VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2024, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 12, Number 2, July-December 2024 *pp: 285-297*

Empowering Sentiment Analysis with Deep Learning Model: Evaluating Social Media's Benefits and Drawbacks

Noor Fatima [1], **Majdah Alvi** [1], **Muhammad Bux Alvi** [1*]

[1] Department of Computer Systems Engineering, Faculty of Engineering, The Islamia University of Bahawalpur, Pakistan

Keywords: *Deep Learning, LSTM Networks, Natural Language Processing (NLP), Sentiment Analysis, Social Media Impacts, Word2Vec*

Journal Info:
Submitted:
December 01, 2024
Accepted:
December 15, 2024
Published:
December 31, 2024

Abstract Online social networks (OSNs) have revolutionized communication by facilitating unprecedented information sharing and global connections. Despite these benefits, OSNs also present significant challenges, including the spread of misinformation, increased distraction, and adverse mental health effects. This study examines a dataset of 3,904 user reviews collected from online sources and personal networks, revealing a polarized sentiment distribution with 56% positive, 43.1% negative and 0.9% neutral views on the impact of social platforms. To capture the nuanced sentiments expressed, Long Short-Term Memory (LSTM) enhanced with preprocessing techniques such as tokenization, lemmatization, and word embeddings with Word2Vec was employed. The LSTM model achieved an accuracy of 86.43% in sentiment classification, significantly outperforming traditional baseline methods. These findings provide valuable information for platform developers, policymakers, and researchers aiming to understand and mitigate the social and psychological effects of digital platforms. Future research will focus on expanding the dataset and addressing class imbalance to further refine and enhance sentiment analysis models.

***Correspondence author email address:** mbalvi@iub.edu.pk
DOI: 10.21015/vtcs.v12i2.2041

1 Introduction

As of October 2024, global Internet users exceeded 5.52 billion, accounting for approximately 67.5% of the world population [1]. This underscores the ubiquitous role of social networks in daily life, where millions of posts, tweets, reviews, and comments are generated daily [2]. These vast amounts of user-generated content serve as a rich source of data on public attitudes and opinions. For businesses, policymakers, and researchers, analyzing such data is critical for understanding consumer sentiment, monitoring trends, and making data-driven decisions [4]. However, the diverse, informal, and unstructured nature of social media data poses significant challenges to effective analysis [3].

Sentiment analysis, a subfield of Natural Language

Processing (NLP), addresses these challenges by identifying and classifying the emotional tone within written content and categorizing opinions as positive, negative, or neutral. While traditional sentiment analysis approaches, such as lexicon-based methods or machine learning models, struggle with the complexities of social media discourse [5]. Features such as informal language, slang, sarcasm, and ambiguous expressions undermine the accuracy and reliability of these approaches [6].

This paper addresses these limitations by employing Long Short-Term Memory (LSTM) networks, a deep learning architecture suited to modeling sequential dependencies in text [7]. Integrating comprehensive preprocessing methods and Word2Vec embeddings with LSTM processing enables effective sentiment classification in informal text, preserving contextual meaning through vector representations.

1.1 Motivation and Significance

Social media plays a dual role in contemporary society, facilitating communication, business promotion, and educational opportunities while also raising concerns about misinformation, addiction, and mental health issues [8]. Public sentiment towards social networks reflects this duality, with users expressing strong appreciation and critical viewpoints. Accurate capture and interpretation of these mixed feelings is essential to improve platform design, address user concerns, and foster healthier digital environments.

Despite the widespread use of social media, sentiment analysis research has predominantly focused on domains such as product reviews or movie ratings, leaving social platforms underexplored [9]. This study fills this gap by analyzing 3,904 user reviews to assess the benefits and drawbacks of social media. The analysis reveals a polarized distribution of sentiments. This imbalance underscores the substantial influence of social media on users' perceptions and behaviors. By applying LSTM networks to this domain, the research advances methodological approaches in sentiment analysis and offers actionable insights to improve digital platforms.

1.2 Research Questions

This study seeks to address the following key research questions:

1. How do semantic word embeddings, such as Word2Vec, enhance sentiment classification accuracy, particularly for nuanced expressions in social media content?
2. How does the performance of LSTM networks compare with traditional sentiment analysis approaches in processing user-generated text?
3. What role does automated hyperparameter optimization, such as Optuna, play in improving the robustness of sentiment analysis models?

1.3 Contributions

This study makes key contributions in response to the research questions described above.

1. **Improved Sentiment Predictions through Word2Vec:** Integrating Word2Vec embeddings significantly boosted the model's ability to discern and interpret subtle expressions, such as sarcasm and ambiguous language. This enhancement led to more reliable sentiment predictions, reflecting a deeper understanding of the context within user-generated content.
2. **Superior Deep Learning Model:** The deployment of LSTM networks resulted in a substantial increase in sentiment classification accuracy, achieving 86.43% compared to the baseline lexicon-based model of 70.00%. This demonstrates LSTM's superior ability to model temporal dependencies and capture intricate patterns within social media text.
3. **Optimized Model Performance:** Leveraged Optuna for hyperparameter tuning, achieving enhanced robustness evidenced by higher Macro F1-scores and lower loss values across all sentiment classes.

2 Literature Review

Sentiment analysis has significantly advanced with the integration of advanced machine learning models. This section highlights pivotal studies that utilize different deep learning methods. Table 1 summarises related work, detailing the datasets used, methodologies employed, and the merits and limitations identified in each study.

