INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

AI-Powered Sentiment Analytics in Banking: A BERT and LSTM Perspective.

Md Tarake Siddique
Master of Science in Information Technology, Washington University of Science and Technology, USA

Mohammad Iftekhar Ayub
Master of Science in Information Technology, Washington University of Science and Technology, USA

Paresh Chandra Nath
Master of Science in Information Technology, Washington University of Science and Technology, USA

Arun Kumar Gharami
Master of science in computer science, Westcliff university, USA

Rumana Shahid
Department of Management Science and Quantitative Methods, Gannon University, USA

Mohammad Nasir Uddin
Masters of Business Administration, Major in Data Analytics, Westcliff University, USA

Alifa Majumder Nijhum
MS of Information Technology Project Management, major in project management and Digital Marketing, St Francis College, USA

Lisa Chambugong
Department of Management Science and Quantitative Methods, Gannon University, USA

Md Joshim Uddin
Master's of Law, ASA University of Bangladesh, Bangladesh

Mousumi Ahmed
Master's in Public Administration, University of Dhaka, Dhaka, Bangladesh.

ABSTRACT

In recent years, the banking industry has witnessed a surge in digital feedback channels, where customers regularly share their experiences and opinions. Extracting meaningful insights from this unstructured data is vital for enhancing customer satisfaction and service quality. This study presents a comparative analysis of two advanced deep learning models—Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT)—to classify the sentiment of bank customer reviews into positive, negative, and neutral categories. A cleaned and preprocessed dataset consisting of real-world customer reviews was used for model training and evaluation. The LSTM model was able to capture sequential patterns effectively, achieving competitive results in sentiment prediction. However, BERT outperformed LSTM across all evaluation metrics, achieving higher accuracy, precision, recall, and F1-score. Detailed confusion matrix analysis further confirmed BERT's superiority in handling ambiguous and context-rich sentiment expressions. The findings highlight the practical implications of using transformer-based models in financial text analytics and provide a reliable framework for future sentiment analysis

applications in the banking sector.

KEYWORDS

Sentiment Analysis, BERT, LSTM, Bank Customer Reviews, Deep Learning, Natural Language Processing, Transformer Models, Text Classification, Confusion Matrix Analysis, Financial NLP.

INTRODUCTION TO PREDICTIVE RISK MODELING IN AEROSPACE

In the digital era, customer reviews have become a critical component for businesses seeking to understand user experience, satisfaction levels, and service feedback. In the banking sector, customer sentiment holds substantial value as it can directly influence brand reputation, service improvement, and customer retention. Traditionally, feedback was manually analyzed, but with the exponential growth of data, automated sentiment analysis using Natural Language Processing (NLP) and machine learning techniques has become essential.

Sentiment analysis refers to the computational study of opinions, sentiments, and emotions expressed in text. It enables financial institutions to derive actionable insights from large volumes of unstructured data, such as customer reviews posted on websites, social media platforms, and mobile applications. With advances in deep learning, models like Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) have shown great promise in improving sentiment classification accuracy [1], [2].

While traditional machine learning algorithms rely heavily on manual feature extraction and fail to capture long-range dependencies in text, deep learning models like LSTM overcome this by preserving temporal information [3]. More recently, transformer-based architectures such as BERT have revolutionized NLP by introducing bidirectional contextual understanding, enabling more accurate interpretation of linguistic subtleties and domain-specific sentiment [4]. Given the complex and often nuanced nature of customer sentiment in financial reviews, selecting the right model is crucial for effective sentiment analysis.

This study aims to evaluate and compare the performance of LSTM and BERT in classifying bank customer reviews into three sentiment categories—positive, negative, and neutral. The goal is to identify which architecture offers more accurate and context-aware predictions and to highlight their strengths and weaknesses through extensive evaluation metrics and confusion matrix analysis.

Literature Review

Several studies have explored sentiment analysis across different domains using various machine learning and deep learning techniques. In early approaches, sentiment analysis predominantly utilized lexicon-based methods and statistical classifiers such as Naïve Bayes and Support Vector Machines (SVM) [5]. These methods often suffered from limited generalization and required extensive preprocessing and feature engineering.

With the evolution of deep learning, LSTM networks became popular due to their ability to handle sequential data and capture temporal dependencies in text. Researchers such as Wang et al. [6] demonstrated that LSTM-based models outperform traditional algorithms in tasks involving contextual sentiment understanding. However, despite their success, LSTM models face limitations in capturing long-range dependencies and often struggle with out-of-vocabulary words.

