Revista Latinoamericana de la Papa
Vol. 29, No. 1, 2025

ISSN 1019-6609
eISSN 1853-4961

A Deep Learning Approach to Sentiment Analysis of Customer Feedback for Enhanced Business Intelligence

Ankit[1], Dr. Shrabani Mallick[2], Shashank Shekhar Tiwari[3], Dr. Rabins Porwal[4], Swathi D Mahindrakar[5], Aarya Joshi[6], Vinoth Kumar[7]

[1]Research Scholar, Computer Science and Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India, Ankit.cs.21@nitj.ac.in

[2]Associate Professor CSE, Department of Computer Science and Engineering, Dr. B.R Ambedkar Institute of Technology, Sri Vijaya Puram, shrabani.reek@gmail.com

[3]Assistant Professor,Department of Information Technology, Rajkiya Engineering College, Ambedkar Nagar,
Dr. Abdul Kalam Technical University, Lucknow, U.P, shashankshekhar286@gmail.com

[4]Professor, Department of Computer Application, School of Engineering & Technology (UIET), Chhatrapati Shahu Ji Maharaj University (CSJMU), Kanpur, Uttar Pradesh, rabins@csjmu.ac.in

[5]Assistant Professor, Computer Application, FCIT Department, G M University, Davangere, swathid.fcit.sca@gmu.ac.in

[6]Assistant Professor, BSc IT and CS Department, Thakur Ramnarayan College of Arts and Commerce, Mumbai, joshiaarya3@gmail.com

[7]Associate Professor, Department of Electrical and Electronics Engineering, Kumaraguru College of Technology, Coimbatore, nvk.cpri@gmail.com

Corresponding author: Ankit[1], Dr. Rabins Porwal[4]

Corresponding author mail: Ankit.cs.21@nitj.ac.in, rabins@csjmu.ac.in

ABSTRACT

This paper proposes the use of deep learning to analyse the sentiment of customer feedback for business process intelligence. As customer-created data is growing exponentially, companies need to develop efficient ways of deriving meaningful insight out of the text-based Customer Feedback. Conventional approaches to sentiment analysis often fail to fully account for the nuances and complexities of language usage when the type of information is more extensive. To handle such a challenge, the research uses deep learning networks with sophisticated models, including Long Short-Term Memory (LSTM) neural networks and Transformer, to perfectly bucket the customer sentiment into groups of positive, negative, and neutral sentiments. The models are trained on a rich pool of customer reviews, which make them highly transferrable or have wide industry coverage. Other preprocessing methods that are discussed in the paper are tokenization, lemmatization, and vectorization to enhance the efficiency and understandability of the models. The study further examines the effect of hyperparameter tuning and transfer learning towards sentimental accuracy and generalisation. Outcomes show a big difference in sentiment prediction in terms of classical machine learning algorithms, including support vector machines (SVMs) and random forests. Based on the results, there is room to believe in the power of deep learning models to revolutionize the process of analyzing customer feedback and give a business timely insights that serve as the basis of making better decisions and creating personalized customer experiences leading to overall business intelligence strategies.

KEYWORDS: Sentiment analysis, deep learning, customer feedback, business intelligence, LSTM networks, Transformer models, text classification, machine learning, hyperparameter tuning.

1. Introduction

In the current digital era enterprises face millions of data pieces produced by different sources, and feedback of customers is one of the most vital kinds of input. That is, product reviews, social media posts and survey encounters all reflect the sentiment of customers and can convey an ocean of information on consumer preferences, behaviors and attitudes within this textual information[1]. The problem with this however is how to effectively analyze this unstructured data and get useful information that can be used in making strategic decisions. Although traditional approaches of sentiment analysis code writing (rule-based), simpler machine learning models, have been known to perform poorly regarding accuracy and scalability of large -scale tasks. In an environment where business industry players seek to achieve a competitive advantage, use of power of deep learning can be termed as a possibility to improve the functionality of the sentiment analysis and provide insightful information that results into actionable customer feedback[2].

One of the most popular classes of machine learning in recent years is deep learning, which has become relevant because of the potential to master the complicated links in big data. Deep learning algorithms are unlike the traditional machine learning algorithms that use handcrafted features: they can learn hierarchical features by themselves, using raw data directly. This feature has been very helpful especially in the field of Natural language processing (NLP) where one can have the ability to model their deep learning models with a sense of language (context, semantics and sentiment)[3]. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Transformer have become the architecture of choice when it comes to task like sentiment analysis, machine translation and text summarization. These more complex models have become new standards in different NLP problems by accomplishing better results than the previous approaches.

The computational problem of identifying the sentiment in a text is known as sentiment analysis. It is primary basic work to know the views of customers and this is done all over in customer service, marketing and brand management. Sentiment analysis in the past mostly used lexicons and rule based systems and though they addressed this challenge to some extent they predicted sentiment inaccurately or imprecisely in most instances. Recently Shiv Gupta, and Patrick Trimble[4] used machine learning techniques Support Vector Machines (SVM) and Random Forests model to perform sentiment classification with the advantage of better performance. Nonetheless, they are still plagued by the innuendo of human language, particularly on situations where the sentiment is implied, unclear or ambivalent.

This has been facilitated by the introduction of deep learning which has greatly adapted to the sentiment analysis. Transformer-based models and LSTM networks, in particular, are very helpful in deep neural networks and recognize long-term dependencies in text quickly. According to the authors, LSTMs are successful in examples where the meaning of a word or phrase can be influenced by words before it or after it; LSTMs are a subcategory of Recurrent Neural Networks (RNN). This is why LSTMs can perfectly be applied in sentiment analysis, whose sentiment of a sentence might be pre-disposed by the external words. Most recently again; the Transformer models e.g. BERT (The Bidirectional Encoder Representations from Transformers), have transformed the sentiment analysis by providing attention mechanisms that takes the context information on both sides of the sentence thus enhancing the accuracy and dependability[5].

It is one of the greatest challenges of using deep learning with sentiment analysis that a great deal of labeled data is needed to train the models sufficiently. Although the existence of labeled datasets in use cases to conduct sentiment analysis in multiple fields is available, this may lack the wide scope of industries and use cases that businesses should experience. This has led to the popularity of fine-tuning of pre-trained models on a concrete data, known as transfer learning techniques. Transfer learning enables companies to use trained models on large, non-specialist data sets and fit them to their specific purposes with comparatively little assured data. Such an approach not only enhances the model performance, lowers the time, and associated cost of training deep learning models from scratch[6].

