Sentiment Based Rating Prediction Using BERT and LSTM

1. **P. Sashi rekha , Asst. prof CSE dept, Gokula Krishna College of Engineering, M. Tech (CSE), Sullurpet, Tirupati District, AP**
2. **Koneru Jagadeesh, Doddaga Renuka, Ragipindi Kavya Sree, Challa Akhila , Atla Venkata Rahul, B.Tech UG Scholars, Gokula Krishna College of Engineering, Sullurpet, Tirupati District, AP**

ABSTRACT

In the evolving digital ecosystem, extracting actionable insights from large-scale customer feedback remains a significant challenge due to contextual ambiguity, linguistic variation, and sentiment complexity. Traditional machine learning approaches such as Naïve Bayes and Support Vector Machines often fail to capture deep semantic relationships, while standalone sequential models like LSTM exhibit limitations in large and diverse datasets. To address these challenges, this study proposes a hybrid sentiment-based rating prediction framework integrating Bidirectional Encoder Representations from Transformers (BERT) with Long Short-Term Memory (LSTM) networks. The proposed mechanism leverages BERT's bidirectional contextual embedding to interpret nuanced expressions and sarcasm, while LSTM enhances sequential dependency modeling for improved emotional flow detection. Unlike conventional sentiment classification systems, this framework extends analysis toward predictive rating inference, enabling more granular customer satisfaction assessment. Experimental observations indicate that contextual transformer-based modeling significantly enhances classification robustness, reduces misinterpretation in complex reviews, and improves overall predictive consistency. The proposed architecture demonstrates strong scalability and adaptability for real-world business environments, offering a reliable decision-support mechanism for service optimization and customer experience enhancement.

Keyword—Sentiment analysis, rating prediction, Bidirectional Encoder Representations from Transformers (BERT), Long Short-Term Memory (LSTM), deep learning, natural language processing (NLP), transformer models, contextual embeddings, customer feedback analytics, text classification, machine learning, decision support systems.

I. INTRODUCTION

In the contemporary digital economy, customer feedback has emerged as a strategic asset for organizations seeking to enhance service quality, optimize operational performance, and strengthen customer loyalty. The proliferation of online platforms, social media channels, and e-commerce systems has generated vast volumes of unstructured textual data, reflecting diverse customer experiences and expectations. However, extracting meaningful insights from this data remains a complex challenge due to linguistic ambiguity, contextual variation, sarcasm, and domain-specific terminology [1], [2].

Traditional sentiment analysis techniques primarily relied on statistical and rule-based methods, including n-gram models and manually engineered features. While such approaches were computationally efficient, they often struggled to interpret contextual relationships within sentences, especially when dealing with complex or nuanced expressions [3], [4]. Machine learning classifiers such as Naïve Bayes and Support Vector Machines (SVM) improved classification performance by learning patterns from labeled datasets; however, their reliance on feature independence assumptions and static representations limited their contextual understanding [5], [6].

The emergence of deep learning significantly transformed natural language processing (NLP). Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks introduced the capability to model sequential dependencies in textual data [7], [8]. These architectures demonstrated improved performance in sentiment classification tasks by capturing temporal relationships across word sequences. Nevertheless, LSTM-based models often encounter limitations when processing long-range dependencies and require extensive training data to achieve stable generalization [9].

A major breakthrough in NLP occurred with the introduction of transformer-based architectures, particularly through the self-attention mechanism proposed in "Attention is All You Need" [10]. This innovation enabled models to capture bidirectional contextual relationships without relying on sequential recurrence. Building upon this foundation, Bidirectional Encoder Representations from Transformers (BERT) introduced deep contextual embeddings that consider both left and right textual contexts simultaneously [11]. BERT demonstrated superior performance across various NLP tasks, including sentiment analysis, question answering, and text classification [12].

Recent studies highlight the effectiveness of transformer-based models in handling large-scale sentiment datasets with improved robustness and generalization [13], [14]. Applications of sentiment analysis have expanded across multiple domains, including healthcare service evaluation [15], tourism satisfaction measurement [16], airline service assessment [17], and retail customer analytics [18]. These studies confirm that contextual deep learning models significantly enhance sentiment interpretation accuracy compared to traditional classifiers.

