Deep Learning-Based Sentiment Analysis: Enhancing IMDb Review Classification with LSTM Models

Vivek Varadharajan [1], Nathan Smith [2], Dinesh Kalla [2*], Fnu Samaah [3], Vishwanadham Mandala [4]

[1] Department of Computer Science, Denver University, Colorado, USA

[2] Department of Computer Science, Colorado Technical University, Colorado, USA

[3] Department of Computer Science, Harrisburg University of Science and Technology, PA, USA

[4] Department of AAI, Data Engineer Leader, Cummins Inc, IN, USA

*Correspondence: Dinesh Kalla (kalladinesh@outlook.com)

Abstract: Sentiment analysis, a vital aspect of natural language processing, involves the application of machine learning models to discern the emotional tone conveyed in textual data. The use case for this type of problem is where businesses can make informed decisions based on customer feedback, identify the sentiments of their employees, and make decisions on hiring or retention, or for that matter, classify a text based on its topic like whether it is about a particular subject like physics or chemistry as is useful in search engines. The model leverages a sequential architecture, transforms words into dense vectors using an Embedding layer, and captures intricate sequential patterns with two Long Short-Term Memory (LSTM) layers. This model aims to effectively classify sentiments in text data using a 50-dimensional embedding dimension and 20 % dropout layers. The use of rectified linear unit (ReLU) activations enhances non-linearity, while the SoftMax activation in the output layer aligns with the multi-class nature of sentiment analysis. Both training and test accuracy were well over 80%.

Keywords: Sentiment Analysis, Natural Language Processing (NLP), Customer Satisfaction, Machine Learning, Long Short-Term Memory (LSTM)

1. Introduction

Opinion mining or sentiment analysis is a crucial aspect of natural language processing (NLP). It involves identifying and classifying the opinions expressed in text. This technique is widely used for data mining of customer satisfaction surveys, social networking sites, product ratings, and other text data to gain insight into public sentiment. The potential of sentiment analysis to provide data on consumer behaviors, enhance customer experiences, and inform strategic decisions is a promising development in the field of NLP (Agarwal et al., 2020) [1]. However, earlier approaches to sentiment analysis, such as the rule-based approach or some forms of classifiers that do not include neural networks, struggle to capture the sequentiality in language. These methods typically work word by word, disregarding the order or context in which the words are used, leading to lower accuracy in sentiment classification (Zhao et al., 2021) [2].

The objectives set for this research are as follows: Main objective: To design an efficient sentiment analysis model for the IMDb movie reviews. The aim is to identify complex approaches within sentiments, handle language intricacies, and most importantly, yield high percentages of successful sentiment classification. The specific objectives of the study are: To apply a machine learning model that captures sequential dependencies in text data, compare the performance of several models including traditional classifiers and Long Short-Term Memory (LSTM) neural network models, fine-

tune the LSTM model for accuracy and feasibility, and ascertain the portability of the trained model to related fields like product and customer feedback surveys.

To mitigate these two limitations, this research employs LSTM networks, a type of RNN used for successive data. LSTMs are more useful when there is a need to factor in the context and sequence of words, making it ideal for sentiment analysis. The study involves several key steps: data preprocessing, model architecture design, training and evaluation, optimization, and comparison with baseline models. Data preprocessing includes cleaning the IMDb dataset by removing null values, duplicates, punctuation, hyperlinks, and stop words, followed by tokenization and padding to ensure uniform input length.

The architecture of this model is robust, consisting of a sequential model that passes the input through an embedding layer to transform the words into vectors and two LSTM layers that help carry out sequential analysis. Dropouts have been included to reduce overfitting. The next step is training and evaluating where the model is trained on the IMDb dataset through the concept of accuracy to determine its correctness, and the validation is done to check for its generality. Optimization is about changing hyper-parameters, including the dimensions of the embedding vector, the number of words per sequence, and the number of LSTM nodes. The study also compares the LSTM model with the other ordinary classifiers, including Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Decision Tree, to demonstrate the efficiency of LSTMs. This work will seek to enhance the accuracy of sentiment classification to above 80% while presenting a robust model that can be adopted for sentiment analysis tasks other than movie reviews, extending beyond the use of LSTM networks.

Study Problem

The problem addressed by this study was the development of an improved sentiment analysis machine learning model for IMDb movie reviews that effectively captures the sequence of events and contextual cues, which conventional techniques often overlook by focusing solely on the word level without considering the critical interaction and order of words.

The following reasons explain the necessity of developing an appropriate sentiment analysis model for movies' IMDb reviews. IMDb is one of the most visited websites, and its users post about their beliefs and attitudes towards movies. Detailed sentiment analysis of such texts can provide additional insights about the audience, enhance recommendation systems, and uncover public recognition of filmmakers and producers. Furthermore, a dependable sentiment analysis model can accurately process the reviews, which would otherwise take enormous time to analyze manually. This model offers practical benefits that can significantly improve your work in this field.

It is also important to note that sentiment analysis can be applied not only to the sample of the IMDb movie reviews. By excluding punctuation marks and stop words and normalizing the words, such a model can also be easily transferred to other domains such as Amazon reviews and ratings, food delivery service reviews, or any other type of social media posts. This versatility of the model opens up a world of possibilities and can inspire new applications and research in the field of sentiment analysis.