Table 1. Summary of Key Studies on Deep Learning for Sentiment Analysis

Authors	Techniques Used	Dataset	Advantages	Limitations
Zhao and Liu (2021) [7]	CNNs, LSTMs	Social Media Posts	Improved performance over traditional methods	Limitations in handling context and long-range dependencies
Alvi et al. (2024) [9]	Xtractor (Two-step Tweet Extractor)	Twitter Data	Enhanced sentiment analysis by preprocessing and filtering	Specific to Twitter; may not generalize to other platforms
Bilen and Horasan (2021) [24]	LSTM Networks	Customer Reviews	High accuracy in sentiment prediction	Potential overfitting; model complexity
Alvi et al. (2023) [10]	RoSET, RBRS3 (Rule-based Scorer)	4,500 bilingual social media sentences	Improved sentiment detection (Positive: +20.8%, Negative: +16%); addressed lack of labeled datasets	Informal language processing challenges; sentiment lexicons may lack domain-specific terms
Alamoudi and Alghamdi (2021) [19]	Deep Learning, Word Embeddings	Yelp Reviews	Higher accuracy in sentiment classification and aspect detection	Computationally intensive due to deep architectures
Asudani et al. (2023) [35]	Analysis of Word Embedding Models	Various Text Datasets	Contextual embeddings improved performance	Increased computational costs
Chowdhury et al. (2021) [4]	Advanced Preprocessing Techniques	Social Media Data	Highlighted challenges and proposed solutions for sentiment analysis on noisy data	Primarily discusses challenges and lacks experimental validation
Liao (2024) [18]	Double Classification, LSTM Models	Takeaway Platform Reviews	Enhanced sentiment accuracy for domain-specific nuances; Improved understanding of customer feedback in the food delivery industry	Limited accessibility due to future publication date; May not generalize well to other domains
Alsharef et al. (2022) [23]	LSTM, Word Embedding	Social Media Data	Automated toxicity detection; Supported content moderation strategies on social platforms	Retracted article; Potential data/methodology issues; Findings may not be reliable
Alharbi et al. (2021) [28]	Word Embeddings, RNN Variants	Amazon Reviews	Hybrid models improved accuracy	Requires significant computational resources
Idrovo-Berrezueta et al. (2024) [29]	LSTM, Linear Regression, SVM	Social Network Comments (Ecuador)	Enhanced sentiment detection in corruption-related comments	Required extensive preprocessing due to dialectal variations and colloquial language
Diwan (2020) [22]	RNNs on Multimodal Inputs	Textual, Visual Data	Comprehensive sentiment understanding	Needs more comparative analyses
Gao et al. (2020) [8]	Statistical Analysis of Survey Data	Online survey from 4,872 participants during COVID-19 outbreak	Identified the impact of social media exposure on mental health; Provided insights during a public health crisis	Cross-sectional study limits causal inference; Self-reported data may have bias
Tamam and Yanik (2021) [11]	Hybrid ML-Lexicon Approaches	Social Media Data (e.g., Twitter)	Improved sentiment analysis accuracy; Combined strengths of multiple techniques	Increased computational complexity; May not generalize well to all datasets
Muhammada et al. (2020) [21]	LSTM with Word2Vec	Indonesian Hotel Reviews	Achieved 85% accuracy; detailed evaluations	Restricted to single dataset; lacks diversity

3 Proposed Deep Learning Architecture

This research presents a comprehensive sentiment analysis framework comprising three main components: advanced preprocessing, optimized LSTM architecture, and systematic hyperparameter tuning. Figure 1 illustrates the overall workflow.

3.1 Data

This research utilizes semi-automated Python-based extraction methods to analyze social media impacts through a dataset of 3,904 user reviews collected from personal networks and online platforms, including Kaggle, Facebook, and various forums. The dataset categorizes reviews into three sentiment classes: positive, negative, and neutral. Positive reviews express benefits such as knowledge enhancement, Negative reviews highlight concerns about misuse, and Neutral reviews acknowledge balanced perspectives or lack of awareness on the topic. Table 2 shows a few samples from the dataset.

Table 2. Sample Data Entries from the Social Media Reviews Dataset

Review ID	Review Text	Sentiment
109	"Every work is completed in a short time due to social media"	positive
3291	Social media facilitates people in wrong activities which can lead them to jail	negative
1155	With social networking sites, you exchange information easily in real-time via a chat	positive
3899	"Social media has good and bad sides. I try to use it carefully."	neutral