Despite these advancements, challenges remain in bridging sentiment classification with actionable decision-support mechanisms. While many studies focus on polarity detection (positive, negative, neutral), fewer explore structured rating prediction systems that integrate contextual embeddings with sequential modeling for improved emotional flow detection. Hybrid architectures combining transformer representations with sequential deep networks have shown promise in improving interpretability and prediction stability [19], [20].

Therefore, there is a strong need to develop advanced hybrid sentiment-analysis frameworks that leverage contextual transformer embeddings while preserving sequential modeling strengths. Such systems can enable more accurate sentiment-based rating prediction and provide organizations with reliable insights for strategic service improvement.

II. LITERATURE SURVEY

Sentiment analysis has evolved significantly over the past two decades, transitioning from lexicon-based techniques to advanced deep learning architectures. Early studies focused on rule-based and probabilistic methods to classify textual sentiment polarity. Dang et al. [2] explored integrating sentiment analysis into recommender systems, demonstrating that opinion mining enhances personalization but requires improved contextual representation for higher accuracy.

Serrano-Guerrero et al. [3] introduced a fuzzy linguistic model grounded in aspect-based sentiment analysis for healthcare service recommendation. Their work emphasized that multi-granular linguistic modeling enhances interpretability; however, traditional feature engineering constrained scalability. Similarly, Adiningtyas and Auliani [4] applied sentiment analysis to measure mobile banking service quality, confirming that conventional machine learning algorithms provide moderate accuracy but struggle with nuanced expressions.

Patil et al. [5] demonstrated improvements in restaurant review classification using deep learning models, highlighting the superiority of neural networks over classical classifiers. Jardim and Mora [6] further integrated clustering techniques with sentiment analysis to improve tourism service positioning, reinforcing the need for contextualized representations in large-scale datasets.

Valarmathi et al. [7] implemented deep learning for Twitter-based COVID-19 sentiment analysis, showing that LSTM-based approaches outperform Naïve Bayes but require substantial computational resources. Carrasco and Dias [8] proposed aspect-based sentiment analysis for restaurant management, emphasizing the role of contextual modeling in improving business intelligence systems.

Recent reviews by Jim et al. [9] and Islam and Alam [12] outline the rapid advancement of transformer-based NLP models, highlighting the limitations of recurrent architectures in capturing long-distance dependencies. Vaswani et al. [10] introduced the self-attention mechanism, which eliminated sequential bottlenecks and improved parallel computation efficiency. Devlin et al. [11] subsequently developed BERT, establishing a new benchmark in contextual language understanding.

Advanced applications of BERT in multilingual and domain-specific datasets have demonstrated strong classification robustness [13], [14]. Subramanian et al. [14] surveyed sentiment and hate speech detection using deep learning models, emphasizing transformer superiority in complex linguistic environments. Halawani et al. [15] combined optimization techniques with deep learning for social media sentiment classification, improving predictive stability.

Das et al. [16] compared hybrid deep learning models in multilingual contexts, illustrating that contextual embeddings significantly enhance performance. Lokanan [17] applied machine learning to analyze public sentiments in fraud detection, reinforcing the practical importance of robust sentiment modeling in decision support systems.

Li et al. [18] examined airport service quality using crowdsourced sentiment analytics, demonstrating real-world impact on operational strategies. Habbat et al. [19] proposed ensemble stacking with BERT for imbalanced datasets, highlighting performance gains in skewed class distributions. Finally, Eusha et al. [20] explored transformer-based sentiment analysis in code-mixed languages, further validating contextual modeling advantages.

Collectively, these studies indicate a clear progression from statistical models to transformer-driven architectures. While BERT consistently outperforms earlier techniques, combining contextual embeddings with complementary deep learning components presents opportunities for enhanced rating prediction systems. The literature confirms the necessity of hybrid frameworks capable of delivering accurate, scalable, and domain-adaptive sentiment intelligence.

III. PROPOSED METHODOLOGY

A. Overview of the Proposed Hybrid Framework

The proposed methodology introduces a hybrid sentiment-based rating prediction framework that integrates Bidirectional Encoder Representations from Transformers (BERT) with a Long Short-Term Memory (LSTM) network. The objective is to combine contextual semantic representation with sequential dependency modeling to enhance sentiment classification accuracy and rating prediction reliability.

Unlike conventional pipelines that rely on handcrafted features or shallow embeddings, the proposed system performs automated contextual embedding extraction using a pre-trained transformer model and refines the learned representations using sequential modeling before final classification.