Study Purpose

The purpose of this quantitative, experimental study was to create and assess a sentiment analysis model based on deep learning, using Long Short-Term Memory (LSTM) networks, to enhance the accuracy of classifying IMDb movie reviews.

This study entailed creating an LSTM model structure that converted words into compact vectors using an embedding layer, processed these vectors through numerous LSTM layers to capture sequential patterns, and used dropout layers to mitigate overfitting. The research was conducted using IMDb as the primary source, and the target

population consisted of a wide range of movie reviews available on the platform. The study utilized a sample size of 50,000 reviews, which were divided equally between positive and negative attitudes. The reviews were selected using a method called stratified random sampling. The data underwent preprocessing by eliminating null values, duplicates, punctuation marks, hyperlinks, and stop words. This was followed by tokenization and padding to ensure consistent input length. The model underwent training and evaluation using a range of performance metrics, including accuracy, precision, recall, and F1-score. Comparisons were performed to baseline models such as Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Decision Tree classifiers. The results highlighted the enhanced proficiency of LSTM models in capturing contextual and sequential information for sentiment analysis. Additionally, they shed light on the versatility of these models across several domains, extending beyond IMDb reviews to encompass product feedback and social media posts.

Guiding Research Questions

1. In what way can the model improve search results as well as movie recommendations?

    The model's precise assignment of labels to IMDb movie reviews will enhance the relevancy of search results and recommendations. It's not just about positive reviews, but also about understanding the user's preferences and mood. For instance, when a user searches for movies, the model can rank films based on positive reviews and sentiments that align with the user's preferences, creating a more engaging and satisfying user interface. Additionally, the model can recommend movies that match the review sentiments likely to interest the user, considering their current state of mind or mood.

2. Can the model be applied to other contexts, such as the comments section for a product on Amazon or restaurant reviews?

    The model's flexibility and effectiveness are not limited to a single domain. It can be applied to various contexts, from Amazon product reviews to restaurant feedback. Its ability to capture both sequence matters and contextual dependencies from the text makes it a versatile tool. For Amazon, it can help recommend products, restock inventory, or identify opportunities for product improvement. In the case of restaurant reviews, it can understand customer feedback, highlight frequently ordered foods, and identify areas needing improvement, thereby boosting customer satisfaction. This adaptability allows businesses to glean insights from customer reviews, enabling better decision-making and service enhancement.

Review of the Literature

The problem addressed by this study was the development of an improved sentiment analysis machine learning model for IMDb movie reviews that effectively captures the sequence of events and contextual cues, which conventional techniques often overlook by focusing solely on the word level without considering the critical interaction and order of words. The purpose of this quantitative, experimental study was to create and assess a sentiment analysis model based on deep learning, using Long Short-Term Memory (LSTM) networks, to enhance the accuracy of classifying IMDb movie reviews. This study addresses two guiding research questions: (1) In what way can the model improve search results as well as movie recommendations? And (2) Can the model be applied to other contexts, such as the comments section for a product on Amazon or restaurant reviews?

Deep Learning and Sentiment Analysis

LSTM-based sentiment analysis has been the subject of recent research, identifying some important issues and potential solutions. Studies have shown that language-specific preprocessing strategies are critical for improving model performance, especially for

morphologically tricky languages like Arabic (Oussous et al., 2019; Zahidi et al., 2023) [3,4]. Word embedding models, such as FastText and Word2Vec, have demonstrated the potential to increase classification accuracy (Zahidi et al., 2023) [4]. CNN and LSTM are two deep learning models that routinely outperform more conventional machine learning techniques (Bhuvaneshwari & Rao, 2022; Aljehane, 2023) [5,6]. To improve LSTM performance, researchers have investigated various optimization strategies, including adaptive particle swarm optimization (Shobana & Murali, 2021) [7]. Additional difficulties with multilingual sentiment analysis arise from the need for ethical considerations and language-specific adjustments (Madhavaram et.al., 2024) [8]. Large language models and pre-trained models are examples of recent advances, and future research will concentrate on overcoming difficulties in various linguistic situations and enhancing model flexibility (Rahman Jim et al., 2024) [9].

Applications of Sentiment Analysis

Sentiment analysis, a power tool for understanding public sentiment and consumer behavior, is of paramount importance in various contexts, including social media, e-commerce, and movie reviews. The numerous case studies that demonstrate the variety of uses and developments in sentiment analysis models serve as a validation of its significance, reassuring professionals in the field of its relevance and impact.

Reviews of Films

IMDb movie reviews are frequently used as a standard sentiment analysis dataset. Research by Sarlis and Maglogiannis (2020), Jnoub et al. (2020), and Xia (2024) [10-12] investigate how sentiment analysis can enhance the comprehension of movie reviews and audience ratings. These studies highlight the significance of domain-specific data for training models and show how advanced machine-learning methods may effectively classify attitudes. Filmmakers and marketers can gain significant insights from these researchers' improved understanding of moviegoers' feelings, which they have achieved by utilizing sophisticated algorithms [13].

Online Shopping

Sentiment analysis is useful for understanding customer views and product reputations in the e-commerce space. Studies by Patra et al. (2022), Xu et al. (2020), and Venkata et.al (2023) [14-16] explore domain adaptation techniques to improve model generalization on various platforms. This research uses sophisticated machine-learning methods, like deep learning and adaptive stochastic networks, to increase sentiment classification accuracy [17]. Businesses can better customize their plans and increase customer happiness by using the insights obtained from these analyses, essential for comprehending consumer preferences and market trends.