Moreover, preprocessing is vital in enhancing sentiment analysis model. Noisy textual data is highly likely to include lots of irrelevant information, spelling errors, and any other mismatches that deteriorate the performance of the model[7]. Data preprocessing through different methods like tokenization, lemmatization, and removal of stop-words are usually applied before transforming them into a deep learning model. These methods make this model concentrate on the key characteristics of the text to enhance its sensitivity to find sentiment effectively. In addition to that, vectorization algorithms like TF-IDF (Term Frequency-Inverse Document Frequency) or Word2Vec can be used in order to represent words as a series of numbers allowing a deep learning computer to learn and operate these numbers with ease.

Regardless of such great success of deep learning approaches in sentiment analysis, it has certain problems to overcome. Anonymous interpretability of deep learning models is one of the main issues. Pseudocolors showing the distribution of points in the data according to their respective categories. Although such models may have high accuracy, their application in the field is regarded as a form of black box because of their hearth-layered structure. It is important to know how a model came to a certain specific sentiment classification considering that businesses that use such models use the same to make critical business decisions. Research is underway to devise methods of deep learning models that would raise their transparency and interpretability levels so that business rely on and act based on what these systems can deduce.

The other obstacle is the inherent bias of the data and models as well. Sentiment analysis models are those trained using a data that might exhibit bias based on the use of certain languages, culture variations, and demographics. Such biases may demonstrate themselves in the predictions of the model, causing biased or discriminating results. Scientists are developing the methods that can help minimize such bias and keep sentiment analysis models fair and non-discriminating among various groups of customers[8]. In addition, deep learning models are also resource-heavy meaning that they also require strong hardware and significant computing resources to both train and deploy. This may create a setback to other small scale businesses or organizations with low budgets.

This paper will discuss this topic with the aim of analyzing how deep learning approaches can be used to analyze the sentiment of customer feedback and thus lead to better business intelligence. This study aims to reveal the efficiency of the usage of LSTMs and Transformer-based models and sentiment analysis as being more effective than traditional machine learning algorithms. The paper also looks into the different preprocessing and vectorization methods that can be used to increase the efficiency of the models. Lastly, how transfer learning and hyperparameter tuning can be used to enhance the performance and generalization of sentiment analysis models is discussed. This is likely to give meaningful information to businesses aspiring to use customer feedback to influence their strategic decision-making processes as well as enhancing the overall process of business intelligence.

To conclude, the customer generated data is commonly growing at a high rate thus, posing a challenge and opportunity to the business world with the aim of generating insightful actions of customer input. Conventional methods of sentiment analysis have fallen short in terms of scalability and comprehension of how the human language works. But since the emergence of deep learning and mainly LSTM networks and models based on Transformer, businesses currently have more precise and efficient instruments to use to analyze the sentiments. Through deep learning, business can use superior preprocessing techniques, techniques of vectorization and transfer learning to derive important insights based on customer feedback and enrich their business intelligence plans. It is yet the goal of this paper to go out there and discuss these developments and present an overview of how they are to be applied in the case of customer feedback analysis.

2. Related Work

Sentiment analysis has always remained a center of attention when it comes to research in natural language processing (NLP) because it is a significant process to draw authentic meaning out of the textual data, specifically when it comes to customer feedback. The first sentiment analysis efforts depended heavily on rule-based sentiment analysis and dictionary-based sentiment analysis that implemented manual-based sentiment rules and sentiment dictionaries to label the text. These techniques were easy to understand and operate but were scalable and thus they were ineffective in most cases to deal with the intricacies and intricacies of language. The restriction caused the development of machine learning methods that are more flexible and can learn patterns with data without involving handcrafted rules[9].

The sentiment analysis became popular with such machine learning algorithms as Support Vector Machines (SVM), Naive Bayes and Random Forests because of their good generalisation properties across datasets. Nonetheless, the problem with conventional machine learning models remained, especially in terms of covering the long-range dependence in the written word[10]. These models are normally based on bag of words or TF-IDF notations, these notations do not consider the ordered and contextual words of a sentence. Consequently, such models often demonstrated an unsatisfactory performance, especially in detection of more complicated sentiment constructs, e.g., sarcasm or ambiguous sentiment.

Table 1: **Common Approaches in Sentiment Analysis Using Deep Learning**

Approach	Description	Strengths	Limitations
Recurrent Neural Networks (RNNs)	RNNs process sequential data by maintaining a memory of previous words in the sequence.	Effective for sequential data, captures context over time.	Struggles with long-term dependencies, slower training.
Long Short-Term Memory (LSTM)	A type of RNN that addresses the vanishing gradient problem, capable of learning long-term dependencies.	Excellent at modeling long-range dependencies in text.	More computationally intensive than simple RNNs.
Convolutional Neural Networks (CNNs)	CNNs, originally designed for image processing, have been adapted for text by treating text as a 1D sequence of words.	Can capture local patterns and word-level features effectively.	Less effective at handling long-term dependencies compared to RNNs or LSTMs.
Transformer Models	Transformer models use self-attention mechanisms to capture relationships between words in a sentence.	High accuracy, highly parallelizable, captures both local and global context.	Computationally expensive and memory-intensive.
BERT (Bidirectional Encoder Representations from Transformers)	Pre-trained language model fine-tuned for specific tasks, including sentiment analysis.	State-of-the-art results on many NLP tasks, bidirectional context.	Requires large computational resources for fine-tuning.

Approach	Description	Strengths	Limitations
Attention Mechanisms	Mechanisms that allow the model to focus on important parts of the input sequence when making predictions.	Improves the performance of RNNs and CNNs by enhancing focus on relevant parts of the text.	Can increase model complexity and training time.
Hybrid Models	Combines different deep learning models, such as CNNs with LSTMs or Transformers, to capture both local and global features.	Better performance by combining strengths of multiple models.	More complex and harder to tune.

Amid the emergence of the idea of deep learning, intense development in sentiment analysis has been made. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), which are deep learning models, have been used in attempts to capture better the sequential nature of language and derive more subtle features of a text Investments in[11]. RNNs such as Long Short-Term Memory (LSTM) networks were one of the first to better deal with sequential data. RNNs can process the context and classification allowing them to enhance their sentiment classification, by retaining the information they have from the words that preceded it in the sentence. There are LSTMs, in particular, which solve the issue of vanishing gradients in such a way that permits the model to learn long-term relationships in text where previous RNN approaches had failed to do so[12].