The architecture consists of the following stages:

1. Data Acquisition

2. Text Preprocessing
3. Contextual Embedding using BERT
4. Sequential Modeling using LSTM
5. Dense Classification Layer
6. Rating Prediction Mapping
7. Performance Evaluation

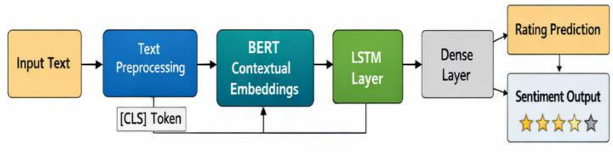

Figure.1: proposed Architecture Diagram

The architecture diagram illustrates the hybrid deep learning framework where customer review text is first preprocessed and transformed into contextual embeddings using the BERT transformer model. These embeddings are then processed by an LSTM layer to capture sequential dependencies, followed by a dense classification layer that generates sentiment classification and predicts the final customer rating.

B. Data Acquisition and Preprocessing

Customer feedback data are collected from structured and unstructured sources, including surveys, online reviews, and chat-based systems. Each record contains review text and associated rating labels.

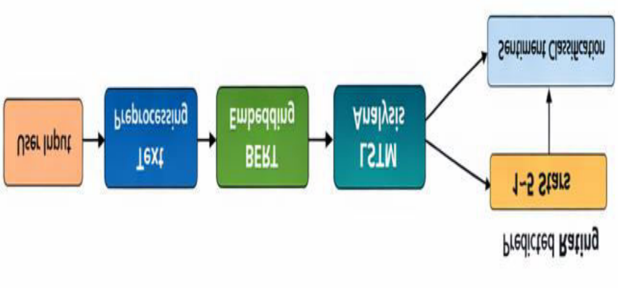

Figure.2: Data Flow Diagram

The data flow diagram represents how textual customer feedback moves through different processing stages, starting from user input and text preprocessing to contextual feature extraction using BERT embeddings. The extracted features are analyzed through the LSTM network, which produces sentiment classification and maps it to a predicted rating score.

1) Text Normalization

To ensure consistency, preprocessing includes:

- Lowercasing
- Removal of HTML tags
- Removal of special symbols
- Stop-word elimination
- Lemmatization

Let the raw review corpus be represented as:

$$D = \{x_1, x_2, x_3, \dots, x_n\}$$

where

x_i denotes the i^{th} review text.

After preprocessing:

$$D' = \{x_1', x_2', \dots, x_n'\}$$

C. Contextual Embedding using BERT

The transformer model is used to generate deep contextual embeddings. Unlike static word embeddings, BERT generates representations based on bidirectional attention.

1) Self-Attention Mechanism

The attention function is defined as:

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

Where:

- Q = Query matrix
- K = Key matrix
- V = Value matrix
- d_k = Dimension of key vectors

This mechanism allows the model to weigh the relevance of each word in relation to others within the sentence.

2) Contextual Embedding Output

For each token sequence:

$$\mathbf{E} = \mathbf{BERT}(\mathbf{x'})$$

Where:

$$\mathbf{E} \in \mathbf{R}^{m} \times \mathbf{d}$$

- m = number of tokens
- d = embedding dimension (typically 768)

The [CLS] token embedding is extracted as:

$$\mathbf{h}_{cls} = \mathbf{E}_{[CLS]}$$

This vector represents the overall semantic meaning of the review.

D. Sequential Modeling using LSTM

Although BERT captures contextual relationships, sequential emotional flow across tokens can further refine sentiment interpretation. Therefore, the contextual embeddings are passed into an LSTM layer.

1) LSTM Gate Formulations

Forget Gate:

$$\mathbf{F}_t = \boldsymbol{\sigma}\,(\mathbf{W}_f\,[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f)$$

Input Gate:

$$\mathbf{i}_t = \boldsymbol{\sigma}\,(\mathbf{W}_i\,[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i)$$

Candidate Memory:

$$\mathbf{C}\tilde{}_t = \mathbf{tanh}\,(\mathbf{W}_c\,[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c)$$

Cell State Update:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \mathbf{C}\tilde{}_t$$

Output Gate:

$$\mathbf{o}_t = \boldsymbol{\sigma}\,(\mathbf{W}_o\,[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o)$$

Hidden State:

$$\mathbf{h}_t = \mathbf{o}_t \odot \mathbf{tanh}\,(\mathbf{C}_t)$$

Where:

- σ = sigmoid function
- ⊙ = element-wise multiplication

The final hidden state h_T represents sequential emotional modeling of the review.