Social Media

Social media sites, a rich source of up-to-date information, are a practical application ground for sentiment research. Researchers Brownfield and Gagan et. al. (2019) , Eswar et.al. (2021) and Sunkara et. al. (2021) [18-20] use state-of-the-art machine-learning approaches to analyze emotions in postings shared on social media, shedding light on how sentiment analysis can be practically used to track public opinion, spot new trends, and manage company reputation [18-20]. Their work, which includes managing large datasets, ongoing education, and adjusting to different domains, significantly improves the resilience and flexibility of sentiment analysis algorithms, enlightening and informing the academic and professional community [21-23].

The preceding case studies highlight the importance of sentiment analysis in obtaining significant insights from diverse data sources. Businesses and researchers can better understand customer sentiment, product reputations, and market trends by

utilizing powerful and flexible sentiment analysis models. This, in turn, makes it easier to make more informed decisions and to plan strategically, which eventually produces better results across a range of industries.

Challenges and Future Directions in Sentiment Analysis

LSTM-based sentiment analysis has been the subject of recent research, identifying some important issues and potential solutions. Studies have shown that language-specific preprocessing strategies are critical for improving model performance, especially for morphologically tricky languages like Arabic (Oussous et al., 2019; Zahidi et al., 2023) [3,4]. Word embedding models, such as FastText and Word2Vec, have demonstrated the potential to increase classification accuracy (Zahidi et al., 2023) [4]. CNN and LSTM are two deep learning models that routinely outperform more conventional machine learning techniques (Bhuvaneshwari & Rao, 2022; Galla, 2023) [24,25]. To improve LSTM performance, researchers have investigated various optimization strategies, including adaptive particle swarm optimization [26]. Additional difficulties with multilingual sentiment analysis arise from the need for ethical considerations and language-specific adjustments [27]. Large language models and pre-trained models are examples of recent advances, and future research will concentrate on overcoming difficulties in various linguistic situations and enhancing model flexibility [28].

Gaps in the Literature

The literature on sentiment analysis highlights significant gaps in model design, performance comparison, and transferability, especially when it comes to handling the sequential and contextual nature of text input.

Model Creation

One of the main gaps in sentiment analysis research is the lack of models that can adequately represent sequential relationships and contextual subtleties in text data. Conventional techniques, like rule-based strategies and basic classifiers, frequently process text word by word, ignoring word context and order. The efficacy and accuracy of sentiment classification are restricted by this method [29].The preservation of sequential information and the ability to capture long-term dependencies in text are two intriguing solutions provided by recent advances in deep learning, especially concerning Long Short-Term Memory (LSTM) networks. LSTM networks are more effective than traditional classifiers in various NLP tasks. However, further research is still needed to fully understand how LSTM networks can be applied to sentiment analysis, particularly in datasets with lengthy and diverse texts like IMDb reviews.

Model Comparison

The absence of thorough performance comparisons between LSTM models and conventional classifiers represents another significant gap. There needs to be more systematic comparison in sentiment analysis, despite some research showing the improved performance of LSTM models in particular circumstances, such as morphological processing in Arabic. Conventional classifiers, such as Decision Trees, Multinomial Naïve Bayes, and Gaussian Naïve Bayes, have been extensively employed in sentiment analysis; however, they encounter difficulties in obtaining the sequential and contextual data required for sophisticated text classification [30]. Consequently, it is imperative to compare LSTM models to these conventional classifiers to determine their efficacy and pinpoint any possible drawbacks or areas in need of development.AI and ML plays significant role in prediction and forecasting based on the trends [31,32].

Transferability

Lastly, there needs to be more research in the literature about the transferability of sentiment analysis models across various domains. The majority of research so far has been on a specific domain, such as movie reviews and has not examined how well these models may be applied to other textual data types, like social media posts or product reviews on Amazon. This gap is important because a sentiment analysis model's usability and application can be greatly increased by being able to generalize across different domains. A model would be a flexible tool in sentiment analysis, increasing its practical applications and leading to more reliable and generalized NLP solutions if it could be easily modified and retrained for diverse datasets [33].

To close these gaps, the present research study presents a complete strategy combining sophisticated deep learning methods, thorough model comparison, and an emphasis on model adaptability is needed. Future research might greatly improve the accuracy and application of sentiment analysis models by developing an LSTM-based model, methodically testing its performance against standard classifiers, and guaranteeing its transferability to multiple domains. This project is useful for businesses that depend on textual data to understand customer sentiment and add to the scholarly conversation.

2. Method

The problem addressed in this study is developing an improved sentiment analysis method for IMDb movie reviews that effectively captures the sequence of events and contextual cues, which conventional techniques often overlook by focusing solely on the word level without considering the critical interaction and order of words. The purpose of this quantitative, experimental study was to create and assess a sentiment analysis model based on deep learning, using Long Short-Term Memory (LSTM) networks, to enhance the accuracy of classifying IMDb movie reviews. This study addresses two guiding research questions: (1) In what way can the model improve search results as well as movie recommendations? And (2) Can the model be applied to other contexts, such as the comments section for a product on e-commerce or restaurant reviews?

2.1. Data Description and Problem Setup

2.1.1. Description of the IMDb Sentiment Dataset

The IMDb Sentiment Dataset is a collection of movie reviews and ratings where each has an associated sentiment score. The IMDb sentiment dataset is now considered a standard in the NLP and sentiment analysis domain. It comprises of movie reviews as obtained from the official website, IMDb for movies. Every review in the dataset is said to be associated with a sentiment which could either be either positive or negative. These labelings help to enable supervised learning and model training.