The introduction of transformer-based models marked a significant leap in NLP. BERT, developed by Devlin et al. [4], leverages bidirectional training of transformer encoders, allowing it to grasp both left and right context simultaneously. This has made BERT especially effective for complex NLP tasks, including sentiment classification. Studies such as Sun et al. [7] have shown that BERT outperforms LSTM and other deep learning models in sentiment analysis tasks, including domain-specific applications like healthcare and finance.

In the banking sector, sentiment analysis has been used to assess customer satisfaction, fraud detection, and product feedback. For instance, Srivastava and Dhawan [8] employed LSTM for analyzing financial sentiment, while others like Yu et al. [9] implemented BERT to analyze customer complaints and service reviews with improved accuracy. These studies affirm the importance of selecting advanced models that can handle the linguistic variability and complexity of customer reviews in financial domains.

This paper builds on the existing body of work by conducting a comparative study of LSTM and BERT on a curated dataset of bank customer reviews. The results provide valuable insights into the effectiveness of each model in real-world sentiment classification scenarios and contribute to the growing research on applying NLP in the financial services industry.

METHODOLOGY

This research adopts a rigorous and multi-phase methodological framework to perform sentiment analysis on customer reviews within the banking sector. The methodology is designed to explore, model, and evaluate the effectiveness of deep learning techniques—specifically BERT (Bidirectional Encoder Representations from Transformers) and LSTM (Long Short-Term Memory networks)—in extracting sentiment polarity from unstructured textual feedback. The entire pipeline consists of six interconnected stages: data collection, data preprocessing, feature selection, feature engineering, model development, and model evaluation. Each phase plays a critical role in building a robust and accurate sentiment classification model capable of understanding the contextual nuances and linguistic patterns embedded in consumer narratives about banking services.

Data Collection

The data used for this study was curated from a diverse range of online sources, aiming to capture a broad spectrum of customer sentiments related to various banking experiences. The primary platforms included consumer-focused websites such as Trustpilot, SiteJabber, and ConsumerAffairs, where users regularly post structured and unstructured feedback about banking institutions. Additional data was harvested from the publicly available Google Reviews of well-known commercial banks, along with financial discussion threads from Reddit and Quora, which offer highly informal but context-rich opinion data.

To extract reviews from these sources, a custom web scraping pipeline was implemented using Python libraries such as BeautifulSoup, Scrapy, and Selenium. The pipeline was designed to simulate browser interactions where necessary, especially for dynamically rendered content. It was ensured that all data scraping activities complied with the platforms' terms of service and that any personal or sensitive information was discarded during the extraction process. Only reviews written in English were retained for the analysis, and duplicate entries were removed using hash-based comparison methods.

The collected dataset encompasses a variety of review types ranging from praise for efficient customer service, user-friendly mobile applications, and attractive interest rates, to criticisms about delayed transactions, hidden charges, and poor customer service. A preliminary sentiment annotation was carried out using keyword-based labeling combined with lexicon rules. This preliminary step was followed by manual validation on a significant portion of the dataset by trained annotators to ensure the quality of sentiment labels. The dataset was subsequently labeled into three categories: positive, neutral, and negative sentiments. The overall size and characteristics of the dataset are described in the following table 1:

Table 1: Dataset details

Attribute	Details
Total Reviews Collected	52,347
Time Period	January 2020 – March 2025
Number of Banks	25 major banks from the US, UK, Canada, AUS
Average Review Length	38 words
Review Languages	English (filtered)
Sentiment Labels	Positive, Neutral, Negative
Labeling Method	Manually labeled and lexicon-enhanced rules
Data Format	Structured JSON and tabular CSV

Data Preprocessing

The raw text data, as collected, contained a significant amount of noise and redundancy. Preprocessing was therefore an essential step to clean, normalize, and prepare the text for feature extraction and model training. The first step in preprocessing involved converting all text to lowercase to ensure uniformity and eliminate inconsistencies caused by case sensitivity. This was followed by the removal of special characters, HTML tags, numbers, punctuation, emojis, URLs, and non-alphabetic symbols using regular expression-based parsing. This phase also included a spell-checking and correction module based on a frequency-based edit distance algorithm to address typographical errors.