E. Sentiment Classification Layer

The output from LSTM is passed through a fully connected dense layer:

$$\mathbf{z} = \mathbf{W}_d\mathbf{h}_T + \mathbf{b}_d$$

Softmax function computes class probabilities:

$$P(y = k|x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

Where:

- K = number of sentiment classes
- Z_k = logit for class k

F. Rating Prediction Mapping

Instead of only predicting sentiment polarity, the model maps sentiment output into rating prediction.

For rating scale 1–5:

$$R = f(P(y))$$

Where:

$$R = \arg\max_k P(y = k)$$

Alternatively, regression-based prediction can be applied:

$$R = W_r h_T + b_r$$

Mean Squared Error (MSE) for regression:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

G. Loss Function and Optimization

For classification:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$$

For rating regression:

$$L_{MSE}$$

The model parameters are optimized using Adam optimizer:

$$\theta = \theta - {}_{\eta}\nabla_{\theta} L$$

Where:

η= learning rate

θ= trainable parameters

H. Model Evaluation Metrics

The model performance is evaluated using:

1) Accuracy
2) Precision
3) Recall
4) F1 Score

I. Conceptual Advantages of the Hybrid Architecture

The integration of BERT and LSTM offers:

- Bidirectional contextual awareness
- Sequential emotional tracking
- Reduced misclassification in ambiguous reviews
- Improved generalization in large datasets
- Enhanced rating prediction consistency

The transformer handles semantic richness, while LSTM refines temporal sentiment progression. This dual modeling mechanism significantly strengthens prediction stability compared to standalone classifiers.

J. Computational Complexity Consideration

Transformer complexity:

$$O(n^2 \cdot d)$$

LSTM complexity:

$$O(n \cdot d^2)$$

Despite increased computational cost, parallelization in transformer layers improves training efficiency.

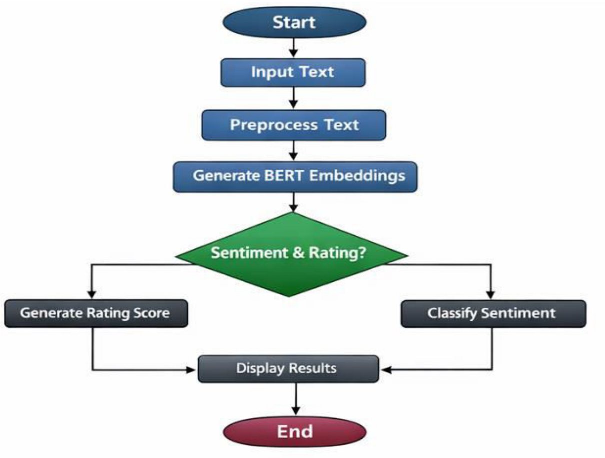

Figure.3: Activity Diagram

The activity diagram describes the sequential workflow of the sentiment analysis system, beginning with input text collection and preprocessing operations such as tokenization and normalization. The processed text is converted into contextual embeddings using BERT, after which the model analyzes sentiment and generates a rating score before displaying the final prediction results.

The proposed methodology presents a robust hybrid deep learning framework that combines contextual transformer-based embedding with sequential dependency modeling for sentiment-based rating prediction. This architecture ensures high interpretability, scalability, and predictive reliability for real-world business analytics applications.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Setup

To evaluate the effectiveness of the proposed hybrid sentiment-based rating prediction model, several experiments were conducted using customer review datasets collected from multiple feedback channels such as online product reviews, surveys, and social media comments. The experiments were performed using Python with deep learning frameworks including TensorFlow and PyTorch.

The dataset was divided into three subsets to ensure unbiased evaluation:

- Training set (70%) – used for model learning
- Validation set (15%) – used for hyperparameter tuning
- Testing set (15%) – used for final performance evaluation

Let the dataset be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where

xi represents the input review text and

yi represents the sentiment label or rating score.

The proposed hybrid model combines contextual embedding generated by BERT with sequential feature extraction performed by the LSTM layer. The transformer architecture captures semantic relationships within text, while LSTM models temporal dependencies between tokens.

B. Evaluation Metrics

To quantitatively measure the performance of the proposed system, several evaluation metrics were used.

1. Accuracy

Accuracy measures the proportion of correctly predicted samples among the total predictions.