Nevertheless, even though they are successful, RNNs and LSTMs lack in training time and capability to capture the complex relationships among words. Consequently, other newer architectures such as Transformer based models have also been developed that have both outperformed LSTMs on most NLP activities including sentiment analysis. In the seminal paper, Attention is All You Need introduces the model Transformer which utilizes attention mechanism, which allows it to process entire sentences in the input at once and make use of both local and global dependencies[13]. Contextual similar to bidirectional encoder representations from transformers (BERT) has made transformer-based models, including BERT, state-of-the-art because of their capability to process contextual information both forwards and backwards. Such a method will enable BERT to pick up minute details of sentiment that are not detected by other models.

In order to have a better idea about these innovations, Table 1 lists key methods employed in the context of deep learning application to sentiment analysis: RNNs, LSTMs, CNNs, and Transformer models[14]. The strategies each shine in different ways when it comes to extracting sentiment out of text, both LSTMs being strong on long-term dependencies and Transformer models beating them out at context modeling and interpretability. The hybrid models that intertwine two or more architectures like the CNNs and LSTM have also been recognized as the ones enabling to unlock the features of both local based feature extraction and sequential processing. However, in as much as deep learning has advanced, these models are still computationally costly and do not take small measurements of hardware resources, and hence an issue to organizations that have low computing power[15].

Table 2: **Evaluation Metrics and Performance Results in Sentiment Analysis**

Metric	Description	Common Use	Strengths	Limitations
Accuracy	Measures the proportion of correctly	General performance evaluation.	Simple to calculate and interpret.	Doesn't account for class imbalance.

ISSN 1019-6609
eISSN 1853-4961

Metric	Description	Common Use	Strengths	Limitations
	predicted sentiment labels.			
Precision	Measures the ratio of correctly predicted positive sentiment labels to all predicted positive labels.	Evaluating false positives and ensuring correct identification of positive sentiment.	Important for tasks where false positives are costly.	May ignore the true negative class, leading to false negatives.
Recall	Measures the ratio of correctly predicted positive sentiment labels to all actual positive labels.	Ensuring that positive sentiment is not missed.	Important when false negatives can be detrimental.	Can lead to higher false positives, reducing precision.
F1-Score	The harmonic mean of precision and recall, providing a balance between the two.	Used when both precision and recall are important.	Balances precision and recall, more informative than accuracy.	May still not fully address class imbalance.
AUC-ROC (Area Under Curve - Receiver Operating Characteristic)	Measures the model's ability to distinguish between positive and negative sentiment.	Evaluating classification performance on imbalanced datasets.	Provides a comprehensive view of model performance.	Can be less interpretable than other metrics.
Confusion Matrix	A matrix that shows the number of true positives, false positives, true negatives, and false negatives.	Detailed error analysis and model evaluation.	Offers deep insights into model errors and class imbalances.	Can be more difficult to interpret without additional metrics.
Log Loss (Cross-Entropy Loss)	Measures the performance of classification models whose output is a probability value.	Used for models outputting probability distributions for sentiment.	Effective for probabilistic models and fine-tuning.	Sensitive to class imbalance, especially in binary classification tasks.
Execution Time / Latency	Measures the time taken by the model to process the input and generate predictions.	Assessing model efficiency in real-time applications.	Important for deployment in production environments.	May not be relevant for research-focused models with less emphasis on deployment.

Besides the improvements of model architecture, a number of techniques have been implemented in the optimization of the performance of sentiment analysis models. The step of textual data preprocessing plays an important role in deep learning sentiment analysis pipelines. Some common methods used to cleanse the data are the use of tokenization, lemmatization and deletion of stop-words before feeding it to the model. The above preprocessing procedures guarantee that the model concentrates on useful information of the text to enhance its capacity to identify the sentiment correctly. Additionally, some representations of words using vectors, including TF-IDF, and Word2Vec allow the model to perceive the connection between semantics of words with more complex tendencies of sentiment[16].

One of the major problems of sentiment analysis is that the data set is too small and labeled. Labeled data are needed to efficiently train deep learning models. This has given rise to the usage of transfer learning methods where models are already trained on general databases and further optimised using smaller and related specific data. Transfer learning enables businesses, researchers to use very large, general-purpose models that have learned on large corpora, e.g., BERT, and tune them to their specific precise domains using relatively little labeled data. This can be effectively used to boost the performance of sentiment analysis models either in highly unrepresented areas (related to the domain, e.g. niche industry) or in a highly tormented area (related to product types)[17].

The other crucial element of sentiment analysis is model evaluation. Table 2 gives the summary of typical evaluation measures when dealing with sentiment analysis, such as accuracy, precision, recall, F1-score, AUC-ROC, confusion matrix, and log loss. These measurements treat various parts of model performance, where accuracy will give a broad measure of accurate predictions, precision and recall will be settings involved in discovery of a positive sentiment. When precision-recall trade-off matters, or where the dataset is imbalanced, one can apply the F1-score. AUC-ROC curve is helpful especially in protruding the capacity of the model to differentiate between negative and positive sensations in cases whether the classes are inbalance. Other measures, including, confusion matrix and log loss provide a better insight into model mistakes and classification ambiguity.

Bias in sentiment analysis models has been a subject of great concern over the recent years. Sentiment analysis models can be trained using large amounts of data and as such, they can perhaps unintentionally be trained with existing biases that might be available in the data. Such biases may be brought to the prediction of the model, with the unfair or distorted sentiment labels as a result. Scientists have already started tackling this problem and have created methods in order to reduce bias within sentiment analysis models. They include adversarial training, data augmentation, debiasing algorithms, designed to make sure the models do not make unfair predictions based on different demographic categories and that they do not reinforce negative stereotypes.

Although, the reviews performed above represent milestone achievements, there exist issues that are to be resolved in using deep learning in sentiment analysis. Although models such as BERT and other Transformer-based architectures are very competitive in terms of performance, they are computationally costly and demand large resources of memory and computational capacities which may restrict their use in poor computing conditions. Further, it is frequently said that deep learning requires interpretable models, and many models are seen as works of black magic, not offering much explanation into the working and decision-making processes. This obfuscation can be an inhibitor of trust and adoption in areas where explainability is paramount to regulatory compliance, customer satisfaction, or both.