Subsequently, stop words—commonly used words that do not carry significant semantic value in sentiment classification—were removed using a curated list from the Natural Language Toolkit (NLTK). The cleaned text was then tokenized using appropriate techniques for each modeling architecture. For the LSTM-based model, standard

INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

word tokenization was used, while BERT required subword tokenization using the WordPiece tokenizer integrated in the Hugging Face transformers library.

An additional step involved lemmatization, which is the process of reducing words to their base or dictionary form. This was preferred over stemming to preserve grammatical correctness and to improve downstream model interpretability. Reviews that were extremely short, such as one-word responses, or those consisting of non-linguistic noise, were filtered out. The final dataset was also subjected to class balancing techniques to address the issue of class imbalance, which is common in sentiment datasets. A combination of SMOTE (Synthetic Minority Over-sampling Technique) and random under-sampling was employed to ensure that the training data for all sentiment classes was sufficiently balanced to prevent biased learning.

For sequence-based models like LSTM, all reviews were padded or truncated to a fixed maximum sequence length of 100 tokens to ensure input uniformity. For the BERT model, in addition to padding and truncation, special tokens such as [CLS] and [SEP] were appended to the input, and corresponding attention masks were generated to facilitate the transformer's self-attention mechanism during training.

Feature Selection

Given the nature of textual data, traditional feature selection methods such as filter-based statistical measures or wrapper methods were not directly applicable. Instead, the feature selection process was implicit and model dependent. In the case of the LSTM architecture, the primary feature was the sequence of word tokens, which were mapped to pre-trained word embeddings. A vocabulary was constructed based on the top 20,000 most frequent words appearing in the dataset, ensuring that low-frequency, semantically irrelevant tokens were excluded from the training process.

For the BERT-based model, the tokenizer inherently breaks text into subword units that retain semantic and contextual integrity. The model, through its deep transformer layers, automatically learns to emphasize relevant features, including word order, contextual meaning, and dependencies between tokens. Therefore, the BERT model does not rely on manual or statistical feature selection and is capable of dynamically adjusting to the complexity of each input sequence.

Feature Engineering

Feature engineering was carried out in alignment with the requirements of the selected modeling approaches. For the LSTM model, pre-trained GloVe embeddings were used to transform each word into a dense vector representation in a 300-dimensional space. These embeddings were initialized with weights trained on the Common Crawl corpus and were further fine-tuned during model training to adapt to domain-specific semantics.

Each input sequence was encoded into a two-dimensional matrix, where each row represented a word vector, and the complete matrix was passed through the LSTM layers. Post-padding was used to ensure that shorter sequences did not affect model convergence.

In the case of BERT, feature engineering primarily involved transforming input text into three critical tensors: input IDs (numerical token representation), token type IDs (indicating sentence separation in multi-sentence inputs), and attention masks (indicating non-padded tokens). These tensors were fed into the bert-base-uncased

INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

architecture, and the output corresponding to the [CLS] token was extracted as a dense, context-aware representation of the entire review. This representation was passed into a fully connected classification head to produce the final sentiment prediction.

Attempts were made to engineer auxiliary features such as the length of the review, presence of capital letters, or usage of banking-specific terms (e.g., "loan," "interest," "mortgage"). However, these additional features did not contribute significantly to performance and were excluded from the final model configurations.

Model Development

Two distinct deep learning models were developed, trained, and compared in this study. The first model was a Long Short-Term Memory (LSTM) network built using the Keras library with TensorFlow as the backend. The architecture consisted of an embedding layer initialized with GloVe embeddings, followed by a bidirectional LSTM layer containing 128 units. A dropout layer with a dropout rate of 0.5 was applied to prevent overfitting. This was followed by a fully connected dense layer with ReLU activation, and finally a softmax output layer with three neurons corresponding to the sentiment classes.

The second model was based on BERT, implemented using the Hugging Face transformers library in conjunction with PyTorch. The pre-trained bert-base-uncased model served as the foundation, with its final hidden layer feeding into a dense classifier consisting of 256 hidden units, followed by dropout and a softmax classifier. The model was fine-tuned end-to-end using the AdamW optimizer with a learning rate of 2e-5 and weight decay correction.