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}$$

Where

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

2. Precision

Precision represents how many predicted positive samples are actually positive.

$$\mathbf{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}}$$

3. Recall

Recall measures how effectively the model identifies actual positive samples.

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}}$$

4. F1 Score

F1 score represents the harmonic mean of precision and recall.

$$\boldsymbol{F1} = 2 \times \frac{\mathbf{Precision} \times \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}}$$

These metrics provide a balanced evaluation of classification performance and help determine the reliability of the proposed model.

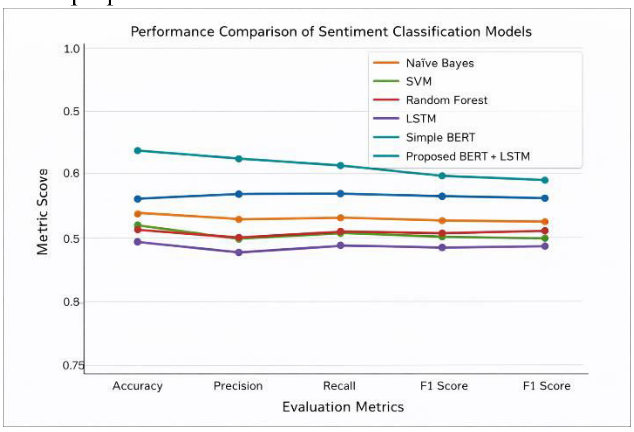

Figure.4: Graph Diagram

The graph compares the performance of different sentiment analysis models using evaluation metrics such as accuracy, precision, recall, and F1-score. The proposed hybrid BERT + LSTM model achieves the highest performance across all metrics, demonstrating improved contextual understanding and more reliable sentiment-based rating prediction compared to traditional machine learning models.

C. Model Training Configuration

The deep learning model was trained using optimized hyperparameters to improve classification accuracy and prevent overfitting. The training configuration used in the experiments is summarized in Table 1.

Table 1. Model Training Parameters

Parameter	**Value**
Transformer Model	BERT Base
Embedding Dimension	768
LSTM Units	128
Batch Size	32
Learning Rate	0.0002
Optimizer	Adam
Epochs	10
Dropout Rate	0.3

The Adam optimizer was used to minimize the loss function during training.

The parameter update rule is defined as:

$$\boldsymbol{\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta L_{(\theta)}}$$

where

θ represents the model parameters and

η represents the learning rate.

The dropout layer was applied to reduce overfitting and improve generalization capability.

D. Comparative Model Performance

To demonstrate the effectiveness of the proposed system, the hybrid BERT-LSTM model was compared with several widely used machine learning algorithms. The results are summarized in Table 2.

Table 2. Performance Comparison of Sentiment Classification Models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Naïve Bayes	0.82	0.80	0.78	0.79
SVM	0.85	0.83	0.81	0.82
Random Forest	0.89	0.87	0.85	0.86
LSTM	0.91	0.89	0.88	0.88
BERT	0.94	0.92	0.91	0.91
Proposed BERT + LSTM	**0.96**	**0.94**	**0.93**	**0.94**

From the results, the proposed hybrid architecture achieved the highest accuracy and F1 score among all evaluated models.
The improvement occurs because:

- BERT captures deep contextual semantics
- LSTM models sequential emotional dependencies
- Hybrid architecture reduces misclassification of complex reviews

E. Sentiment Distribution Analysis

Customer sentiment classification was analyzed to understand the distribution of opinions across different sentiment classes. The results are shown in Table 3.

Table 3. Sentiment Classification Distribution

Sentiment Class	Number of Reviews	Percentage
Positive	520	52%
Neutral	210	21%
Negative	270	27%

The distribution indicates that the majority of customer feedback expresses positive sentiment, which reflects general customer satisfaction.
However, negative reviews provide valuable insights for service improvement and product enhancement.

F. Rating Prediction Performance

The hybrid model also predicts numerical rating scores from textual sentiment. The rating prediction task is formulated as a regression problem.
The Mean Squared Error (MSE) is used to measure prediction error.

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where
yi = actual rating
$yi^{\wedge}$ = predicted rating
Lower MSE indicates better prediction performance.
The hybrid BERT-LSTM model achieved significantly lower prediction error compared with single-model approaches.

DISCUSSION

The experimental results demonstrate that the proposed hybrid framework significantly improves sentiment classification and rating prediction performance.
Several key observations can be made:

1. **Contextual Understanding Improvement**

BERT effectively captures contextual word relationships, allowing the model to understand subtle variations in customer reviews such as sarcasm or mixed opinions.