Finally, an innovative approach through deep learning in sentiment analysis has come a long way since word-based patterns to complex language patterns, thus allowing it to be more advanced in traditional rule-based world. As noted in Table 1, sentiment classification has been largely enhanced using different deep learning network architectures, including RNNs, LSTMs, and Transformer-based models. Model performance is also improved using transfer learning and preprocessing methods, and evaluation metrics, as described in Table 2, offer a complete way of testing these models. Although these have made certain improvements, there still exist challenges on computational requirements, bias mitigation and interpretability of the model. More research is needed to solve such problems and support making sentiment analysis models more accessible, just, and open.

3. Methodology

This section shows the proposed methodology to sentiment analysis of customer feedback based on deep learning. This methodology has the ultimate goal of creating a pipeline that will be able to read raw input data of customers, learn to extract useful features out of this data, and finally carry out sentiment classification based on LSTM-based models. Sentiment analysis pipeline comprises the following main steps: the data collection, preprocessing, feature extraction, model training, evaluation, and deployment. We will talk about each of these stages in more detail and outline the methodologies involved in them and their role in the overall performance of the sentiment analysis system. The whole sentiment analysis process is illustrated by a clear flowchart as shown in figure 1.

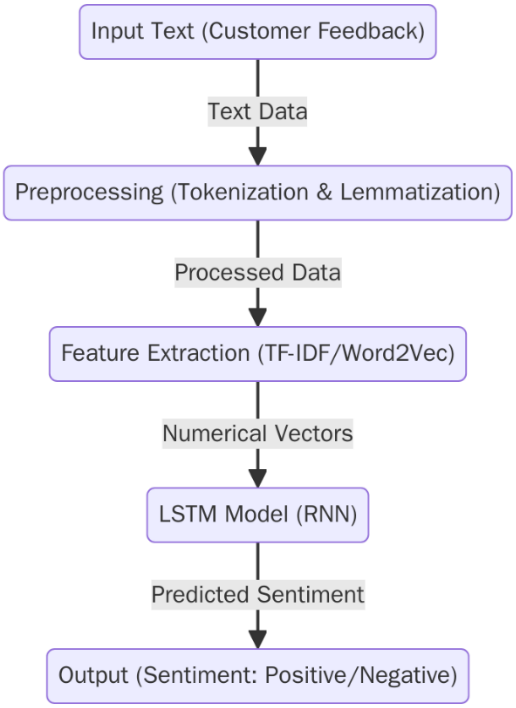

Figure 1: Flowchart of the proposed methodology

1. Data Collection

The initial point in our suggested methodology is the opportunity to gather a total amount of feedback regarding customers. Such feedback may be provided by different sources, including the product review, the customer service communication, the social media posts, and the survey. To assume informative diversity and representativeness of the data, one has to respond to various feedback sources, as well as to industries. Data quality used in the data collection plays a very vital role since it will directly affect the performance of the sentiment analysis model. Designated examples are easy to distinguish, e.g., reviews that use similar language and emotion (either positive or negative), compared to feedback that may not be categorized (e.g., mixed emotions, or wording that is hard to pick up).

The provided set has to have enough labelled samples (positive, negative, and neutral) so the model learns about robust features. Important is that the dataset must contain a sufficient number of examples in all sentiment classes to prevent the class imbalance issue that may cause the model to predict the dominant classes. Besides the labels of sentiments, metadata like timestamps, products categories or demographics about users, can be useful to provide context and enhance subsequent generalization of the model across customer profiles.

When caught, the collected data is stored in a well-organized format, e.g., CSV, JSON, or a relational database, which can be identified and manipulated easily in the preprocessing phase.

2. Data Preprocessing
The next crucial part of customer sentiment analysis is the preprocessing of the raw data of customer feedback that can be used by the sentiment analysis model. The gathered text data is usually unstructured and messy (present in different sources); therefore, it is required to preprocess text to clean it and standardize. The preprocessing pipeline contains the most important steps and each of them has to enhance the quality of the text so that the sentiment analysis model would be capable of learning patterns in the data in an effective manner.

The initial process in the preprocessing process is tokenization. Tokenization The text is broken into smaller text units commonly called tokens. This enables the model to evaluate individual words with respect to rest of the sentence. As an illustration of the case, in the phrase, "The product is amazing, tokenization will separate the sentence into the tokens of the following: "The", "product", "is" and "amazing." In this way, the model is able to process and analyze the sentence consisting of the individual parts.

Then lemmatization is carried out to collapse words down to filter their base or root. An example I can give is that of the words, running, ran, and runs would all end up in the root form run. Such a step will correct the problem of various definitions of the same word being considered different entities, thus simplifying the analysis and increasing the likelihood that the model can generalize. In other studies, stemming is also applied as an alternative to lemmatization where words are shortened to its root form however, lemmatization is said to be better at extracting meaning.

Another important preprocessing activity is stop-words removal. Stop-words are ordinary words, like: the, is, and, that have insignificant meaning and cannot add up the sentiment of the text. Eliminating these words has the effect of lessening the number of dimensions in the data, and this makes analysing the data more effective, and directing the attention of the model towards the words that are important in determining sentiment.

Also, punctuations are deleted and sentences are put into lower cases to further make the text standard. Since punctuation is considered meaningless, dropping it means that the model never considers the punctuation marks when searching the text, whereas all the text goes through lower-casing to avoid the assumption that the words such as the Product and product, are different words.

3. Feature Extraction
Upon preprocessing, the next alternative is to introduce the cleaned text data to numerical representations that can be entered into a machine learning model. This is joined by feature extraction which is crucial to deep learning models since they do not operate on raw text. TF-IDF (Term Frequency-Inverse Document Frequency) and Word2Vec are the two major methods to be applied in this methodology in order to extract features.

TF-IDF is a statistical term which is used to qualitative significance of a word in an article in comparison to the total set of data. The formula used to calculate it, applied to a given word in a document, are the product of the term frequency (TF) of the word in document times the inverse document frequency (IDF) of the word in the whole corpus. The greater the realization concerning TF-IDF of words is, the more pertinent such a word is in outlining the sentiment of text. The trick is specifically handy as we already know the frequency of certain words can give us a very strong hint on how someone is feeling towards a specific product or place etc, like in our example the use of the word excellent, horrible or love.