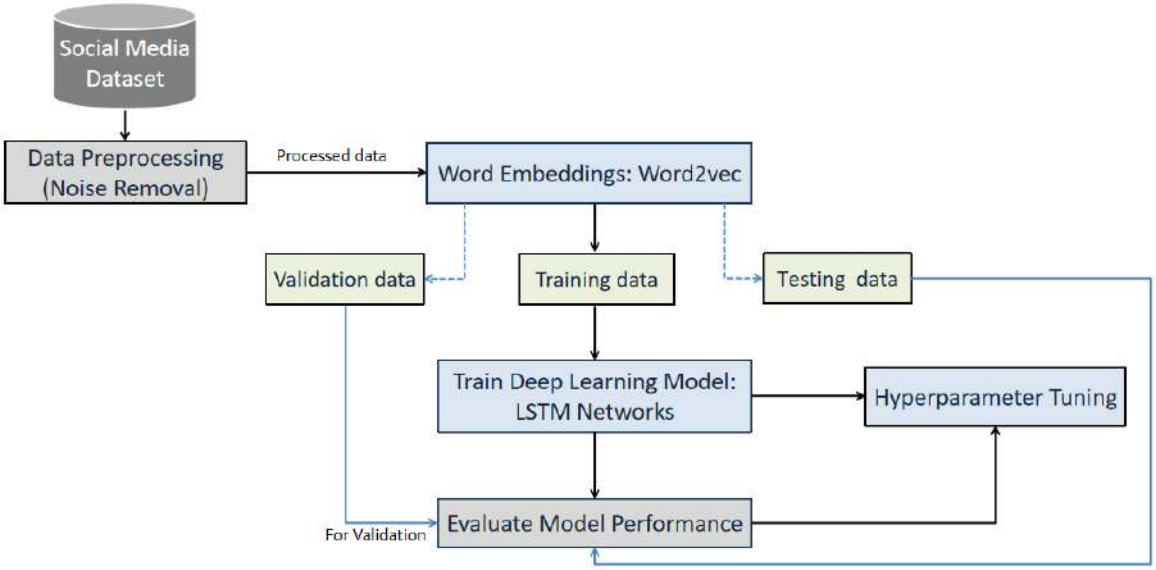

Figure 1. Deep Learning Model Architecture to effectively classify sentiments from social media reviews

The dataset's significance lies in capturing real-world perspectives, sentiment diversity, and quantitative and qualitative analysis. The class imbalance presents unique challenges for model training and evaluation, requiring specialized preprocessing approaches to ensure suitability.

Figure 2. Word Cloud Visualization for Social Media Impact Analysis

A Word Cloud visualization was created to highlight the most prominent words associated with user reviews' sentiments, as shown in Figure 2. This visualization aids in identifying common themes and emphasizing the prevalence of specific terms within each sentiment category.

3.2 Data Preprocessing: The Foundation of Robust Sentiment Analysis

The effectiveness of any sentiment analysis system depends on meticulous preprocessing of textual data to ensure accuracy and relevance. This study's preprocessing involved a carefully designed pipeline using advanced techniques and Python libraries such as `NLTK`, `re`, `contractions`, `emoji`, and `spellchecker`. These steps were tailored specifically to address noise and inconsistencies inherent in social media reviews while preserving context and sentiment [11].

- **Contraction Expansion:** To address the informal nature of social media language, contractions (e.g., "can't") were expanded using the `contractions` library.
- **URL and Hashtag Removal:** Social media reviews often contain non-linguistic noise, such as URLs and hashtags. These elements were removed using powerful regular expressions (`re.sub`) to focus the analysis on the textual content.
- **Incorporating Emojis:** Emojis, which convey strong emotional tones, were either removed using the `emoji` library or mapped to sentiment scores to enhance contextual understanding.

- **Punctuation Normalization:** Punctuation and special characters, such as commas, full stops, and quotation marks, were cleaned using a comprehensive regular expression-based approach [12]. This step ensured consistency across all text data.
- **Tokenization:** The cleaned text was split into tokens (words) using `word_tokenize` from NLTK to facilitate granular processing, enabling the model to analyze individual words and their relationships within the text. e.g., `['people', 'health', 'effect', 'due', 'social', 'media']`
- **Lowercasing:** Tokens were converted to lowercase to ensure that variations in case (e.g., "Social" vs. "social") did not affect the analysis.
- **Stopword Removal:** Non-informative words such as "is" and "the" were removed using `stopwords` from NLTK. This significantly reduced textual noise, allowing the model to focus on more meaningful words that contribute to sentiment classification.
- **Lemmatization:** Each word was reduced to its base form using `WordNetLemmatizer`, relying on Part-of-Speech (POS) tagging to preserve linguistic context. For instance, "running" was reduced to "run" [13].
- **Spelling Correction:** Erroneous spellings were corrected using `SpellChecker` from the spellchecker library, enhancing the quality and accuracy of tokenized data [14].
- **Mapping and Label Cleaning:** Labels were cleansed of extraneous characters and grouped into actionable sentiment categories—positive, negative, and neutral—using Python's `applymap` function.

Table 3. Dataset Statistics After Preprocessing

Metric	**review**	**sentiment**	**cleaned_review**	**lemmatized_review**	**corrected_review**
Count	3904	3904	3904	3904	3904
Unique	3685	3	3333	3304	3286
Top	wastage of time	positive	[wastage, time]	wastage time	wastage time
Freq	3	2187	10	10	10

As illustrated in Table 3, the preprocessing pipeline effectively cleaned and standardized the initial dataset of 3,904 social media reviews. The key transformations included cleaning (removing URLs, hashtags, etc.), tokenization, lemmatization (reducing words to their base form), and spell correction. This resulted in a refined dataset with 3,286 unique reviews after correction, a significant noise reduction and improved consistency for subsequent analysis. The most frequent review phrase, "wastage of time," appeared 10 times post-processing, highlighting the impact of the cleaning steps. The sentiment labels were also standardized into three categories (positive, negative, and neutral), with a majority of 2,187 positive labels. These preprocessing steps significantly enhanced the consistency and quality of our dataset by reducing noise.

3.2.1 Innovations and Challenges in Data Preparation

While traditional preprocessing frameworks may suffice for uniform datasets, social media reviews introduce unique complexities:

- **Richness of Emojis:** Emojis serve as emotional amplifiers in social media conversations. Instead of outright removal, future iterations of our model could integrate pre-annotated emoji sentiment scores for greater precision [15].
- **Handling Slang and Abbreviations:** Social media jargon, acronyms (e.g., "LOL" or "BTW"), and shorthand (e.g., "u" for "you") highlight areas for improvement. Embedding custom vocabulary mappings could add significant contextual understanding [16].
- **Multilingual Text:** Social media transcends language barriers, often mixing multiple languages in a single review. Expanding the preprocessing pipeline to include multilingual corpora would elevate global sentiment understanding [17].

