiment class.

FINE TUNING OF THE SYSTEM:

Within this section lies the architectural blueprint of our BERT-based sentiment analysis model. We delve into the structural nuances, from integrating additional layers to fine-tuning processes enabling our model to discern complex sentiment cues within financial texts. Additionally, the description of our training environment and computational resources provides a practical perspective on the scalability and reproducibility of our system design.

4.1 Architecture of the BERT-based Model

The core of the system's architecture is the BERT model, designed to understand the nuances and context of language by pre-training on a large corpus of text (Devlin et al., 2019). For our sentiment analysis task, the 'bert-base-uncased' model was utilized, which consists of 12 layers of Transformer blocks, an attention mechanism that considers the context from both the left and the right side of a token within the text.

4.2 Integration of Additional Layers and Fine-tuning

We appended additional neural network layers to adapt the pre-trained BERT model for our binary classification task, as shown in Figure 4. This fine-tuning stage involved adding a GlobalMaxPool1D layer on top of the BERT output to condense the feature representation from the entire sequence to a fixed-size representation. Following this pooling layer, we introduced two fully connected dense layers with 128 and 32 neurons, respectively, and a dropout layer with a dropout rate of 0.1 to reduce the risk of overfitting. The final output layer is a dense layer with two neurons corresponding to the binary sentiment classes, utilizing a softmax activation function to output a probability distribution over the classes.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

The entire BERT model and the additional layers were trained for fine-tuning. This end-to-end training allows the BERT model to adjust its weights to the specifics of the sentiment analysis task.

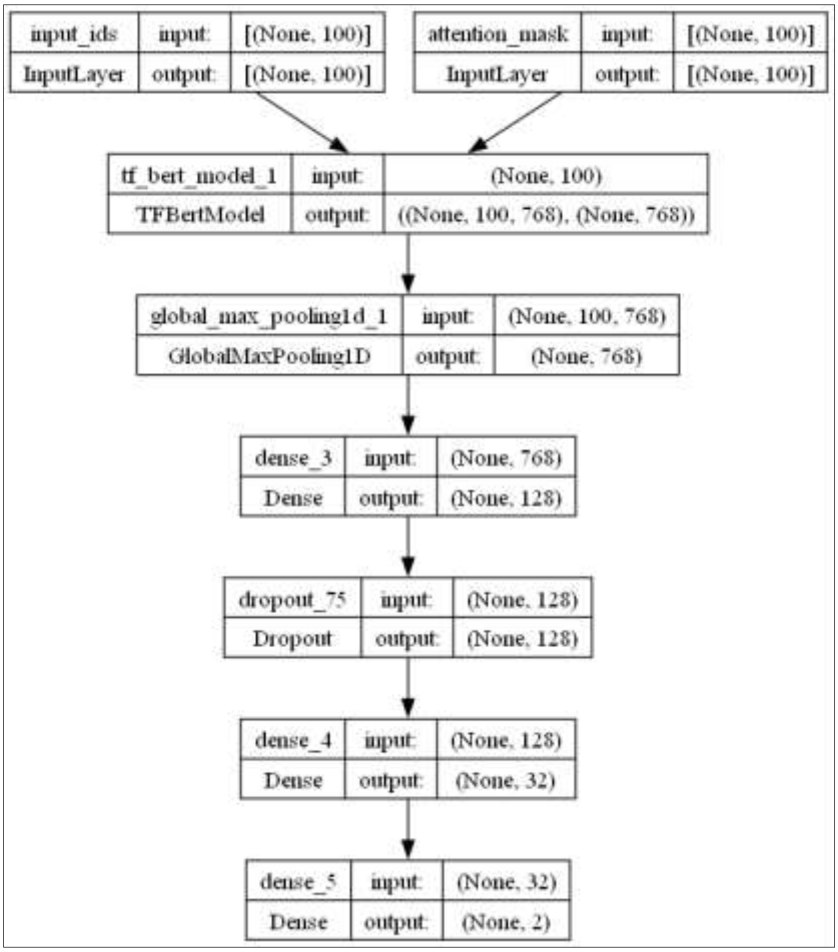

Figure 4: Fine-tuning of the BERT model

Source: Author's own

4.3 Description of the Training Environment and Computational Resources

The training was performed on a cloud-based environment accessed through a Jupyter Notebook interface. This environment provided a Tesla K80 GPU, which allowed for accelerated matrix computations essential for deep learning model training. The software stack included TensorFlow as the main machine learning framework and the Hugging Face Transformers library, which provided the pre-trained BERT model and the necessary infrastructure for implementing transformers.

TensorFlow facilitated the handling of tensors, backpropagation, and optimization tasks, while the Transformers library allowed for seamless integration of BERT into the TensorFlow ecosystem. The model was compiled using the Adam optimizer with parameters

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

such as a learning rate of 0.00005, epsilon of 0.0000008, and a decay rate of 0.01, as per the recommendations from the BERT authors.