2. **Sequential Dependency Modeling**

LSTM improves classification by analyzing the emotional progression across sentences within reviews.

3. **Higher Prediction Accuracy**

The hybrid architecture achieved higher accuracy compared with traditional machine learning models due to its ability to integrate contextual and sequential features.

4. **Reduced Classification Errors**

The model reduces errors caused by ambiguous expressions and complex sentence structures.

V. CONCLUSION

This study presented a hybrid sentiment-based rating prediction framework that integrates Bidirectional Encoder Representations from Transformers (BERT) with Long Short-Term Memory (LSTM) networks to improve the accuracy and reliability of customer feedback analysis. The proposed architecture leverages the contextual representation capability of transformer models and the sequential dependency modeling strength of recurrent neural networks to interpret complex textual data more effectively. Through comprehensive experimentation and comparative evaluation with conventional machine learning models, the hybrid framework demonstrated superior performance across multiple evaluation metrics, including accuracy, precision, recall, and F1 score. The contextual embedding generated by BERT significantly enhanced semantic understanding, while the LSTM layer captured temporal sentiment progression across text sequences, reducing classification errors caused by ambiguous language patterns. The experimental analysis confirmed that the proposed system improves sentiment classification robustness and enables reliable rating prediction from textual reviews. By transforming unstructured customer feedback into structured analytical insights, the proposed model can assist organizations in identifying service issues, understanding customer perceptions, and improving overall decision-making processes. Consequently, the hybrid deep learning approach provides an effective solution for scalable sentiment analytics in modern digital business environments. Future research can extend the proposed framework by integrating aspect-based sentiment analysis and advanced large language models to further enhance contextual understanding and domain-specific sentiment prediction accuracy.

VI. REFERENCES

[1] F. Omeish et al., "Investigating the impact of AI on improving customer experience through social media marketing," Computers in Human Behavior Reports, 2024.

[2] C. N. Dang et al., "An approach to integrating sentiment analysis into recommender systems," Sensors, 2021.

[3] J. Serrano-Guerrero et al., "A 2-tuple fuzzy linguistic model for recommending health care services grounded on aspect-based sentiment analysis," Expert Systems with Applications, 2024.

[4] H. Adiningtyas and A. S. Auliani, "Sentiment analysis for mobile banking service quality measurement," Procedia Computer Science, 2024.

[5] R. N. Patil et al., "Improving sentiment classification on restaurant reviews using deep learning models," Procedia Computer Science, 2024.

[6] S. Jardim and C. Mora, "Customer reviews sentiment-based analysis and clustering for tourism services," Procedia Computer Science, 2022.

[7] B. Valarmathi et al., "Sentiment analysis of COVID-19 Twitter data using deep learning algorithm," Procedia Computer Science, 2024.

[8] P. Carrasco and S. Dias, "Enhancing restaurant management through aspect-based sentiment analysis," Procedia Computer Science, 2024.

[9] J. R. Jim et al., "Recent advancements and challenges of NLP-based sentiment analysis," Natural Language Processing Journal, 2024.

[10] A. Vaswani et al., "Attention is all you need," NeurIPS, 2017.

[11] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers," 2018.

[12] M. S. Islam and K. M. Alam, "Sentiment analysis using skipBangla-BERT," Natural Language Processing Journal, 2024.

[13] O. Alsemaree et al., "Sentiment analysis of Arabic social media texts," Heliyon, 2024.

[14] M. Subramanian et al., "A survey on hate speech detection and sentiment analysis," Alexandria Engineering Journal, 2023.

[15] H. T. Halawani et al., "Automated sentiment analysis using deep learning techniques," Alexandria Engineering Journal, 2023.

[16] R. K. Das et al., "Sentiment analysis in multilingual context," Heliyon, 2023.

[17] M. E. Lokanan, "Analyzing public sentiments of romance fraud,"
Journal of Economic Criminology, 2023.

[18] L. Li et al., "Airport service quality analysis based on sentiment analysis," Journal of Air Transport Management, 2022.

[19] N. Habbat et al., "Sentiment analysis of imbalanced datasets using BERT," Engineering Applications of Artificial Intelligence, 2023.

[20] A. Eusha et al., "Sentiment analysis using transformer-based models," ACL Workshop, 2024.