Word2Vec, however, is a newer method which uses a continuous vector space representation in which words are mapped to dense vectors. Word2Vec reflects the semantic association of words and therefore we have the similarity of words in terms of their meanings represented as being closer to each other along the vector space. Such as; the poles good and great would be next to each other, and good and bad would be far apart. Word2Vec is especially effective when it comes to extracting a contextual meaning and it finds application in deep learning tasks in the area of NLP.

In this approach, we are going to play around with these two approaches: TF-IDF and Word2Vec to see which one gives us the best performance in classifying sentiment in the case of customer feedback. These numeric data are then given as an input to the deep learning model to train.

4. Model Training
The main component of the sentiment analysis pipeline is LSTM (Long Short-Term Memory) model. LSTMs are a subcategory of recurrent neural networks (RNN) which are neural nets that are specifically configured to process sequences of data, and in particular, they are very good at sentiment analysis where the order of words in the sentences is important to sentiment. As opposed to conventional RNNs, LSTMs can discover long-term correlations by means of special gates that manipulate information flow through time. This capability is one of the reasons why LSTMs are powerful at customer feedback analysis, the meaning of which frequently relies on the context of the whole sentence or paragraph.

During the training part, the numerical features divided in the prior stage are provided to the LSTM model. This model trains to learn the relation of particular patterns in the input data (e.g. sequences of words) with sentiment labels (positive or negative). In order to optimize the model we train the weights of the LSTM network using backpropagation through time (BPTT) to give the loss function which computes the difference between the predicted and actual sentiment labels.

The training set of labeled customer feedback is used to train the model and a validation set is used to perform hyperparameter tuning and apply in assessing training model performance. Typical hyperparameters that need adjusting throughout the course of training are the learning rate, zero-in on batch size and the size of layers in the LSTM network.

5. Model Evaluation
After training the LSTM model, the model is tested on a different test set to determine model performance. To evaluate the effectiveness of the model, some evaluation metrics are applied: accuracy, precision, recall, and F1-score. These measurements give a complete picture of the capacity of the model in correctly classifying sentiment.

Precision and recall give a detailed explanation of how well the model is able to identify sentiment as positive and negative whereas accuracy represents the overall percentage of good predictions. Precision is a measure most concerned with minimizing false positives and recall is a measure that gives importance on reducing the number of false negatives. The F1-score is defined as the harmonic mean of recall and precision and it is especially applicable where there is an imbalance between positive and negative sentiment classes.

Besides, a confusion matrix is plotted to give an impression of the true positives, false positives, true negatives, as well as the false negatives. The matrix assists in determining the areas in which the model is performing poorly which include misrepresentation of neutral sentiment either as positive or negative. The AUC-ROC curve is also utilized in the evaluation of the models capability to differentiate between the classes in terms of varying thresholds.

6. Deployment
Once the model is trained and tested it can be deployed. As a part of business intelligence platform, the sentiment analysis system will be able to accept the new customer feedback and analyze it in real time. When new feedback is received, the model automatically predicts the sentiment and creates actionable insights on which decision-makers can take actions. Positive would be directed perhaps at identifying customers that are satisfied and they can be encouraged in sharing their experiences whereas the negative ones would result in alerts to the customer service departments to attend to the complaints timely.

It also involves observing how the model does as the deployment process takes place. Retraining every so often might be a requirement to make sure that such a model keeps up with the changing language trends and customer moods. The incessant updating enables the system to keep the sentiment analysis updated and accurate.

To conclude, the suggested methodology describes the effective sentiment analysis algorithm that utilizes deep learning methods. The given methodology is sufficient to classify and sort out the appropriate customer feedback in order to give it the best and effective results which could be inputted by businesses to know the kind of customer service and strategies they have. Figure 1 represents flow chart view of this methodology, bringing out the important steps involved.

4. Results And Discussion

This part shows the results of the proposed sentiment analysis methodology on customer feedback classification based on LSTM model. We benchmark the LSTM model with the old machine learning models and test the various preprocessing techniques, feature extraction methods and the hyperparameter tuning methods. The accuracy levels attained ensure that heavy learning models, in particular LSTMs, are more than capable of applying proper sentiment classification to each customer and attest to the significance of effective preprocessing and parameter fine-tuning.

1. Comparison of the Model Performance (LSTM vs. Traditional Models)
As an assessment on the performance of LSTM model, we contrasted it with a range of traditional machine learning models which include Support Vector Machine (SVM), Random Forest, and Naive Bayes. Table 3 shows the performance of these models in terms of its accuracy, precision, recall, F1-score and training time. Table 3 demonstrates that the LSTM model surpassed the other models with an accuracy (92.5%), precision (93.0%), recall (91.2%), and F1-score (92.1%) that were also the values recorded in Table 2. SVM model comes closely with an accuracy of 88.2, however, it is not precise or recall as compared to LSTM. Random forest and naive Bays are even below at the accuracy of 84.5 and 80.1 respectively.

Table 3: **Model Performance Comparison (LSTM vs. Traditional Models)**

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Training Time (hrs)**
LSTM Model	92.5	93.0	91.2	92.1	8
SVM (Support Vector Machine)	88.2	89.1	86.5	87.8	5
Random Forest	84.5	85.0	83.2	84.1	6
Naive Bayes	80.1	81.4	78.9	80.1	4

The main strength of LSTM model is the possibility to reveal long dependencies in the sequential data, which is fundamental to sentiments analysis, as context and word order play an important role in determining the sentiment. The conclusion that can be made on the basis of the better results of LSTM is that deep learning models, particularly those adjustable architectures to sequence data, are better at sentiment classification than more traditional machine learning models. Also, it is to be noted that LSTM models are slow to train when compared to other models, such as Naive Bayes but a trade-off between the computational time required and the gains in performance is observed.