3.3 Word Embeddings with Word2Vec

Word2Vec, a shallow neural network-based embedding technique, was employed to generate dense vector representations of words to capture the semantic nuances of social media language. This technique facilitates the learning of contextual relationships between words. Using the `gensim` library (v4.3.3), the model was configured with a vector size of 300, a context window of 10 words, and a minimum word count threshold of 1. These parameters were chosen to reflect the linguistic characteristics of social media text, which often consists of short conversational messages. Compared to alternative embedding methods such as GloVe or BERT, Word2Vec offers computational efficiency and semantic interpretability, which are critical for analyzing large, noisy datasets like social media posts[35]. While BERT provides deep contextual embeddings, it is computationally intensive and may be unnecessary for datasets where Word2Vec can effectively capture semantic relationships. GloVe, on the other hand, is a count-based method that may not capture contextual nuances as effectively in short, informal texts prevalent in social media.

3.3.1 Semantic Insights from Word2Vec

The Word2Vec embeddings revealed nuanced semantic relationships in the dataset. Analysis of the term "`people`" highlights its contextual associations and contrasts:

- **Most Similar Words:** Using `w2v_model.wv.most_similar("people", topn=50)`, words such as "talk", "communicate", and "socially" emerged as highly similar. These terms reflect the contextual use of "people" in interaction, social behavior, and communication discussions. Terms like "isolated" and "education" further highlight societal themes, such as isolation or learning environments.
- **Semantic Similarity:** The cosine similarity between "`people`" and "`interact`" was extremely high at 0.965555, highlighting their closely intertwined context in the discourse of human interaction and relationships.
- **Contrasting Terms:** Using `w2v_model.wv.most_similar(negative=["people"])`, contrasting terms such as "unavoidable" and "insomnia", as well as "emotionless" and "risk", were identified, reflecting semantically unrelated or divergent associations. These terms suggest indirect links to non-human or adverse contexts, such as emotional detachment or danger.

Table 4. Word2Vec Similarity Results for "people"

w2v_model.wv.most_similar('people')	w2v_model.wv.most_similar(negative =['people'])
• (interact, 0.9655548334121704)	• (unavoidable, 0.350969523191452)
• (talk, 0.9643059372901917)	• (hypersonic, 0.3029715120792389)
• (socially, 0.959194004535675)	• (impersonate, 0.007015147712081671)
• (communicate, 0.9550158381462097)	• (responsive, -0.003983135335147381)
• (whatsapp, 0.9546887278556824)	• (offensive, -0.004036257974803448)
• (isolated, 0.9534667134284973)	• (insomnia, -0.048521097749471664)
• (global, 0.9507717490196228)	• (tangible, -0.10373144596815109)
• (education, 0.9402336478233337)	• (multitasking, -0.20172405242919922)
• (addict, 0.9399133920669556)	• (emotionless, -0.6086081266403198)
• (emotionally, 0.7472774982452393)	• (risk, -0.7061274647712708)

To visualize the semantic relationships, t-distributed Stochastic Neighbor Embedding (t-SNE) was employed to project embeddings into a two-dimensional space. As depicted in Figure 3, semantically similar words clustered together by t-SNE, providing an intuitive visualization of the model's effectiveness in mapping contextual similarities within social media discourse.

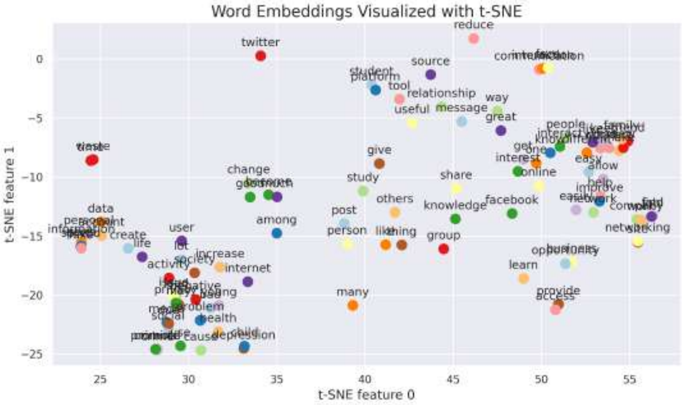

Figure 3. t-SNE Visualization of Word2Vec Embeddings

3.4 Preparation of Data for the Keras Embedding Layer

The dataset was meticulously prepared to integrate Word2Vec embeddings with Keras' neural network framework through the following steps:

3.4.1 Tokenization and Sequencing

Using Keras' `Tokenizer`, text reviews were converted into numerical sequences by mapping each unique word to a distinct integer. This transformation is essential for converting textual data into a format that neural network models can process effectively.

```
tokenizer.index_word
{1: 'social', 2: 'media', 3: 'people', 4: 'use',
5: 'help', 6: 'time', 7: 'make', 8:'information'
, 9: 'world', 10: 'share', ...}
```

3.4.2 Sequence Padding

Sequences were standardized to ensure uniform input lengths using Keras' `pad_sequences`, applying post-padding with zeros. This uniformity is crucial for maintaining consistent input dimensions across the neural network, facilitating efficient batch processing and model training.

```
padded_sequences :
[[  31   19   28 ...    0    0    0]
 [  82    1    2 ...    0    0    0]
 [  22   68    1 ...    0    0    0]
 ...
 [ 285  485    6 ...    0    0    0]
 [   4  140  644 ...    0    0    0]
 [1420   31   45 ...    0    0    0]]
```

3.4.3 Embedding Matrix Construction

A custom embedding matrix was constructed to map vocabulary words to their corresponding pre-trained Word2Vec vectors [21]. The embedding matrix $\mathbf{E} \in \mathbb{R}^{|V| \times d}$ has a shape of `(vocab_size, embed_dim)`, where $|V|$ represents the size of the vocabulary and $d = 300$ denotes the dimensionality of each word vector. Each row in the embedding matrix corresponds to a unique word in the vocabulary, mapping it to its respective Word2Vec vector. This construction leverages the semantic and syntactic information captured by Word2Vec during its training on extensive textual corpora.