2.1.2. Characteristics of the Dataset

The IMDb sentiment dataset is relatively large with records amounting to 50,000, which makes it suitable for developing and testing sentiment analysis models. Splitting the dataset into two equal parts: positive and negative ensures that the model does not lag or favour a particular sentiment more than the other. The reviews are highly diverse with some containing less than 100 words while others are about 2,000 words long. Specifically, the length of the reviews varies within the dataset, which makes it difficult in the preprocessing phase and while feeding data to the model as it needs to be normalized.

2.2.2. Data Cleaning and Preprocessing Steps

To prepare the IMDb sentiment dataset for analysis and model training, several data cleaning and preprocessing steps were undertaken:

1. Removal of Nulls: The dataset was checked for null values, and any records containing null values were removed to ensure data integrity. Fortunately, the IMDb dataset did not contain any null values.
2. Removal of Duplicates: A total of 418 duplicate records were identified and removed from the dataset to avoid redundancy and ensure that each review is unique.
3. Punctuation Removal: Punctuation marks, which do not contribute to the sentiment of the text, were removed. This step helps in simplifying the text and reducing the feature space.
4. Hyperlinks Removal: Any hyperlinks embedded within the reviews were removed as they are irrelevant to the sentiment analysis task.
5. Stop Words Removal: Common stop words (e.g., "the," "is," "and") were removed as they do not carry significant sentiment information. This step helps in reducing noise in the data.
6. Lemmatization: Words were lemmatized to their root forms (e.g., "running" to "run") to ensure that different forms of the same word are treated as a single feature. This step helps in reducing the dimensionality of the feature space and improving model performance.
7. Normalization of Case: All words were converted to lowercase to ensure uniformity and to prevent the model from treating words with different cases as separate entities.
8. Tokenization and Padding: The text was tokenized, converting each word into an integer value based on its frequency of occurrence in the corpus. The reviews were then padded or truncated to a fixed length to ensure consistent input dimensions for the model.

2.2. Baseline Models

2.2.1. Description of Baseline Models

Gaussian Naïve Bayes (GNB):

The Gaussian Naïve Bayes as the name suggests is a Naïve Bayes classifier that follows the Bayes' theorem and assumes features in the data to be Gaussian distributed (Joseph et al., 2022). Highly useful for high-dimensional data but has the drawback of implying that all features are independent and uncorrelated which is not ideal for text datasets.

Multinomial Naïve Bayes (MNB):

Multinomial Naïve Bayes is another kind of probabilistic classifier derived from Bayes' theorem but it is particularly suitable for discrete data like word frequencies. This model believes the features (words) are generated through the multinomial distribution and is popular for use in text categorization (Rani & Kumar, 2019).

Decision Tree (DT):

Decision Tree is a model among non-parametric models where the data is divided into subsets based on the input features and the splits forms a trees like structure. Nodes are features, branches are decision rules, and the end points or leaves signify results. Decision Trees are easily understandable by human beings but there is a tendency of them to overfit in the model (Yuan et al., 2020).

Performance Measures of Baseline Models with Tokenizers and TF-IDF Vectorization

The baseline models were first trained using a basic Tokenizer which mapped every word to an integer according to its frequency in the given categories. Subsequently, the efficacy of these models was assessed using Term Frequency-Inverse Document

Frequency (TF-IDF) vectorization where the words were weighted in relation to their occurrences in the document in comparison to the number of documents in the collection.

Gaussian Naïve Bayes:
- Tokenizer:
  - Accuracy: ~50%
- TF-IDF:
  - Accuracy: 78%

Multinomial Naïve Bayes:
- Tokenizer:
  - Accuracy: ~50%
- TF-IDF:
  - Accuracy: 86%

Decision Tree:
- Tokenizer:
  - Accuracy: ~50%
- TF-IDF:
  - Accuracy: 73%

Summary of Classification Reports for Each Baseline Model

Gaussian Naïve Bayes (TF-IDF):
- Accuracy: 78%
- Precision, Recall, F1-Score:
  - Negative: Precision 0.76, Recall 0.81, F1-Score 0.79
  - Positive: Precision 0.80, Recall 0.75, F1-Score 0.77

Support: 9,917 samples

Multinomial Naïve Bayes (TF-IDF):
- Accuracy: 86%
- Precision, Recall, F1-Score:
  - Negative: Precision 0.86, Recall 0.85, F1-Score 0.86
  - Positive: Precision 0.86, Recall 0.86, F1-Score 0.86

Support: 9,917 samples

Decision Tree (TF-IDF):
- Accuracy: 73%
- Precision, Recall, F1-Score:
  - Negative: Precision 0.72, Recall 0.73, F1-Score 0.73
  - Positive: Precision 0.73, Recall 0.72, F1-Score 0.72

Support: 10,000 samples

2.3. LSTM-Based Sequential Recurrent Neural Network Model

2.3.1. Architecture of the LSTM Model

The architecture of the Long Short-Term Memory (LSTM) model used for sentiment analysis consists of the following layers:

1. Embedding Layer: This layer transforms the input words into dense vectors of fixed size (50 dimensions). It helps in converting words into a numerical format that can be fed into the neural network.
2. First LSTM Layer: The first LSTM layer contains 64 units. This layer processes the embedded word vectors sequentially and captures the temporal dependencies and patterns in the text data.