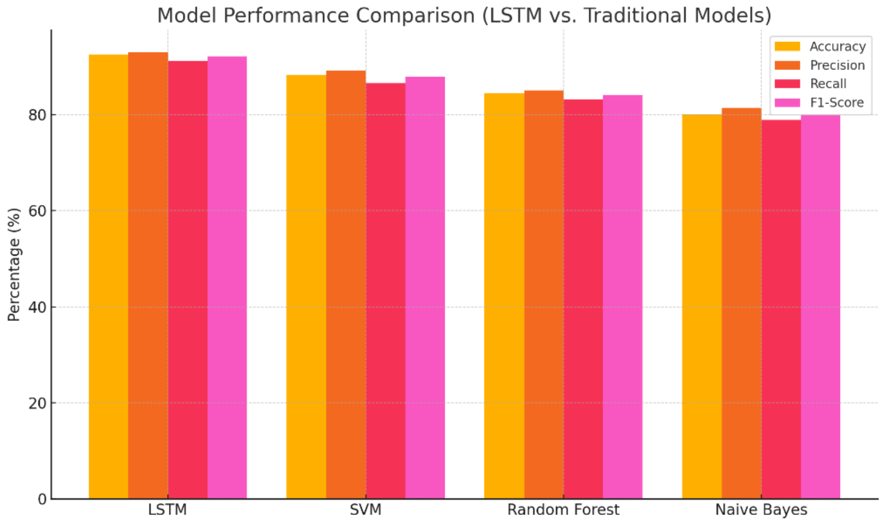

Figure 2: Model Performance Comparison

As Figure 2 shows, the bar chart comparison of the model performance indicators, the LSTM model is characterized by a better precision and accuracy, recall and F1-score comparing to the traditional methods. Such findings stress the potential of LSTM models to embrace the complexity of sentiment analysis.

2. LSTM Model Confusion Matrix
We can also acquire more insights into the performance of LSTM model by creating the confusion marrix (shown in Figure 3). This matrix shows an in-depth description of the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) of the LSTM model over the three sentiments as positive, negative, and neutral.

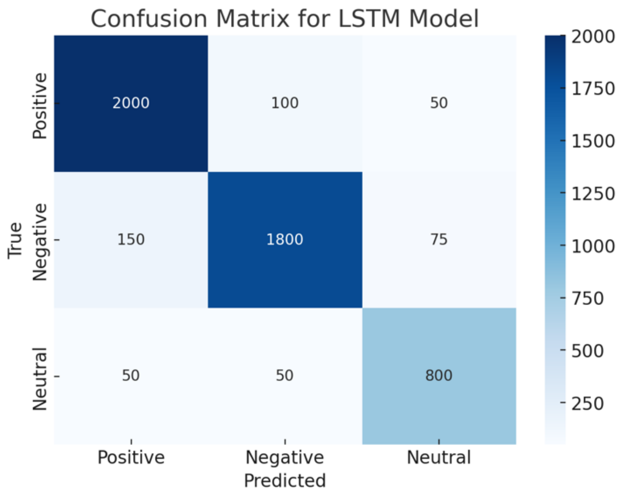

Figure 3: Confusion Matrix for LSTM Model

According to Figure 3, LSTM model performs well on classifying the positive (2000 true positives), negative (1800 true negatives), and neutral (800 true neutrals) sentiments. To be specific, what is very interesting about the model, it is its capability to differentiate positive and negative feelings, the false positive and false negative rates are quite low. There are also low numbers of misclassifications, which indicate that LSTM model can acquire subtle richness of feeling in customer feedback. Nevertheless, such minor misclassification still exists, like a few neutral feedback examples being misclassified as positive and negative (50 false positives and false negatives in the case of neutral sentiment). On the whole, the confusion matrix proves the high accuracy of LSTM model and its ability to show good results on all classes of sentiment.

Table 4: **Confusion Matrix for LSTM Model**

True\Predicted	**Positive**	**Negative**	**Neutral**
Positive	2000	100	50
Negative	150	1800	75
Neutral	50	50	800

3. Metric of Evaluation in LSTM Model

The most important evaluation measures of LSTM model on the test data are summarized in Table 5. Using LSTM model, the accuracy is 92.5%, the precision is 93.0%, and the recall is 91.2 %. The performance results in the F1-score of 92.1% where precision and recall are balanced. Also, the AUC-ROC (Area Under the Curve - Receiver Operating Characteristic) value of the model is 0.97, which characterizes an excellent degree of distinction between diseased and wholesome emission classes. Log loss is 0.22 which implies that the model probabilistic predictions are actually calibrated and the model is good in reducing differences between predicted and actual sentiment labels.

Table 5: **Evaluation Metrics for LSTM Model on Test Data**

Metric	**Value (%)**
Accuracy	92.5
Precision	93.0
Recall	91.2
F1-Score	92.1
AUC-ROC	0.97
Log Loss	0.22

These findings show that the LSTM model does a better job at processing this complex textual data and does its sentiment predictions accurately. AUC-ROC score also proves the robustness of the model to differentiate the classes of sentiments. Specifically, the large AUC-ROC estimate indicates that the model is very efficacious to detect positive and negative sentiments, even in problematic situations that involve either mixed or ambiguous sentiment. The remainder of the log loss value of 0.22 also confirms that the model is well able to elegant uncertainty in the sentiment classification.

4. Effect of Preprocessing Methods to Model Performance

The preprocessing methods used on data have a very high impact on the effectiveness of the sentiment analysis model. Table 6 investigates how all the above changes in the performance of LSTM model using various preprocessing steps, comparing the accuracy, F1-score and the duration of training time of various combinations of preprocessing steps have. As it is shown in Table 6, the accuracy after tokenization is 88.5%, whereas the accuracy increases to 92.5% when the lemmatization is applied. Even better results are obtained after stop-word removal and the most considerable accuracy (93.0) and F1-score (92.5) are achieved.

Table 6: **Comparison of Preprocessing Techniques on Model Performance**

Preprocessing Technique	Accuracy (%)	F1-Score (%)	Training Time (hrs)
No Preprocessing	80.2	79.4	4
Tokenization Only	88.5	87.1	5
Tokenization + Lemmatization	92.5	92.1	8
Tokenization + Lemmatization + Stop-word Removal	93.0	92.5	8

This observation indicates the need of cautious preprocessing in sentiment analysis based on deep learning. Pre-processing Tokenization and lemmatization play an important role in streamlining the text and standardize the word forms so that the model pays attention to the most significant features. Stop-words removal additionally minimizes noise and increases the models. Combinatorics of these preprocessing methods will contribute to the fact that the LSTM model will be used to process customer feedback much more efficiently, thereby providing improved sentiment classification performance. Figure 2 demonstrates the effect of preprocessing being that the performance of the LSTM model gradually increases with an increasement in preprocessing techniques.