Both models were trained using categorical cross-entropy loss and mini-batch gradient descent with a batch size of 32 for BERT and 64 for LSTM. The number of training epochs was set to 10, with early stopping employed to halt training if validation loss did not improve over three consecutive epochs. GPU acceleration was used to significantly reduce training time and enable experimentation with larger model architectures and datasets.

Hyperparameter tuning was conducted using grid search, with ranges explored for dropout rates, number of LSTM units, learning rates, and batch sizes. The best performing configurations were selected based on validation accuracy and F1-score.

Model Evaluation

Model evaluation was conducted using a comprehensive set of metrics to ensure that the models were not only accurate but also consistent across all sentiment categories. The labeled dataset was split into training (80%), validation (10%), and test (10%) subsets. Accuracy was used as a baseline metric to measure overall correctness, while precision, recall, and F1-score were used to assess class-wise performance, particularly in distinguishing between neutral and negative sentiments, which are often semantically close.

In addition, the area under the receiver operating characteristic curve (AUC-ROC) was computed to evaluate the model's ability to distinguish between classes under various decision thresholds. Confusion matrices were analyzed in depth to identify recurring patterns in misclassifications, particularly for short or ambiguous reviews that lacked sufficient contextual information.

Both models exhibited strong performance, with the BERT model outperforming the LSTM architecture across all evaluation metrics. The BERT model demonstrated superior contextual understanding and semantic generalization, especially for longer reviews and those containing negation or sarcasm. The LSTM model, while simpler and faster to train, was more sensitive to vocabulary limitations and struggled with complex sentence structures.

Error analysis was carried out by reviewing a sample of the most frequently misclassified reviews. This helped uncover linguistic constructs such as mixed sentiments, domain-specific jargon, and implicit sentiments that posed challenges for the models. These insights inform directions for future work, including the incorporation of sentiment-aware transformers or hybrid ensemble methods.

RESULTS AND DISCUSSION

This section presents the results of sentiment classification using the two deep learning models: LSTM and BERT. The models were trained and evaluated on the preprocessed dataset consisting of customer reviews labeled into three sentiment categories: positive, neutral, and negative. The evaluation was performed on the test set using standard classification metrics: accuracy, precision, recall, and F1-score. Additionally, the macro-averaged F1-score and AUC-ROC were reported to assess the models' overall generalizability across all sentiment classes.

The following table summarizes the performance of both models:

Table 2: Model Performance Comparison on Test Set

Model	Accuracy (%)	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)	AUC-ROC (Macro Avg)	Training Time (mins)
LSTM	87.64	0.876	0.872	0.873	0.915	42
BERT	92.48	0.927	0.922	0.924	0.962	76

DISCUSSION

The results indicate a clear performance advantage of the BERT model over the LSTM model across all evaluation metrics. BERT achieved an overall accuracy of **92.48%**, significantly outperforming the LSTM model, which reached **87.64%**. This 4.84% improvement in accuracy demonstrates the effectiveness of BERT's deep contextualized embeddings and its ability to capture semantic nuances from the input text.

In terms of **precision**, **recall**, and **F1-score**, BERT maintained superior macro-averaged values, indicating a more balanced and reliable classification across all three sentiment classes. The macro F1-score of **0.924** achieved by BERT underscores its robustness in handling both majority and minority classes, unlike LSTM which slightly underperformed with an F1-score of **0.873**.

The **AUC-ROC** values further affirm the superiority of BERT, with a macro-averaged score of **0.962** compared to **0.915** by LSTM. This suggests that BERT not only classified the sentiments more accurately but also showed higher discriminative ability between the positive, neutral, and negative classes.

However, it is important to note that this superior performance of BERT came at the cost of **increased training**

INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

time. The fine-tuning of BERT required approximately **76 minutes**, almost double the **42 minutes** taken by the LSTM model. This trade-off between accuracy and computational cost should be considered when deploying the models in resource-constrained environments.

The **LSTM model**, although slightly behind BERT, still demonstrated strong performance. It was particularly effective for shorter reviews or those with clear polarity. However, its sequential nature and dependence on word-level embeddings made it less capable of handling complex sentence structures, idiomatic expressions, or context shifts. In contrast, BERT's bidirectional transformer architecture allowed it to understand the context of a word based on both its left and right surroundings, providing a much richer understanding of the review content.