3.4.4 Label Encoding and Data Splitting

Sentiment labels were encoded numerically using Scikit-learn's `LabelEncoder` and subsequently transformed into one-hot vectors. One-hot encoding is essential for enabling the neural network to perform multi-class classification. The dataset was then split into training (80%) and testing (20%) sets using `train_test_split`, ensuring that the model has sufficient data to learn from while retaining a representative subset for evaluating its performance on unseen data.

Table 5. Shape of Processed Data

Data	Shape
`x_train` (Padded Sequences)	(3123, 24)
`y_train` (One-Hot Labels)	(3123, 3)
`x_test` (Padded Sequences)	(781, 24)
`y_test` (One-Hot Labels)	(781, 3)

This comprehensive data preparation enhanced the model's capacity to accurately classify sentiments by embedding semantic relationships and ensuring consistency in input formats.

4 Deep Learning Model Architecture

A deep learning architecture was designed to classify sentiments from social media reviews using the `Keras` library. The model leverages an Embedding layer initialized with pre-trained Word2Vec vectors, an LSTM layer to capture sequential dependencies, and Dense layers for classification. Regularization techniques and hyperparameter optimization were employed to enhance performance and generalization.

4.1 Embedding Layer

The embedding layer transforms input words into 300-dimensional vectors, capturing their semantic and syntactic properties. Initialized with a custom embedding matrix $\mathbf{E} \in \mathbb{R}^{|V| \times d}$ derived from pre-trained Word2Vec embeddings [32], this layer leverages rich linguistic information to enhance the model's understanding of language nuances. For example, vector arithmetic in the embedding space can illustrate relationships such as: King – Man + Woman $\approx$ Queen

This capability allows the model to infer meaningful relationships beyond direct word associations, thereby improving sentiment classification accuracy.

4.2 LSTM Layer

An LSTM layer processes the embedded sequences, effectively capturing long-term dependencies and mitigating the vanishing gradient problem inherent in traditional RNNs [27, 33]. The LSTM architecture employs forget, input, and output gates to regulate information flow:

$$
\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\
\tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\
h_t &= o_t \odot \tanh(C_t),
\end{aligned}
$$

where f_t, i_t, and o_t are the forget, input, and output gates, respectively; C_t is the cell state; h_t is the hidden state; and x_t is the input at time step t.

4.3 Dense Layers

Following the LSTM layer, two fully connected dense layers are employed. The first Dense layer captures learned features, while the second uses a `softmax` activation function to output probability distributions over the three sentiment classes:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \tag{1}$$

where p_i is the predicted probability for class i, and z_i is the input to the softmax function for class i.

4.4 Loss Function

Categorical cross-entropy was utilized as the loss function to measure the divergence between true labels and predicted probabilities:

$$\mathcal{L} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i), \tag{2}$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability for class i.

4.5 Optimizer

The Adam optimizer was selected for its adaptive learning rate capabilities, combining the benefits of momentum and RMSProp[19]. Its update rules are defined as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad \hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \tag{3}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \tag{4}$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{5}$$

where g_t is the gradient at time t, θ_t are the parameters, α is the learning rate, and β_1, β_2, and ϵ are hyperparameters controlling the optimization process.

4.6 Regularization Techniques

To prevent overfitting and enhance generalization, the following regularization methods were used [34]:

- **Dropout:** Randomly deactivates 20% of neurons during training, reducing interdependent learning and improving feature robustness:
  $\tilde{h}_i = h_i \cdot r_i, \quad r_i \sim \text{Bernoulli}(p = 0.2)$,
  where $\tilde{h}_i$ is the output after dropout, h_i is the original activation, and r_i is the dropout mask.
- **L2 Regularization ($\lambda = 0.01$):** Adds the squared magnitude of weights to the loss function, promoting smaller weight values:
  $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{model}} + \lambda \sum_j w_j^2$,
  where $\mathcal{L}_{\text{total}}$ is the total loss, $\mathcal{L}_{\text{model}}$ is the original loss, and w_j are the network weights.
- **Early Stopping:** Monitors validation loss and halts training if no improvement is observed over 3 consecutive epochs, preventing overfitting and saving computational resources.

4.7 Hyperparameter Optimization

Optuna, an automatic hyperparameter optimization framework, was employed to optimize the model's hyperparameters. Utilizing the Tree-structured Parzen Estimator (TPE) algorithm, Optuna conducted 12 trials within the defined search space to identify the configuration that maximizes validation performance. Table 6 shows the optimal hyperparameters.

Table 6. Optimal Hyperparameters Identified by Optuna

Hyperparameter	Optimal Value
Learning Rate	0.001
Number of LSTM Units	107
Dropout Rate	0.2
Batch Size	48
L2 Regularization Coefficient	0.01

These optimized hyperparameters enhanced the model's ability to classify sentiments accurately while maintaining computational efficiency.

5 Results and Analysis

This section presents the findings of the study, demonstrating the improvements achieved through advanced deep learning techniques in sentiment analysis of social media data, and highlighting the benefits and drawbacks revealed by analyzing social media reviews.