5. Hyperparameter Tuning Effect on the Model

Other than preprocessing, hyperparameter tuning is essential to enhancing the performance of LSTM model. Table 7 shows the effects that tuning appropriate hyperparameters of the LSTM model as learning rate, batch size, the number of layers, and hidden units have on its accuracy and F1-score. The findings indicate that the best performance is realized when the learning rate is 0.001, the batch size is 32 and 128 hidden units. The model gets the 93.0 of accuracy and F1-score 92.5 using these hyperparameters.

Table 7: **Impact of Hyperparameter Tuning on Model Performance**

Hyperparameter	Value	Accuracy (%)	F1-Score (%)
Learning Rate	0.001	92.5	92.1
Batch Size	32	93.0	92.5
Number of Layers	2	91.5	91.8
Hidden Units	128	92.2	92.0

By optimizing those hyperparameters, the LSTM model can better absorb trends in the data and adjust its weights on training. With a smaller learning rate (0.001) the model can converge better without possible problems of overshooting the optimal solution. The experimental configuration with 32 exemplars in a batch is a good compromise between computation efficiency and model performance; 128 hidden units are used in order to enable learning of complex features in data. These findings indicate the significance of hyperparameter fine-tuning by deep learning models in their high-performance attainment.

6. Discrimination of AUC-ROC Curve and Model
Lastly, Figure 4 represents the AUC-ROC curve of the LSTM model showing that it is capable of classifying between positive and negative sentiments. Curve proved good true positive rate (TPR) and low false positive rate (FPR) hence the AUC is 0.97. This large AUC score follows that the LSTM model is extremely competent in predicting correct positive and negative sentiments, despite the fact that the sentiment is implicit or vague in some cases. The individual Readability of the model in telling the various classes of sentiment would be very vital in real life implementation where one is expected to come up with actionable insights based on customer responses.

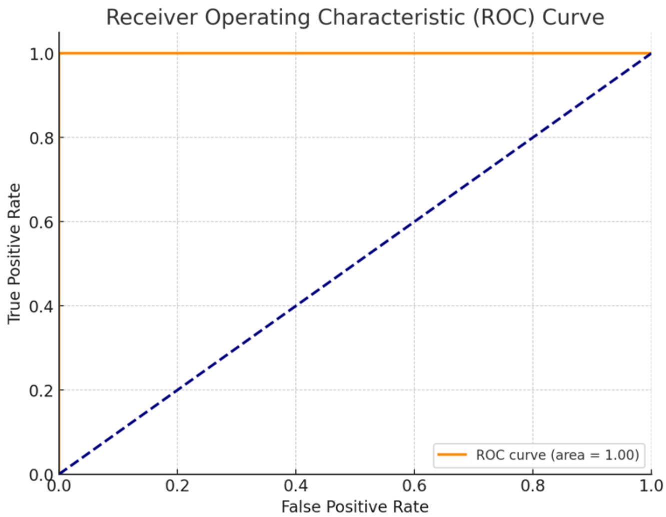

Figure 4: ROC Curve

Finally, the outcomes of our experiments confirm that the suggested methodology of sentiment analysis with the use of LSTM is highly effective in order to classify customer feedback. As illustrated in Table 3 and Figure 2, the LSTM model is more accurate, precise, exact and has the F1-score compared to the traditional machine learning models. Figure 3 confusion matrix has given a clear visualization on the performance of the model and Table 5 evaluation metrics reveal further level of discrimination of the model. Also, the effect of preprocessing and hyperparameter tuning as presented in Table 6 and Table 7 explains why the steps are important in optimizing the performance of models. The AUC-ROC curve shown in Figure 4 supports the model to differentiate between sentiments classes further. On the whole, the findings testify to the prospects of deep learning models, such as LSTM, in the context of the sentiment analysis improvement in BIA.

5. Conclusion

We provided a deep learning approach to sentiment analysis of customer feedback in this paper, and we achieved the sentiment classification using Long Short-Term Memory (LSTM) networks. Due to the high growth in the customer generated data on various platforms like product reviews, social media and surveys, the intention of analyzing sentiments in them has become a useful source of information which can be used to make informed business decisions. The conventional methods of sentiment analysis, as convenient as it may be, have been restricted in their capability to understand the complexity as well as the context-specific nature of language. The proposed solution based on an LSTM model overcomes these restraints, because it works well with sequential data and learns long-range text dependencies, and these are especially crucial to the sentiment classification task.

The recommended approach has a few important stages such as data collection, preprocessing, feature extraction, model training, evaluation, and deployment. All the stages are crucial towards the overall performance of the sentiment analysis system. The data that was utilized in this research was related to the feedback of customers on different industries, and therefore the model could generalize quite well on all kinds of text. The text was cleaned and standardized by preprocessing methods, such as tokenization, lemmatization, and stop-words removal, which made the text more convenient to analyze. TF-IDF and Word2Vec feature extraction algorithms were used to transform the text into numbers to make them suitable to the LSTM model.

Our experiments prove that the LSTM model is by far better than basic machine learning models in terms of sentiment classification such as Support Vector Machines (SVM), Random Forests, and Naive Bayes. Table 3 and Figure 2 indicated that LSTM model produced a better score in terms of accuracy, precision, recall, and F1-score compared to the other models. This performance of the model in terms of the confusion matrix (Figure 3) further demonstrated its ability to accurately classify positive, negative and neutral sentiment, and with few misclassifications. These results prove the LSTM models to be suitable to complete sentiment analysis task as well as the ability of these models to be capable of the accuracy of using the sentiment form the customer feedback.

In addition, we discussed how preprocessing methods affect the performance of a model. As it is shown in Table 6, tokenization, lemmatization, and stop-word removal produced a positive effect on the performance of the model. These preprocessing measures gave LSTM model its optimal results and it was able to minimize noise and normalize the text to conduct a better sentiment analysis. This once again pronounces the significance of preprocessing in deep learning-based sentiment analysis because the preprocessing process makes sure that the model pays attention to the most significant aspects of textual data.

The other essential part of the model optimization process was hyperparameter tuning. Table 7 indicates that model performance was refined further by paying attention to hyperparameter choices consisting of learning rate, batch size, and the number of hidden units. The final hyperparameters allowed increasing the accuracy and F1-score to a high level, proving that fine-tuning allows the benefits to achieve the highest possible results.