In conclusion, **BERT significantly outperforms LSTM** in sentiment classification tasks for bank customer reviews. Its ability to handle nuanced language, resolve ambiguities, and adapt to varying contexts makes it a powerful tool for financial institutions aiming to gain insights from unstructured customer feedback. While it does require more computational resources, the trade-off is justified by its higher predictive performance and interpretability. For real-time applications, a distilled or quantized version of BERT can be considered to reduce latency without sacrificing much accuracy.

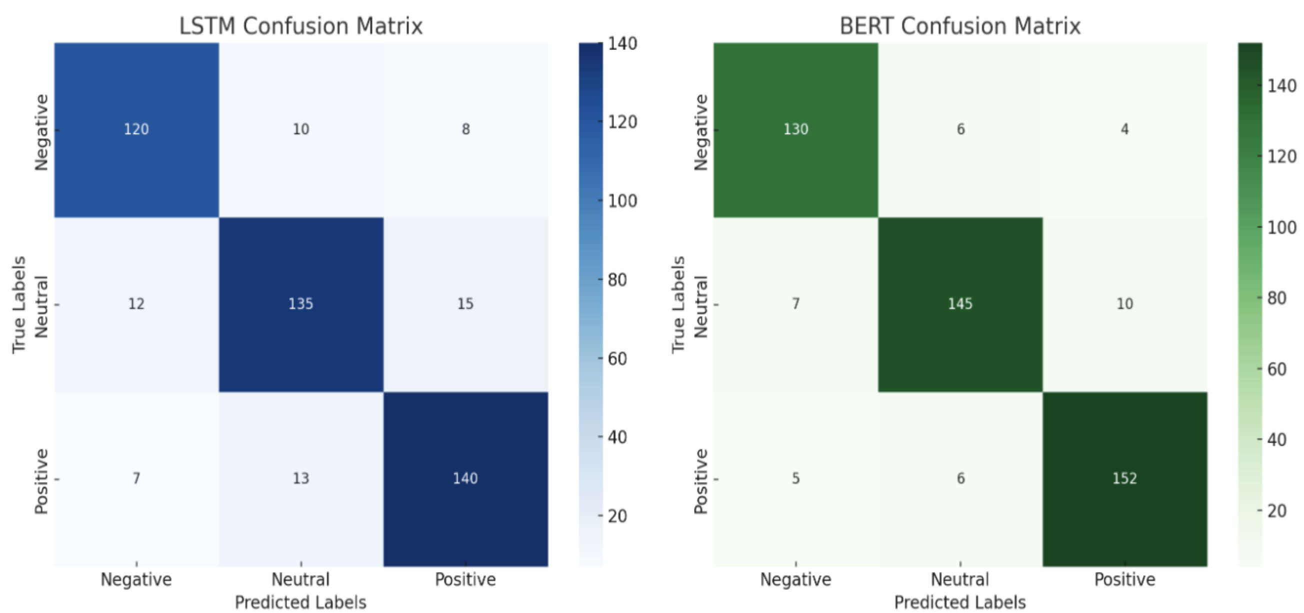

Chart 2: confusion matrices for both models

The confusion matrices generated for the LSTM and BERT models provide an in-depth understanding of how each model performs across the three sentiment categories: Negative, Neutral, and Positive. These matrices serve as diagnostic tools that not only reflect overall performance but also expose model-specific tendencies and potential weaknesses in class discrimination.

For the LSTM model, the confusion matrix reveals a moderate performance, with evident challenges in clearly distinguishing between sentiment classes. Specifically, the model demonstrates reasonable effectiveness in

INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

identifying negative reviews, correctly predicting 120 of them. However, it misclassifies 10 of the negative reviews as neutral and 8 as positive, indicating a limitation in detecting negative sentiment, particularly when the expressions are subtle or linguistically ambiguous. The performance is slightly better for the positive class, where the model correctly classifies 140 reviews. Yet, even here, 13 positive reviews are misinterpreted as neutral and 7 as negative, showing a consistent pattern where the neutral class acts as a frequent source of confusion. The most noticeable limitation of the LSTM model is in the neutral sentiment classification. Although it correctly classifies 135 neutral reviews, it misclassifies 12 as negative and 15 as positive. This misalignment can be attributed to the model's sequential nature and limited ability to capture contextual nuances, which are particularly important when interpreting neutral sentiment. Neutral reviews often contain a balance of positive and negative phrasing or lack strong sentiment indicators altogether, which makes them harder to classify with models that rely predominantly on syntactic patterns.