5.1 Impact of Data Preprocessing and Word Embeddings

Rigorous data preprocessing played a pivotal role in enhancing model performance. The initial dataset comprised 4,376 unique tokens, which was reduced to 2,535 tokens after removing noise and redundancies. This preprocessing step significantly streamlined the training process and improved word representations.

Table 7. Impact of Data Cleaning on Sentiment Analysis Performance

Dataset Version	Accuracy	Weighted Precision	Weighted F1-Score
Raw Data	78.34%	77.89%	76.53%
Cleaned Data	**86.43%**	**84.34%**	**84.92%**

Table 7 shows that transitioning from raw to cleaned data resulted in a notable increase in all key performance metrics, underscoring the importance of data preprocessing. Incorporating Word2Vec embeddings further enhanced the model's ability to capture semantic and contextual nuances, reflecting the sentiment complexity inherent in social media data.

5.2 Effectiveness of LSTM Networks

The performance of the LSTM model was compared to a baseline lexicon-based model for sentiment classification [26, 27].

Table 8. Performance Comparison Between LSTM Network and Baseline Model

Model	Accuracy
Baseline Lexicon-Based Model	70.00%
LSTM Networks (Proposed Model)	86.43%

Table 8 shows the LSTM model achieved a significantly higher accuracy 86.43% compared to the lexicon-based model 70.00%. The results confirm that LSTM networks are highly effective for classifying social media sentiments. Their strength lies in capturing sequential dependencies and contextual nuances in text data, significantly enhancing their performance.

5.3 Model Performance and Optimization

Different LSTM model configurations were evaluated to determine an optimal balance between accuracy, robustness, and generalizability.

Table 9. Performance Comparison Across Model Variants

Model	Accuracy	Macro F1	Weighted F1	Loss
Baseline LSTM	85.28%	58.09%	80.00%	0.3847
LSTM + Dropout + L2	86.30%	57.58%	83.00%	0.4012
LSTM + EarlyStopping	**87.32%**	57.71%	83.00%	0.3923
Optuna-Tuned LSTM	86.43%	**67.33%**	**84.00%**	**0.3556**

The experimental results in Table 9 reveal several key insights:

- **Baseline LSTM Model**: Achieved an accuracy of 85.28%, validated the effectiveness of the fundamental LSTM architecture in capturing sequential patterns in sentiment data.
- **LSTM + Dropout + L2 Regularization**: Improved accuracy to 86.30%, indicated that incorporating regularization techniques could enhance model generalization by preventing overfitting.
- **LSTM + EarlyStopping**: Achieved the highest accuracy of 87.32%, demonstrating that early stopping effectively enhanced model performance by halting training before overfitting occurs.
- **Optuna-Tuned LSTM**: While achieving a comparable accuracy of 86.43%, the Optuna-tuned model attained the highest Macro F1-score of 67.33% and the lowest loss of 0.3556, suggesting that hyperparameter optimization through Optuna enhances the model's balance across classes and overall robustness.

Figure 4 shows the training and validation loss curves for the optimized LSTM model, demonstrating effective learning and convergence with minimal overfitting. The model's stability is a key indicator of its ability to generalize well to unseen social media reviews.

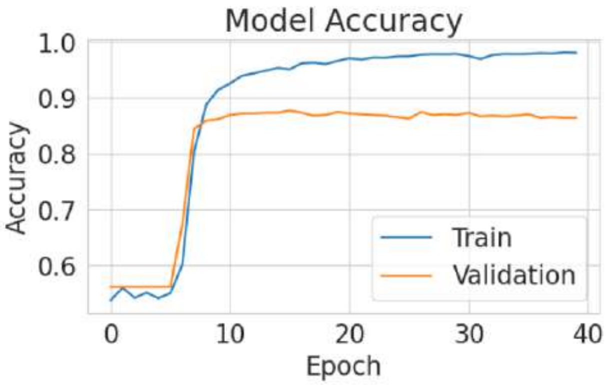

Figure 4. Training and Validation Loss Curves of the Optimized LSTM Model

5.4 Sentiment Class Performance: Benefits and Drawbacks

To provide a comprehensive evaluation, Table 10 presents performance metrics broken down by sentiment class (positive, negative, neutral).

Table 10. Sentiment Class Performance for the Optuna-Tuned LSTM Model

Metric	Positive	Negative	Neutral
Precision	0.87	0.90	0.60
Recall	0.90	0.88	0.15
F1-Score	0.88	0.89	0.24

This analysis shows that the model exhibits high performance in classifying positive and negative sentiments but struggles with the neutral class. The low recall and F1-score for the neutral class suggest a potential drawback of the model, highlighting a need for more balanced data to improve performance across all sentiment classes. The class imbalance in the dataset is likely the primary factor driving the model's lower performance for neutral sentiments, leading to higher uncertainty in determining truly neutral opinions within social media reviews.

The confusion matrix (Figure 5) validates the performance metrics in Table 10, particularly the model's comparatively lower performance on the neutral sentiment class.