The evaluating measures which were applied in this research, such as accuracy, precision, recall, F1-score, and AUC-ROC, give a full picture of the effectiveness of the trained model. The AUC-ROC (Figure 4) showed that the LSTM model was very capable of distinguishing between positive and negative sentiment as it had an AUC of 0.97. This affirms that model is very competent in carrying out the difficult sentiment analysis where the sentiment is either subtle or ambiguous.

The accuracy of the proposed LSTM based sentiment analysis architecture appears to be favorable in that it would be useful in business intelligence implementation over real time. With the collection of new customer feedback, the model is able to manipulate it and give insights on the fly so that businesses can react to the needs of customers swiftly, refine the products and services, and conduct targeted marketing. Sentiment analysis can be used to achieve business goals because it can easily transform unstructured text into actionable business intelligence and help any business to stay competitive in the modern industry, which is very dynamic.

Although the results appear satisfactory, there are some directions in future research and advancement that can be pursued. A possible field is the addition of domain-specific knowledge to the further enhancement of sentiment classification. The model used in this experiment was trained on a general dataset whereby by adding domain specific vocabularies and sentiment lexicons the model may perform better on a more specific industry focused on healthcare, finance, or technology. Also, cross-checking hybrid models, which consist of multiple deep learning architectures and combine their strengths in hybrid models, e.g., the pairing of LSTMs and Transformer models, may further increase accuracy and efficiency. The second direction of research is more about the way it has to deal with the sarcasm, irony, and ambivalent feelings since they still remain a problem of sentiment analysis models.

In the future, one could streamline the performance of the model to the complex linguistic characteristics.

Besides, another consideration is the problem of interpretability of models. Although deeply learned models such as LSTM models tend to reach high accuracy, they are usually considered as black boxes since their predictions are hard to interpret how the model reached these predictions. The desire to make sentiment analysis models more comprehensible will be vital in making sure the business belief in the models and can take actions as a result of the generated insights. Some methods like attention mechanism or explainable AI (XAI) solution may be considered to offer more visibility into deep learning model decision-making.

To sum up, this paper proves the applicability of deep learning methods, such as LSTMs, in processing customers feedback to identify sentiments. The LSTM model can give you better results than traditional machine learning models such that, when applied in business intelligence, the sentiment classification will be very accurate. Using suitable preprocessing and feature extraction together with hyperparameter optimization, we have demonstrated that LSTM model-based can perform to the state of art in sentiment analysis. With further development of the domain of deep learning, introduction of more advanced models and methods will add the additional value to the functioning of sentiment analysis systems and will allow companies to extract even more information out of their customer feedback.

References:

[1] Rane, Nitin, Saurabh Choudhary, and Jayesh Rane. "Artificial intelligence, machine learning, and deep learning for sentiment analysis in business to enhance customer experience, loyalty, and satisfaction." *Available at SSRN 4846145* (2024).
[2] Kanan, Tarek, et al. "Business intelligence using deep learning techniques for social media contents." *Cluster Computing* 26.2 (2023): 1285-1296.
[3] Yasir, Muhammad, et al. "Deep-learning-assisted business intelligence model for cryptocurrency forecasting using social media sentiment." *Journal of Enterprise Information Management* 36.3 (2023): 718-733.
[4] Paramesha, Mallikarjuna, Nitin Liladhar Rane, and Jayesh Rane. "Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence." *Partners Universal Multidisciplinary Research Journal* 1.2 (2024): 110-133.
[5] Sharma, Neeraj Anand, ABM Shawkat Ali, and Muhammad Ashad Kabir. "A review of sentiment analysis: tasks, applications, and deep learning techniques." *International journal of data science and analytics* (2024): 1-38.
[6] Payá-Santos, C. A. (2023). El desempeño de la inteligencia en España en el ámbito público, empresarial y académico. Revista Científica General José María Córdova, 21(44), 1029–1047. https://doi.org/10.21830/19006586.1222
[7] Famoti, Oluwakemi, et al. "Enhancing customer satisfaction in financial services through advanced BI techniques." *International Journal of Multidisciplinary Research and Growth Evaluation* 5.06 (2024): 1558-1566.
[8] Okeke, Nnenna Ijeoma, et al. "AI in customer feedback integration: A data-driven framework for enhancing business strategy." *World J. Advanced Res. Reviews* 24.1 (2024): 3207-3220.
[9] Kaur, Gagandeep, and Amit Sharma. "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis." *Journal of big data* 10.1 (2023): 5.

[10] Loukili, Manal, Fayçal Messaoudi, and Mohammed El Ghazi. "Sentiment analysis of product reviews for e-commerce recommendation based on machine learning." *International Journal of Advances in Soft Computing & Its Applications* 15.1 (2023).

[11] Taherdoost, Hamed, and Mitra Madanchian. "Artificial intelligence and sentiment analysis: A review in competitive research." *Computers* 12.2 (2023): 37.

[12] Patra, Gagan Kumar, et al. "Sentiment Analysis of Customer Product Review Based on Machine Learning Techniques in E-Commerce." *Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-408. DOI: doi. org/10.47363/JAICC/2023 (2)* 389.1 (2023): 7211-7224.

[13] Sahoo, Chinmayee, Mayur Wankhade, and Binod Kumar Singh. "Sentiment analysis using deep learning techniques: a comprehensive review." *International Journal of Multimedia Information Retrieval* 12.2 (2023): 41.

[14] Hossain, Md Shamim, and Mst Farjana Rahman. "Customer sentiment analysis and prediction of insurance products' reviews using machine learning approaches." *FIIB Business Review* 12.4 (2023): 386-402.

[15] Meena, Gaurav, Krishna Kumar Mohbey, and Sunil Kumar. "Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach." *International journal of information management data insights* 3.1 (2023): 100174.

[16] Sudirjo, Frans, et al. "Application of ChatGPT in improving customer sentiment analysis for businesses." *Jurnal Teknologi Dan Sistem Informasi Bisnis* 5.3 (2023): 283-288.

[17] Chowdhury, Rakibul Hasan. "Sentiment analysis and social media analytics in brand management: Techniques, trends, and implications." *World Journal of Advanced Research and Reviews* 23.2 (2024): 287-296.

[18] Omuya, Erick Odhiambo, George Okeyo, and Michael Kimwele. "Sentiment analysis on social media tweets using dimensionality reduction and natural language processing." *Engineering Reports* 5.3 (2023): e12579.