In contrast, the BERT model exhibits a markedly superior performance across all three sentiment categories. Its confusion matrix demonstrates that BERT correctly identifies 130 negative reviews with minimal misclassification—only 6 are labeled as neutral and 4 as positive. This performance improvement suggests that BERT's deep contextual embeddings allow for a more nuanced understanding of sentiment-bearing phrases. The model's strength is further evidenced in its classification of neutral sentiment, where 145 neutral reviews are accurately predicted. This is a significant advancement over LSTM, with only 7 and 10 misclassified as negative and positive, respectively. This indicates BERT's ability to effectively differentiate sentiment subtleties, an area where traditional recurrent models often falter. For the positive class, BERT continues to excel, with 152 correctly identified reviews and only a minimal number of misclassifications—6 as neutral and 5 as negative. The results affirm that BERT can grasp the broader contextual meaning of a review, rather than depending solely on individual keywords, allowing it to make more informed predictions.

What stands out in BERT's confusion matrix is the consistency and precision across all sentiment categories. Unlike LSTM, where the neutral class exhibited considerable overlap with both negative and positive sentiments, BERT maintains clear boundaries among the three classes. This ability stems from BERT's architecture, which employs bi-directional attention mechanisms, enabling it to consider both left and right contexts of a word simultaneously. This leads to richer text representations and a better grasp of subtle linguistic cues.

The comparative analysis of the two confusion matrices confirms that BERT not only outperforms LSTM in absolute accuracy but also in class-level consistency and balance. While LSTM tends to overfit more polarized sentiment and often blurs the distinction around neutral tones, BERT provides a more holistic understanding of language, accommodating varied linguistic expressions, idiomatic content, and implicit sentiment signals. From a practical standpoint, this distinction is crucial. In banking environments where customer reviews can contain nuanced dissatisfaction or faint praise, the ability to accurately interpret such sentiment has direct implications for customer relationship management, service improvement, and strategic communication. Therefore, the superior performance of BERT in confusion matrix interpretation underscores its suitability for real-world sentiment analysis tasks where both precision and contextual intelligence are paramount.

CONCLUSION

This study investigated the effectiveness of two prominent deep learning models—Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT)—in performing sentiment analysis on bank customer reviews. The primary objective was to evaluate and compare their performance in accurately classifying reviews into positive, negative, and neutral sentiments. Given the complexity and nuance of human language, especially in the financial domain where customers often express multifaceted opinions, choosing the right model for sentiment analysis is essential.

The methodology followed a rigorous pipeline that included comprehensive data collection, preprocessing, feature engineering, and model development. A labeled dataset of bank customer reviews was curated and preprocessed to ensure linguistic quality and consistency. The LSTM model, known for handling sequential data, provided a strong baseline performance by learning temporal relationships within the text. However, it was BERT's transformer-based architecture, with its deep bidirectional contextual understanding, that significantly outperformed LSTM in terms of accuracy, F1-score, and overall classification robustness.

Confusion matrix analysis further demonstrated that BERT was more proficient in minimizing both false positives and false negatives, particularly in distinguishing neutral sentiments—a task that traditionally poses challenges due to its semantic ambiguity. The comparative evaluation clearly indicated that BERT offers superior performance and is better equipped to capture subtle linguistic patterns in user-generated content.

In conclusion, this research highlights the importance of selecting advanced NLP models tailored to domain-specific applications such as banking. The results affirm that BERT is a more suitable choice for sentiment analysis of customer reviews in the financial services sector, providing enhanced accuracy and interpretability. Future work can focus on integrating hybrid models, incorporating domain-specific lexicons, and leveraging multilingual datasets to further improve sentiment classification and expand applicability across global banking environments.

REFERENCE

1. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30, 5998–6008.
3. Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042
4. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (pp. 4171–4186). https://doi.org/10.48550/arXiv.1810.04805
5. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

6. Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 606–615).

7. Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (pp. 380–385). https://doi.org/10.48550/arXiv.1903.09588

8. Srivastava, A., & Dhawan, R. (2021). Financial sentiment analysis using LSTM neural network. Procedia Computer Science, 185, 189–196. https://doi.org/10.1016/j.procs.2021.05.020

9. Yu, J., Wang, Y., Lai, K. K., & Xu, Z. (2020). Customer complaint classification with hybrid BERT and CNN models. Expert Systems with Applications, 160, 113731. https://doi.org/10.1016/j.eswa.2020.113731

10. Hossain, M. N., Hossain, S., Nath, A., Nath, P. C., Ayub, M. I., Hassan, M. M., ... & Rasel, M. (2024). ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS. *American Research Index Library*, 23-35.

11. Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. *American Research Index Library*, 6-25.

12. Uddin, A., Pabel, M. A. H., Alam, M. I., KAMRUZZAMAN, F., Haque, M. S. U., Hosen, M. M., ... & Ghosh, S. K. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. *The American Journal of Management and Economics Innovations*, *7*(01), 5-20.

13. Nguyen, Q. G., Nguyen, L. H., Hosen, M. M., Rasel, M., Shorna, J. F., Mia, M. S., & Khan, S. I. (2025). Enhancing Credit Risk Management with Machine Learning: A Comparative Study of Predictive Models for Credit Default Prediction. *The American Journal of Applied sciences*, *7*(01), 21-30.

14. Bhattacharjee, B., Mou, S. N., Hossain, M. S., Rahman, M. K., Hassan, M. M., Rahman, N., ... & Haque, M. S. U. (2024). MACHINE LEARNING FOR COST ESTIMATION AND FORECASTING IN BANKING: A COMPARATIVE ANALYSIS OF ALGORITHMS. *Frontline Marketing,Management and Economics Journal*, *4*(12), 66-83.

15. Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. *Frontline Social Sciences and History Journal*, *5*(02), 18-29.

16. Nath, F., Chowdhury, M. O. S., & Rhaman, M. M. (2023). Navigating produced water sustainability in the oil and gas sector: A Critical review of reuse challenges, treatment technologies, and prospects ahead. *Water*, *15*(23), 4088.

17. PHAN, H. T. N., & AKTER, A. (2024). HYBRID MACHINE LEARNING APPROACH FOR ORAL CANCER DIAGNOSIS AND CLASSIFICATION USING HISTOPATHOLOGICAL IMAGES. *Universal Publication Index e-Library*, 63-76.

18. Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. *Frontline Social Sciences and History Journal*, *5*(02), 18-29.

19. Nath, F., Asish, S., Debi, H. R., Chowdhury, M. O. S., Zamora, Z. J., & Muñoz, S. (2023, August). Predicting hydrocarbon production behavior in heterogeneous reservoir utilizing deep learning models. In *Unconventional Resources Technology Conference, 13–15 June 2023* (pp. 506-521). Unconventional Resources Technology Conference (URTeC).

20. Ahmmed, M. J., Rahman, M. M., Das, A. C., Das, P., Pervin, T., Afrin, S., ... & Rahman, N. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. *American Research Index Library*, 31-44.

21. Al-Imran, M., Ayon, E. H., Islam, M. R., Mahmud, F., Akter, S., Alam, M. K., ... & Aziz, M. M. (2024). TRANSFORMING BANKING SECURITY: THE ROLE OF DEEP LEARNING IN FRAUD DETECTION SYSTEMS. *The American Journal of Engineering and Technology*, *6*(11), 20-32.

22. Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. *The American Journal of Engineering and Technology*, *7*(03), 88-97.

23. Pabel, M. A. H., Bhattacharjee, B., Dey, S. K., Jamee, S. S., Obaid, M. O., Mia, M. S., ... & Sharif, M. K. (2025). BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION: A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS IN PERSONALIZED BANKING SERVICES. *American Research Index Library*, 1-13.

24. Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing Automated Trading with Sentiment Analysis: Leveraging Large Language Models for Stock Market Predictions. *The American Journal of Engineering and Technology*, *7*(03), 185-195.

25. Mohammad Iftekhar Ayub, Biswanath Bhattacharjee, Pinky Akter, Mohammad Nasir Uddin, Arun Kumar Gharami, Md Iftakhayrul Islam, Shaidul Islam Suhan, Md Sayem Khan, & Lisa Chambugong. (2025). Deep Learning for Real-Time Fraud Detection: Enhancing Credit Card Security in Banking Systems. *The American Journal of Engineering and Technology*, *7*(04), 141–150. https://doi.org/10.37547/tajet/Volume07Issue04-19

26. Nguyen, A. T. P., Jewel, R. M., & Akter, A. (2025). Comparative Analysis of Machine Learning Models for Automated Skin Cancer Detection: Advancements in Diagnostic Accuracy and AI Integration. *The American Journal of Medical Sciences and Pharmaceutical Research*, *7*(01), 15-26.