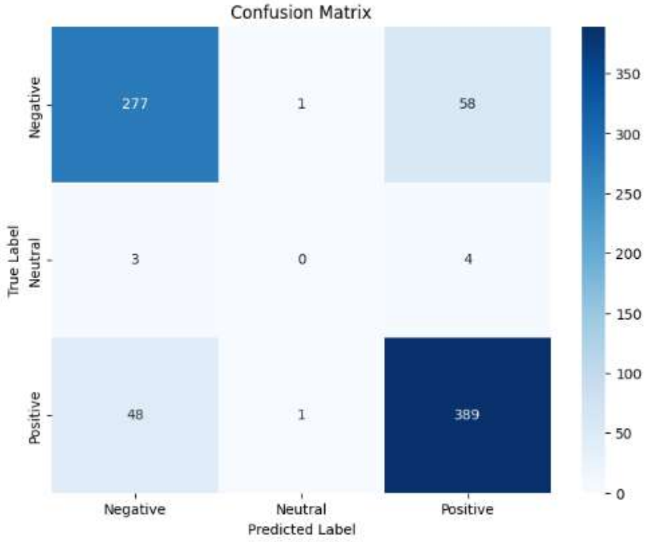

Figure 5. Confusion Matrix for Sentiment Classes

6 Discussion and Conclusion

This study highlights the significant advancements in sentiment analysis achieved through advanced deep learning techniques, specifically in analyzing social media reviews to assess their benefits and drawbacks. Key findings include:

- **Data Preprocessing and Word Embeddings** significantly enhanced model performance by reducing noise and capturing semantic relationships.
- **LSTM Networks** effectively classified social media sentiments, outperforming traditional lexicon-based methods.
- **Model Optimization** through regularization techniques and hyperparameter tuning improved generalization and robustness.

Analyzing social media reviews revealed benefits (e.g., facilitating productive communication and business growth) and drawbacks (e.g., a tendency for addiction and misinformation). The analysis of misclassification patterns shows that sentiment biases, subtle language nuances, and polarization complicate accurate classification. The poor performance in the neutral sentiment class, with a recall of 0.15, reflects the need to address class imbalances and refine techniques for nuanced sentiment detection.

While Word2Vec effectively generalized semantic relationships, it has notable limitations. It lacks the ability to dynamically adapt embeddings based on word context, making it less effective in resolving polysemy or domain-specific terms. This limitation may contribute to the model's challenges in accurately classifying sentiments where context plays a crucial role.

Despite the challenges, the approach presented here has broader implications for real-world industries such as customer feedback analysis, brand monitoring, and public sentiment evaluation, enabling real-time, data-driven decision-making. Overall, this research underscores the dual-edged nature of social media, amplifying both constructive communication and harmful content. The results pave the way for further exploration in this critical area.

7 Future Work

To further enhance the capabilities and applicability of sentiment analysis systems, future research should consider the following:

- **Transition to Contextual Word Embeddings:** Replacing Word2Vec with state-of-the-art contextualized embeddings, such as BERT, ELMo, or RoBERTa, will allow the model to dynamically adapt to context, addressing limitations in handling polysemy and domain-specific language.
- **Expansion of Dataset Scope:** Collecting larger, multilingual, and culturally diverse datasets will enhance model robustness and support multilingual sentiment analysis. Expanding datasets to reflect a wider spectrum of sentiments will address challenges related to class imbalance, particularly for the neutral sentiment class.
- **Mitigating Class Imbalance:** Employ techniques such as oversampling, SMOTE (Synthetic Minority Oversampling Technique), or creating synthetic examples using generative models like GANs. Such efforts will reduce skewness in class performance and enhance the model's ability to capture neutrality.
- **Exploration of Advanced Architectures:** Future studies should investigate the integration of attention mechanisms, transformers, or hybrid models (e.g., CNN-LSTM) to enhance contextual understanding. These architectures are adept at identifying complex patterns, such as sarcasm, irony, and sentiment shifts.
- **Multimodal Analysis and Bias Mitigation:** Extend the framework to include multimodal data (e.g., images, videos, emojis) and investigate methods to identify and mitigate biases, ensuring equitable sentiment predictions across diverse user demographics.

By following these directions, future studies can overcome current limitations, adapt to a wider range of real-world scenarios, and further advance the field of sentiment analysis in social media data.

Author Contributions: Noor Fatima conceptualized the study, performed data preprocessing, developed the models, and drafted the report. Majdah Alvi conceptualized the study, supervised data preprocessing, assisted in model development, and reviewed the initial draft. Dr. Muhammad Bux Alvi provided the seed dataset, conceptualized the study, supervised the research work, reviewed the results, and suggested improvements.

Compliance With Ethical Standards: The authors declare no conflict of interest. This study does not involve human or animal experiments.

References

[1] Statista. (2024). *Number of internet and social media users worldwide as of October 2024* [Online]. Retrieved January 5, 2025, from https://www.statista.com/statistics/617136/digital-population-worldwide/.

[2] Kemp, S. (2023). *Digital 2023: Global Overview Report* [Online]. Retrieved January 4, 2025, from https://datareportal.com/.

[3] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, p. e5107, 2020.

[4] N. R. Chowdhury, M. Z. Islam, and F. Alam, “Challenges of sentiment analysis on social media data,” in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, pp. 1–5.

[5] S. A. Alotaibi, “Challenges in sentiment analysis for Arabic social media,” *Journal of Information Science*, vol. 46, no. 6, pp. 816–827, 2020.

[6] C. J. Hutto and E. Gilbert, “VADER: A parsimonious rule-based model for sentiment analysis of social media text,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 216–225.

[7] K. Zhao and J. Liu, “Deep learning for sentiment analysis of social media,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–9, 2021.

[8] J. Gao, P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, and J. Dai, “Mental health problems and social media exposure during COVID-19 outbreak,” *PloS One*, vol. 15, no. 4, p. e0231924, 2020.

[9] M. B. Alvi, N. A. Mahoto, M. A. Unar, and M. A. Shaikh, “Xtractor: A two-step tweet extractor for sentiment analysis,” in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, 2024, pp. 1–6.