3. Dropout Layer: A dropout rate of 20% is applied to the first LSTM layer to prevent overfitting by randomly setting a fraction of the input units to zero during training.
4. Second LSTM Layer: The second LSTM layer contains 32 units and further captures the sequential dependencies in the text data.
5. Dropout Layer: Another dropout rate of 20% is applied to the second LSTM layer to prevent overfitting.
6. Dense Output Layer: The final layer is a Dense layer with a Softmax activation function, which outputs the probability distribution over the sentiment classes (positive and negative).

Training Setup

1. Sequence Length: The input text is tokenized and padded to a fixed sequence length of 80. This means each input sentence is either truncated or padded to 80 tokens.
2. Embedding Dimensions: The words are embedded into dense vectors of 50 dimensions.
3. Number of Epochs: The model is trained for 5 epochs. Each epoch represents a complete pass through the entire training dataset.

Performance Metrics

1. Training Accuracy: After 5 epochs, the training accuracy of the LSTM model was approximately 87.3%.
2. Validation Accuracy: The validation accuracy of the model was around 83.7%, indicating its ability to generalize to unseen data.

2.3.2. Optimization and Tuning of the Model

1. Hyperparameters: Different hyperparameters of the model were adjusted within the range for obtaining the highest accuracy: the number of LSTM layers, dropout rate, and the learning rate.
    *   The first designed LSTM layer comprised 64 units, an embedding dimension of 50, and dropout of 20%.
    *   To make the model better, a new LSTM layer with 32 nodes was incorporated into the model.
    *   The length of the sentences was varied from 50 to 100 tokens, where the authors between 80 tokens in the text.
2. Sentence Length: This brought about the most important parameter of sentence length. First, testing was conducted with longer contexts, up to 150 tokens, but they greatly affected the training speed. Concerning the sequence length, different values were tested and analysing the outcomes, the authors selected 80 tokens, as this length offered good results and did not require many hours to be trained.
3. Learning Rate and Epochs: In the second stage, a range of values was used to properly determine the learning rate in order to have the best convergence rate. Also, the epoch was set to 5 considering the way the model performed on the validation set to avoid overfitting and underfitting.

3. Results and Discussion

3.1. Comparison of Performance Between Baseline Models and the LSTM Model

The performance of the Long Short-Term Memory (LSTM) model was compared with three baseline models: Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Decision Tree. These baseline models were evaluated using both a basic tokenizer and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization.

Baseline Models Performance (TF-IDF):

- Gaussian Naïve Bayes:
  - Accuracy: 78%
  - Precision (Negative/Positive): 0.76/0.80
  - Recall (Negative/Positive): 0.81/0.75
  - F1-Score (Negative/Positive): 0.79/0.77
- Multinomial Naïve Bayes:
  - Accuracy: 86%
  - Precision (Negative/Positive): 0.86/0.86
  - Recall (Negative/Positive): 0.85/0.86
  - F1-Score (Negative/Positive): 0.86/0.86
- Decision Tree:
  - Accuracy: 73%
  - Precision (Negative/Positive): 0.72/0.73
  - Recall (Negative/Positive): 0.73/0.72
  - F1-Score (Negative/Positive): 0.73/0.72

LSTM Model Performance:

- Training Accuracy: 87.3%
- Validation Accuracy: 83.7%

The LSTM model outperformed the baseline models in terms of training accuracy and validation accuracy. While the Multinomial Naïve Bayes model showed strong performance among the baseline models with an accuracy of 86%, the LSTM model demonstrated superior ability in capturing the sequential dependencies in text data, leading to more accurate sentiment classification.

3.2. Discussion on the Improvements Observed with the LSTM Model

The improvements seen with the LSTM model are largely due to its capability to capture the context and relationships between words, which is essential for understanding sentiment. Unlike traditional models that treat each word in isolation, the LSTM model utilizes its memory cells to retain information over sequences, allowing it to grasp the order and connections between words.

Key Improvements:

1. Contextual Understanding: The sequential processing capabilities of the LSTM model enable it to comprehend the context in which words appear, leading to more accurate sentiment predictions (Ahmet & Abdullah, 2020).
2. Handling Long Texts: The model's ability to handle variable-length sequences and capture long-term dependencies is particularly advantageous for analyzing movie reviews, which often have lengthy and complex sentences.
3. Reduction of Overfitting: The incorporation of dropout layers in the LSTM model helps prevent overfitting, ensuring the model generalizes well to new, unseen data.
4. Example Predictions on Custom Test Data

To demonstrate the effectiveness of the LSTM model, we tested it on custom sentences that were not part of the training data:

1. "Worst movie I have seen or will ever watch"
   Prediction: Negative
2. "Great comedy, go see it with a friend"
   Prediction: Positive

3. "Do not watch that movie, it is horrible"
    Prediction: Negative
4. "It's the best movie made in its genre"
    Prediction: Positive

The LSTM model accurately classified these custom test sentences, showcasing its ability to predict sentiments based on context and word order. These examples emphasize the model's practical applicability in real-world situations where understanding nuanced sentiments is crucial.
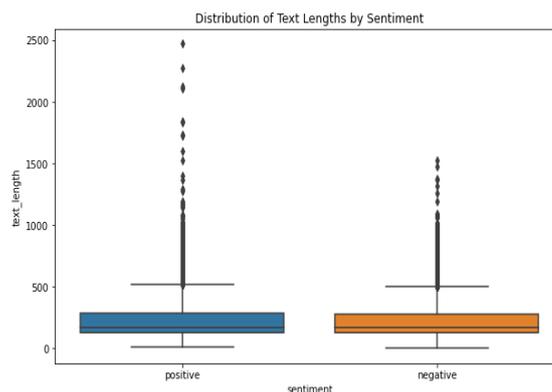
Graph 1: Distribution of Sentiment Text Length Before and After Removing Stop Words
- Left Boxplot (Blue): Distribution of sentiment text lengths before removing stop words.
- Right Boxplot (Orange): Distribution of sentiment text lengths after removing stop words.