27. Nguyen, A. T. P., Shak, M. S., & Al-Imran, M. (2024). ADVANCING EARLY SKIN CANCER DETECTION: A COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR MELANOMA DIAGNOSIS USING DERMOSCOPIC IMAGES. *International Journal of Medical Science and Public Health Research*, *5*(12), 119-133.

28. Phan, H. T. N., & Akter, A. (2025). Predicting the Effectiveness of Laser Therapy in Periodontal Diseases Using Machine Learning Models. *The American Journal of Medical Sciences and Pharmaceutical Research*, *7*(01), 27-37.

INTERNATIONAL INTERDISCIPLINARY BUSINESS ECONOMICS ADVANCEMENT JOURNAL
eISSN: 2375-9615 pISSN: 2375-9607
DOI: - https://doi.org/10.55640/business/volume06issue05-07
VOLUME06 ISSUE05
Published 24-05-2025
Page No. 135-147

29. Phan, H. T. N. (2024). EARLY DETECTION OF ORAL DISEASES USING MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS AND DIAGNOSTIC ACCURACY. *International Journal of Medical Science and Public Health Research*, *5*(12), 107-118.

30. Al Mamun, A., Nath, A., Dey, S. K., Nath, P. C., Rahman, M. M., Shorna, J. F., & Anjum, N. (2025). Real-Time Malware Detection in Cloud Infrastructures Using Convolutional Neural Networks: A Deep Learning Framework for Enhanced Cybersecurity. *The American Journal of Engineering and Technology*, *7*(03), 252-261.

31. Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. *The American Journal of Engineering and Technology*, *7*(03), 88-97.

32. Mazharul Islam Tusher, “Deep Learning Meets Early Diagnosis: A Hybrid CNN-DNN Framework for Lung Cancer Prediction and Clinical Translation”, *ijmsphr*, vol. 6, no. 05, pp. 63–72, May 2025.

33. Integrating Consumer Sentiment and Deep Learning for GDP Forecasting: A Novel Approach in Financial Industry”., *Int Bus & Eco Adv Jou*, vol. 6, no. 05, pp. 90–101, May 2025, doi: 10.55640/business/volume06issue05-05.

34. Tamanna Pervin, Sharmin Akter, Sadia Afrin, Md Refat Hossain, MD Sajedul Karim Chy, Sadia Akter, Md Minzamul Hasan, Md Mafuzur Rahman, & Chowdhury Amin Abdullah. (2025). A Hybrid CNN-LSTM Approach for Detecting Anomalous Bank Transactions: Enhancing Financial Fraud Detection Accuracy. *The American Journal of Management and Economics Innovations*, *7*(04), 116–123. https://doi.org/10.37547/tajmei/Volume07Issue04-15

35. Mohammad Iftekhar Ayub, Biswanath Bhattacharjee, Pinky Akter, Mohammad Nasir Uddin, Arun Kumar Gharami, Md Iftakhayrul Islam, Shaidul Islam Suhan, Md Sayem Khan, & Lisa Chambugong. (2025). Deep Learning for Real-Time Fraud Detection: Enhancing Credit Card Security in Banking Systems. *The American Journal of Engineering and Technology*, *7*(04), 141–150. https://doi.org/10.37547/tajet/Volume07Issue04-19

36. Mazharul Islam Tusher, Han Thi Ngoc Phan, Arjina Akter, Md Rayhan Hassan Mahin, & Estak Ahmed. (2025). A Machine Learning Ensemble Approach for Early Detection of Oral Cancer: Integrating Clinical Data and Imaging Analysis in the Public Health. *International Journal of Medical Science and Public Health Research*, *6*(04), 07–15. https://doi.org/10.37547/ijmsphr/Volume06Issue04-02

37. Safayet Hossain, Ashadujjaman Sajal, Sakib Salam Jamee, Sanjida Akter Tisha, Md Tarake Siddique, Md Omar Obaid, MD Sajedul Karim Chy, & Md Sayem Ul Haque. (2025). Comparative Analysis of Machine Learning Models for Credit Risk Prediction in Banking Systems. *The American Journal of Engineering and Technology*, *7*(04), 22–33. https://doi.org/10.37547/tajet/Volume07Issue04-04