[10] M. B. Alvi, N. A. Mahoto, M. S. A. Reshan, M. Unar, M. Elmagzoub, and A. Shaikh, “Count Me Too: Sentiment Analysis of Roman Sindhi Script,” *SAGE Open*, vol. 13, no. 3, p. 21582440231197452, 2023.

[11] E. Tamam and O. F. Yanik, “Hybrid approaches for sentiment analysis in social media,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 4, pp. 4053–4065, 2021, publisher=Springer.

[12] L. Jiang and W. Zhang, “Deep learning for punctuation normalization in social media text,” *Neural Computing and Applications*, vol. 32, no. 24, pp. 18073–18083, 2020, publisher=Springer.

[13] Y. Tang and H. Wang, “Contextual Lemmatization for Improved Sentiment Analysis,” *Journal of Artificial Intelligence Research*, vol. 70, pp. 1–20, 2021, publisher=AI Access Foundation.

[14] L. Wang and M. Sun, “A Comprehensive Study on Spell Correction Techniques for NLP Applications,” *Expert Systems with Applications*, vol. 185, p. 115598, 2022, publisher=Elsevier.

[15] I. Habernal and I. Gurevych, “Understanding Emoji Usage in Social Media: A Sentiment Analysis Approach,” *Journal of Computer-Mediated Communication*, vol. 25, no. 6, pp. 414–429, 2020, publisher=Oxford University Press.

[16] R. Agarwal and A. Gupta, “SlangNet: A Deep Learning Approach to Handle Slang and Abbreviations in Social Media,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 5678–5688, 2021, publisher=ACL.

[17] P. Meakem and E. Bender, “Multilingual Preprocessing Techniques for Sentiment Analysis in Social Media,” *Computational Linguistics*, vol. 47, no. 2, pp. 345–368, 2021, publisher=MIT Press.

[18] Y. Liao, “Sentiment Analysis by Double Classification of Takeaway Platform Reviews Based on Deep Learning LSTM Models,” *Science and Technology of Engineering, Chemistry and Environmental Protection*, vol. 1, no. 5, 2024.

[19] E. S. Alamoudi and N. S. Alghamdi, "Sentiment classification and aspect-based sentiment analysis on Yelp reviews using deep learning and word embeddings," *Journal of Decision Systems*, vol. 30, no. 2-3, pp. 259–281, 2021.

[20] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 10345–10425, 2023.

[21] R. K. Putra Fissabil Muhammada, Adi Wibowoa Putra Fissabil Muhammada, Retno Kusumaningruma, and Adi Wibowo, “Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews,” *Journal of Information and Communication Technology*, Elsevier, vol. 25, no. 4, pp. 567–580, 2020.

[22] J. V. T. T. Diwan, “Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks,” *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 18, no. 3, pp. 45–52, 2020.

[23] A. Alsharef, K. Aggarwal, S., D. Koundal, D. Alyami, H. Alyami, and D. Ameyed, “[Retracted] An Automated Toxicity Classification on Social Media Using LSTM and Word Embedding,” *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 8467349, 2022.

[24] B. Bilen and F. Horasan, “LSTM network based sentiment analysis for customer reviews,” *Politeknik Dergisi*, vol. 25, no. 3, pp. 959–966, 2021, Gazi University.

[25] M. Xuanyuan, L. Xiao, and M. Duan, “Sentiment classification algorithm based on multi-modal social media text information,” *IEEE Access*, vol. 9, pp. 33410–33418, 2021.

[26] R. Nawaz, H. Afzal, and A. Rauf, “Comparative analysis of deep learning-based techniques and traditional models for sentiment classification in social media,” *Journal of Ambient Intelligence and Humanized Computing*, 2022.

[27] M. Hasnain, U. Ghani, F. Naqvi, and M. Pervaiz, “Comparative performance analysis of deep learning and traditional machine learning models for sentiment analysis,” *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 10663–10686, 2021.

[28] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, “Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews,” *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 5536560, 2021.

[29] P. S. Idrovo-Berrezueta, D. A. Dutan-Sanchez, and G. A. León-Paredes, “Enhancing Sentiment Analysis of Corruption-Related Comments on Social Networks in Ecuador Using Techniques of Machine Learning: LSTM, Linear Regression, and Support Vector Machine,” in *International Conference on Information Technology & Systems*, Springer, 2024, pp. 278–287.

[30] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data," *Information Processing & Management*, vol. 58, no. 1, p. 102435, 2021.

[31] A. Chopra, V. Rajput, D. Pareek, and M. Haarstad, "An efficient text classification framework using LSTM and pre-trained embedding layers on natural language datasets," *Journal of Big Data*, vol. 8, no. 1, pp. 1–13, 2021.

[32] J. Carrell, Z. Zhang, and A. Ignacio, "Evaluation of deep neural networks for sentiment classification in text," *International Journal of Data Science and Analytics*, vol. 12, no. 3, pp. 205–218, 2021.

[33] Z. Wang, Q. Zhao, and K. J. Cios, "Sentiment analysis using LSTM-CNN hybrid architectures and data augmentation," *Information Sciences*, vol. 571, pp. 1–9, 2021.

[34] S. Reddy and S. K. Kavala, "Regularization techniques to improve generalization in sentiment analysis models," *International Journal of Artificial Intelligence and Applications*, vol. 12, no. 5, pp. 11–23, 2022.

[35] D. Asudani, T. Smith, and A. Khanna, "The Impact of Word Embedding Models on Text Analytics: GloVe, Word2Vec, and BERT in Noisy Domains," *Journal of Applied Computational Linguistics*, vol. 15, no. 2, pp. 45–59, 2023.