This visualization shows the reduction in the average text length and the variability in the dataset after the removal of stop words. The left boxplot represents the original text lengths, while the right boxplot represents the text lengths after stop words have been removed.
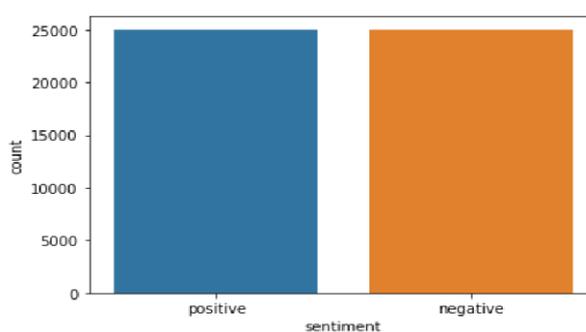
Graph 2: Comparison of Sentiment Classes Distribution
- Left Bar (Blue): Number of negative sentiment reviews.
- Right Bar (Orange): Number of positive sentiment reviews.

This bar chart illustrates the distribution of sentiment classes in the dataset, showing an equal number of negative and positive reviews.

Distribution of text length before and after removing stop words
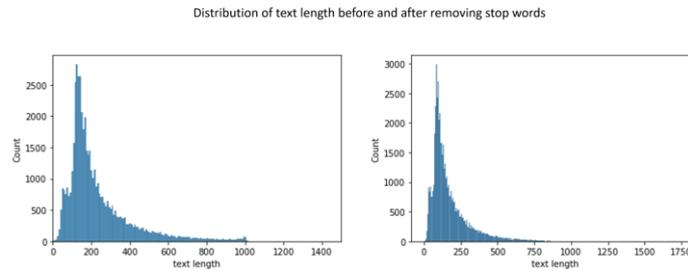
Graph 3: Distribution of Text Length Before and After Removing Stop Words
- Left Histogram: Distribution of text lengths before removing stop words.
- Right Histogram: Distribution of text lengths after removing stop words.

This visualization shows the distribution of text lengths in the dataset. The left histogram represents the text lengths before stop words were removed, and the right histogram shows the text lengths after stop words were removed. The histograms illustrate how removing stop words reduces the average text length and overall word count.

3.3. Limitations of the Model

Despite the advancements in sentiment analysis with the Long Short-Term Memory (LSTM) model, several key limitations need to be considered:
- Training Time: The LSTM model requires extensive training time and significant computational resources due to its complexity and the need to process sequential data. This makes it challenging to implement in resource-limited environments.
- Scalability: The model faces scalability issues, as larger datasets increase the training and inference time and resources required. Its performance may decline when applied to datasets with different characteristics from the training data, necessitating retraining.
- Language Variations: The LSTM model struggles with dialects, slang, and regional expressions, reducing its effectiveness with multilingual datasets or varied linguistic nuances.
- Contextual Challenges:
  - Sarcasm: Difficult to detect due to reliance on context and tone not explicitly present in text.
  - Proper Nouns: Can introduce ambiguity as the model may not distinguish between different entities or understand their associated sentiment without additional context.
  - Other Nuances: Idiomatic expressions, humor, and cultural references can impact sentiment but may not be adequately captured by the model.

4. Conclusion

This study developed and evaluated a Long Short-Term Memory (LSTM) model for sentiment analysis of IMDb movie reviews, demonstrating significant improvements over traditional models like Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Decision Tree. The LSTM model's ability to capture context and sequential dependencies in text resulted in more accurate sentiment classification. Key contributions include effective data preprocessing steps such as tokenization, padding, and lemmatization, which enhanced the model's performance.

The LSTM model has several practical applications, including improving recommendation systems on platforms like IMDb by providing personalized movie suggestions, analyzing product reviews on e-commerce sites to help businesses

understand customer sentiments, and monitoring social media for brand management and marketing strategies.

Future research should focus on optimizing the model to reduce training time and improve scalability for real-time applications. Additionally, efforts should be made to adapt the model for handling multiple languages and dialects, and to better capture complex language features like sarcasm, humor, and proper nouns. Testing and refining the model in diverse real-world scenarios will ensure its robustness and adaptability. Overall, this study highlights the potential of LSTM models to significantly enhance sentiment analysis and their impact on various fields that require a nuanced understanding of textual data.

Ethics

The authors declare that there are no ethical issues associated with the publication of this manuscript. All data used in this research were publicly available and properly cited. The study adhered to ethical guidelines and standards in handling and analyzing the data, ensuring integrity and transparency throughout the research process.