The batch size for the training was set to 32, which balanced the computational resource usage and learning stability. The training was conducted over two epochs, sufficient for the model to converge, given the robust pre-trained foundations of BERT and the additional fine-tuning layers designed for task-specific adjustments.

RESULTS:

Table 2: Model Performance

	Precision	Recall	f1score
Negative	0.72	0.61	0.66
Positive	0.8	0.87	0.83
Macro Avg	0.76	0.74	0.75
Weighted Avg	0.77	0.77	0.77

The model training was conducted over two epochs, engaging a BERT-based neural network architecture fine-tuned for the sentiment analysis task on financial news data. The training process aimed to optimize the classification of textual data into 'Negative' and 'Positive' sentiment categories. Upon application to the test set, the model yielded an overall accuracy of 77 percent. This indicates the model's robustness in correctly identifying the sentiment of the given financial texts in most instances. Although not exhaustive, the accuracy metric provides a quick snapshot of model effectiveness in general terms. A deeper dive into the performance metrics reveals a more nuanced understanding of the model's predictive capabilities:

Negative Sentiment Classification: A precision of 72 percent signifies that when the model predicts a text to have a Negative sentiment, it is correct 72 percent of the time. Recall at 61 percent indicates that of all actual Negative instances, the model successfully identifies 61 percent. The F1-score of 66 percent for the Negative category shows a balance between precision and Recall but suggests a potential area for improvement, particularly in increasing the recall rate.

Positive Sentiment Classification: Precision was notably higher for Positive instances, at 80 percent, suggesting the model is more reliable in its Positive predictions. The Recall was also strong at 87 percent, indicating the model's effectiveness in capturing the actual Positive instances within the test set. The F1-score for the Positive sentiment stood at a robust 83 percent, reflecting a well-balanced precision and Recall for this class.

The macro-average F1-score across both categories was 75 percent, which, while reflective of a relatively balanced performance across classes, also underscores the need for further model

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022**

refinement to ensure even performance across both sentiment classifications. The weighted average F1-score matched the overall accuracy at 77 percent, underscoring the model's consistency in prediction across the class distribution within the dataset.

These metrics collectively provide a comprehensive picture of the model's current state, affirming its efficacy in sentiment classification while highlighting areas where the model could be improved, such as enhancing the Recall for Negative sentiments without a significant trade-off in precision.

DISCUSSION:

The results obtained from the BERT-based model on stock market sentiment analysis indicate a promising direction in applying advanced NLP techniques within financial domains. With an overall accuracy of 77 percent, the model demonstrates substantial capability in discerning sentiment from financial texts, a complex task given the nuanced language often employed in such discourse.

When compared with traditional machine learning approaches or less sophisticated NLP models, our BERT-based approach shows a marked improvement, especially in its ability to understand context and semantics within financial news texts. Baseline models like Naïve Bayes or linear classifiers often lack the depth of understanding that a context-aware model like BERT possesses, leading to lower precision and recall rates.

The findings imply that transformer-based models, such as BERT, are suitable and superior for sentiment analysis in finance. The application of such models can be seen as a significant step forward in developing trading algorithms, risk assessment tools, and market sentiment analysis that can operate with greater accuracy and depth of understanding.

However, the dataset used, while substantial, may not fully capture the diversity of language used across all financial texts. Future research could focus on expanding the dataset to include a more diverse set of financial documents, including more global sources. Additionally, experimenting with different configurations of BERT, or even newer models in the transformer family, could improve results. Furthermore, the fine-tuning of BERT, while effective, did not explore the full extent of hyperparameter optimization, which could potentially yield better results. The computational resources were also a constraint, limiting the ability to train the model over a more extended period or with a more substantial dataset.

REFERENCES:

1. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

2. Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *0123456789*. https://doi.org/10.1007/s10462-023-10442-2

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022**

3. Chalkidis, I., & Kampas, D. (2019). Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law*, *27*(2), 171–198. https://doi.org/10.1007/s10506-018-9238-9

4. Chaudhary, Y. (2020). *Stock-Market Sentiment Dataset*. https://www.kaggle.com/datasets/yash612/stockmarket-sentiment-dataset

5. Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Naacl-Hlt 2019*, *Mlm*.

6. Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T.-Y. (2018). Listening to Chaotic Whispers. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, *2018-Febua*, 261–269. https://doi.org/10.1145/3159652.3159690

7. Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. In *International Review of Financial Analysis* (Vol. 33, pp. 171–185). https://doi.org/10.1016/j.irfa.2014.02.006

8. Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

9. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, *36*(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682

10. Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

11. Sousa, M. G., Sakiyama, K., Rodrigues, L. D. S., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for Stock Market Sentiment Analysis. *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, *2019-Novem*, 1597–1601. https://doi.org/10.1109/ICTAI.2019.00231

12. Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x