References

[1] Agarwal, B., Nayak, R., Mittal, N., & Patnaik, S. (Eds.). (2020). Deep Learning-Based Approaches for Sentiment Analysis. Springer. https://doi.org/10.1007/978-981-15-1216-2

[2] Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Information Processing & Management, 58(5), 102656. https://doi.org/10.1016/j.ipm.2021.102656

[3] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2019). Big Data technologies: A survey. Journal of King Saud University-Computer and Information Sciences, 30(4), 431-448. doi.org/10.1016/j.jksuci.2017.06.001

[4] Zahidi, Y., Al-Amrani, Y., & El Younoussi, Y. (2023). Improving Arabic Sentiment Analysis using LSTM based on Word Embedding models. *Vietnam Journal of Computer Science*, 10(03), 391-407.

[5] Bhuvaneshwari, P., Rao, A. N., Robinson, Y. H., & Thippeswamy, M. N. (2022). Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model. *Multimedia Tools and Applications*, *81*(9), 12405-12419.

[6] Aljehane, N. O. (2023, September). A New Approach to Sentiment Analysis on Twitter Data with LSTM. In *2023 3rd International Conference on Computing and Information Technology (ICCIT)* (pp. 657-663). IEEE.

[7] Shobana, J., & Murali, M. (2021). An efficient sentiment analysis methodology based on long short-term memory networks. *Complex & Intelligent Systems*, *7*(5), 2485-2501.Kumaresan, C., & Thangaraju, P. (2023). ELSA: Ensemble learning based sentiment analysis for diversified text. *Measurement: Sensors*, *25*, 100663.

[8] Madhavaram, C. R., Sunkara, J. R., Kuraku, C., Galla, E. P., & Gollangi, H. K. (2024). The future of automotive manufacturing: Integrating AI, ML, and Generative AI for next-Gen Automatic Cars. *IMRJR*, *1*(1). https://doi.org/10.17148/imrjr.2024.010103

[9] Jim, M. M. I., Rauf, M. A., Rahman, M. M., & Tariquzzaman, M. (2024). AI-Powered Predictive Analytics For Intellectual Property Risk Management In Supply Chain Operations: A Big Data Approach. *Available at SSRN 4986872*.

[10] Sarlis S, Maglogiannis I. On the Reusability of Sentiment Analysis Datasets in Applications with Dissimilar Contexts. Artificial Intelligence Applications and Innovations. 2020 May 6;583:409–18. doi: 10.1007/978-3-030-49161-1_34. PMCID: PMC7256387.

[11] Jnoub, N., Al Machot, F., & Klas, W. (2020). A domain-independent classification model for sentiment analysis using neural models. *Applied Sciences*, *10*(18), 6221.

[12] Xia, Z., Chen, J., & Sun, A. (2024). Mining the relationship between COVID-19 sentiment and market performance. *PloS one*, *19*(7), e0306520.

[13] Sun, C., Tian, Y., Gao, L., Niu, Y., Zhang, T., Li, H., ... & Yu, J. (2019). Machine learning allows calibration models to predict trace element concentration in soils with generalized LIBS spectra. *Scientific reports*, *9*(1), 11363.

[14] Patra, G. K., Rajaram, S. K., Boddapati, V. N., Kuraku, C., & Gollangi, H. K. (2022). Advancing Digital Payment Systems: Combining AI, Big Data, and Biometric Authentication for Enhanced Security. *International Journal of Engineering and Computer Science*, *11*(08), 25618–25631. https://doi.org/10.18535/ijecs/v11i08.4698

[15] Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *Ieee Access*, *7*, 51522-51532.Mejova, Y., & Srinivasan, P. (2021). Crossing Media Streams with Sentiment: Domain Adaptation in Blogs, Reviews and Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *6*(1), 234-241. https://doi.org/10.1609/icwsm.v6i1.14242

[16] Venkata Nagesh Boddapati, Eswar Prasad Galla, Gagan Kumar Patra, Chandrakanth Rao Madhavaram, & Janardhana Rao Sunkara. (2023). AI-Powered Insights: Leveraging Machine Learning And Big Data For Advanced Genomic Research In Healthcare. *Educational Administration: Theory and Practice*, *29*(4), 2849–2857. https://doi.org/10.53555/kuey.v29i4.7531

[17] Shravan Kumar Rajaram, Eswar Prasad Galla, Gagan Kumar Patra, Chandrakanth Rao Madhavaram, & Janardhana Rao. (2022). AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. *Educational Administration: Theory and Practice*, *28*(4), 285–296. https://doi.org/10.53555/kuey.v28i4.7529

[18] Gagan Kumar Patra, Shravan Kumar Rajaram, & Venkata Nagesh Boddapati. (2019). Ai And Big Data In Digital Payments: A Comprehensive Model For Secure Biometric Authentication. *Educational Administration: Theory and Practice*, *25*(4), 773–781. https://doi.org/10.53555/kuey.v25i4.7591

[19] Eswar Prasad Galla.et.al. (2021). Big Data And AI Innovations In Biometric Authentication For Secure Digital Transactions Educational Administration: Theory and Practice, 27(4), 1228 –1236Doi: 10.53555/kuey.v27i4.7592

[20] Janardhana Rao Sunkara, Sanjay Ramdas Bauskar, Chandrakanth Rao Madhavaram, Eswar Prasad Galla, Hemanth Kumar Gollangi, Data-Driven Management: The Impact of Visualization Tools on Business Performance, International Journal of Management (IJM), 12(3), 2021, pp. 1290-1298. https://iaeme.com/Home/issue/IJM?Volume=12&Issue=3

[21] Hemanth Kumar G. et. al.(2024). Data Engineering Solutions: The impact of AI and ML on ERP systems and supply chain management. (2024). *Nanotechnology Perceptions*, *20*(S9). https://doi.org/10.62441/nano-ntp.v20is9.47

[22] Bauskar, S. R., Madhavaram, C. R., Galla, E. P., Sunkara, J. R., & Gollangi, H. K. (2022). Predicting disease outbreaks using AI and Big Data: A new frontier in healthcare analytics. *European Chemical Bulletin*. https://doi.org/10.53555/ecb.v11:i12.17745

[23] Siddharth K, Gagan Kumar P, Chandrababu K, Janardhana Rao S, Sanjay Ramdas B, et al. (2023) A Comparative Analysis of Network Intrusion Detection Using Different Machine Learning Techniques. J Contemp Edu Theo Artific Intel: JCETAI-102.

[24] Galla, Eswar Prasad and Gollangi, Hemanth Kumar and Boddapati, Venkata Nagesh and Sarisa, Manikanth and Polimetla, Kiran and Rajaram, Shravan Kumar and Reddy, Mohit Surender, Enhancing Performance of Financial Fraud Detection Through Machine Learning Model (October 23, 2023).

[25] Eswar Prasad G, Hemanth Kumar G, Venkata Nagesh B, Manikanth S, Kiran P, et al. (2023) Enhancing Performance of Financial Fraud Detection Through Machine Learning Model. J Contemp Edu Theo Artific Intel: JCETAI-101, Available at SSRN: https://ssrn.com/abstract=4980350 or http://dx.doi.org/10.2139/ssrn.4980350

[26] Manikanth Sarisa, Venkata Nagesh Boddapati, Gagan Kumar Patra, Chandrababu Kuraku, & Siddharth Konkimalla. (2022). Deep Learning Approaches To Image Classification: Exploring The Future Of Visual Data Analysis. *Educational Administration: Theory and Practice*, *28*(4), 331–345. https://doi.org/10.53555/kuey.v28i4.7863

[27] Sarisa, M., Boddapati, V. N., Patra, G. K., Kuraku, C., Konkimalla, S., & Rajaram, S. K. (2020). An Effective Predicting E-Commerce Sales & Management System Based on Machine Learning Methods. *Journal of Artificial Intelligence and Big Data*, 1(1), 75–85. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1110

[28] Gollangi, H. K., Bauskar, S. R., Madhavaram, C. R., Galla, E. P., Sunkara, J. R., & Reddy, M. S. (2020). Exploring AI Algorithms for Cancer Classification and Prediction Using Electronic Health Records. *Journal of Artificial Intelligence and Big Data*, 1(1), 65–74. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1109

[29] Rajaram, S. K., Konkimalla, S., Sarisa, M., Gollangi, H. K., Madhavaram, C. R., Reddy, M. S., (2023). AI/ML-Powered Phishing Detection: Building an Impenetrable Email Security System. ISAR Journal of Science and Technology, 1(2), 10-19.

[30] Mohit Surender Reddy, Manikanth Sarisa, Siddharth Konkimalla, Sanjay Ramdas Bauskar, Hemanth Kumar Gollangi, Eswar Prasad Galla, Shravan Kumar Rajaram, 2021. "*Predicting tomorrow's Ailments: How AI/ML Is Transforming Disease Forecasting*", *ESP Journal of Engineering & Technology Advancements,* 1(2): 188-200.

[31] Galla, Eswar Prasad and Gollangi, Hemanth Kumar and Boddapati, Venkata Nagesh and Sarisa, Manikanth and Polimetla, Kiran and Rajaram, Shravan Kumar and Reddy, Mohit Surender, Prediction of Financial Stock Market Based on Machine Learning Technique (October 23, 2023). Eswar Prasad G, Hemanth Kumar G, Venkata Nagesh B, Manikanth S, Kiran P, et al. (2023) Prediction of Financial Stock Market Based on Machine Learning Technique. J Contemp Edu Theo Artific Intel: JCETAI-102, Available at SSRN: https://ssrn.com/abstract=4975786

[32] Galla, Eswar Prasad and Gollangi, Hemanth Kumar and Boddapati, Venkata Nagesh and Sarisa, Manikanth and Polimetla, Kiran and Rajaram, Shravan Kumar and Reddy, Mohit Surender, Enhancing Performance of Financial Fraud Detection Through Machine Learning Model (October 23, 2023). Eswar Prasad G, Hemanth Kumar G, Venkata Nagesh B, Manikanth S, Kiran P, et al. (2023) Enhancing Performance of Financial Fraud Detection Through Machine Learning Model. J Contemp Edu Theo Artific Intel: JCETAI-101, Available at SSRN: https://ssrn.com/abstract=4980350 or http://dx.doi.org/10.2139/ssrn.4980350

[33] Bauskar, Sanjay and Boddapati, Venkata Nagesh and Sarisa, Manikanth and Reddy, Mohit Surender and Sunkara, Janardhana Rao and Rajaram, Shravan Kumar and Polimetla, Kiran, Data Migration in the Cloud Database: A Review of Vendor Solutions and Challenges (July 22, 2022). Available at SSRN: https://ssrn.com/abstract=4988789 or http://dx.doi.org/10.2139/ssrn.4988789