Journal of Education for Pure Science- University of Thi-Qar
Vol.14, No.4 (2024)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: https://doi.org/10.32792/jeps.v14i4.452

# Social Media Comment Classification with Long Short-Term Memory Networks

**Noor Abdul Wahid Thwiny [1*], and Kadhim Hasenn Alibraheemi [2]**

[1, 2] Department of Computer Science, College of Education for Pure Sciences, Thi-Qar University, IRAQ.

[*] Corresponding email: noorabdalwahed.23co2@utq.edu.iq

**Received 8/6/2024, Accepted 7/8/2024, Published 1/12/2024**



## Abstract

People can now communicate with each other and share ideas thanks to social media by posting content on many communication platforms every day. Text comments have become the most widely used in posts across various social media platforms, through which users express their opinions. Manually classifying these comments is a time-consuming process, so this study aims to use artificial intelligence techniques to solve this problem effectively. In this study, we rely on natural language processing (NLP) techniques to classify comments into three categories: positive, negative, and neutral, while using deep learning techniques to increase classification accuracy. The study focuses on classifying tweets posted in the English language on the Twitter platform, using a dataset obtained from Kaggle. A long-short-term memory (LSTM) network architecture improved the model performance. The results showed that the proposed model was able to achieve up to 87% accuracy in classifying tweets, highlighting its effectiveness in this field. This model enables reviews to be automatically categorized, which helps customers in their search for goods or services before transacting and saves time. Additionally, marketers can use it to find out what audiences think about their brands and products.

**Keywords:** Classification, Social media, Deep learning, Twitter, LSTM

## 1. Introduction

Over time, we have seen tremendous growth in the use of social media and social networking as a result of increased access to the Internet. Nowadays, social media has become an indispensable part of people's lives, as they consider it the main means of daily communication [1]. Individuals can express their opinions freely and openly on social media without having to reveal their identity, which sometimes leads to disagreements. However, some people abuse this freedom excessively, which can ultimately reduce meaningful online discussions and seeking the opinions of others [2].

The number of users of social media networks is witnessing tremendous growth, which increases the volume of data generated daily and makes textual content that greatly influences the decisions of individuals and groups the main type of information available on these platforms [3]. Although significant advances in social media comment classification models, considerable challenges remain regarding classification accuracy. Many existing models suffer from high error rates when dealing with diverse and constantly changing feedback. This poses a major challenge, as inaccurate classification can lead to misleading results and limit the system's effectiveness. Our study aims to improve classification accuracy by introducing an innovative LSTM model to effectively address this problem. With this approach, we seek to fill the current research gap and provide a more effective and accurate solution for classifying comments on social media. The models were trained using data from the Twitter network. It is one of the social networks that has been researched more than others compared to other social networks [4]. In addition to saving time, feedback grading procedures will protect users, by determining whether they are positive, negative, or neutral. The following a contributions of the study:

- A word embedding layer was used to convert words into digital representations.
- The LSTM model was used for text classification, which can effectively understand and analyze text word sequences. It works to identify the context and chronological order of words.
- Increased accuracy of comment classification compared to previous research using the same dataset.

This paper is organized as follows: Previous studies are included in Section 2. Section 3 details the method used. The study's experimental findings are shown in Section 4. The study's discussion and future duties are covered in Section 5.

## 2. Related Works

This section presents relevant work by a group of authors and the procedures they used:

Rustam, et al. (5), the study examines the growing use of social network data, especially Twitter, it suggests a voting classifier (VC) based on stochastic gradient descent classifier (SGDC) and logistic regression (LR). The VC uses soft voting to combine the LR and SGDC forecasts. With TF and TF-IDF, the voting classifier achieves an accuracy of 78% and 79%, respectively.

Tusar, et al. (6), this study used multiple machine learning approaches and NLP to analyze sentiment in US airline Twitter data. The paper introduces NLP approaches like BoW and TF-IDF, along with ML classification algorithms such as SVM, LR, MNB, and RF, to analyze sentiment in datasets with multiple classes. The best-performing approaches achieve 77% accuracy

Kumar, et al. (7), research uses an airline Twitter dataset from Kaggle to investigate sentiment analysis in the airline sector using machine learning techniques. Three machine learning approaches are used in the study: neural networks, SVM, and decision trees. The assessment indicates that the neural network-oriented method attains a maximum accuracy of 75.99%. The main objective is to assess sentiment polarity (positive, negative, or neutral) in tweets better to understand customer comments in the ever-changing aviation industry.

Abdullah et al. (8), to classify comments, the authors integrate deep learning techniques with machine learning techniques, such as LR, SVM, DT, AdaBoost, RF, RBF SVM, Gradient Boosting, Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), and LSTM. This results accuracy of 85.76% for the study.

Sengul, et al. (9), the study explore sentiment analysis on Twitter data using different algorithms, including Naive Bayes (NB), SVM, LR, LSTM, and CNN. After pre-processing English tweet messages from the Sentiment140 dataset, the LSTM+CNN hybrid approach achieved the highest accuracy of 85%. The evaluation was based on the F1 score, which demonstrates the effectiveness of these methods in classifying emotions on digital social networks.

Kalluri. (10), this thesis explores sentiment analysis using deep learning models of textual data, with a focus on social media tweets. The study compares the performance of the algorithms, noting challenges in natural language processing. Experience indicates that the CNN model achieves a balance between speed and accuracy, while the Recurrent Neural Networks (RNN) model, with word embedding, is more accurate but slower. RNN shows average results of 86%. The research aims to understand the basics of deep learning, optimal methods for classifying tweets, address time gaps, and determine the best model group.

Patel, et al. (11), in this research, the airline reviews dataset is analyzed using sentiment analysis. Multiple machine learning methods, including NB, SVM, and DT, were employed to classify sentiment performance into negative, positive, and neutral categories. Since the BERT method outperformed the other machine learning algorithms in the research and earned the greatest accuracy of 83% on the airline data set, its performance was compared to that of the other algorithms.

## 3. Methodology of proposed methods

This part explains the social media comment classification dataset and pre-processing for tweets before starting classification as well as a suggested methodology for classifying comments on the selected dataset. The basic stages in this system are depicted in Figure 1, each part in the Figure is illustrated in the next sections.

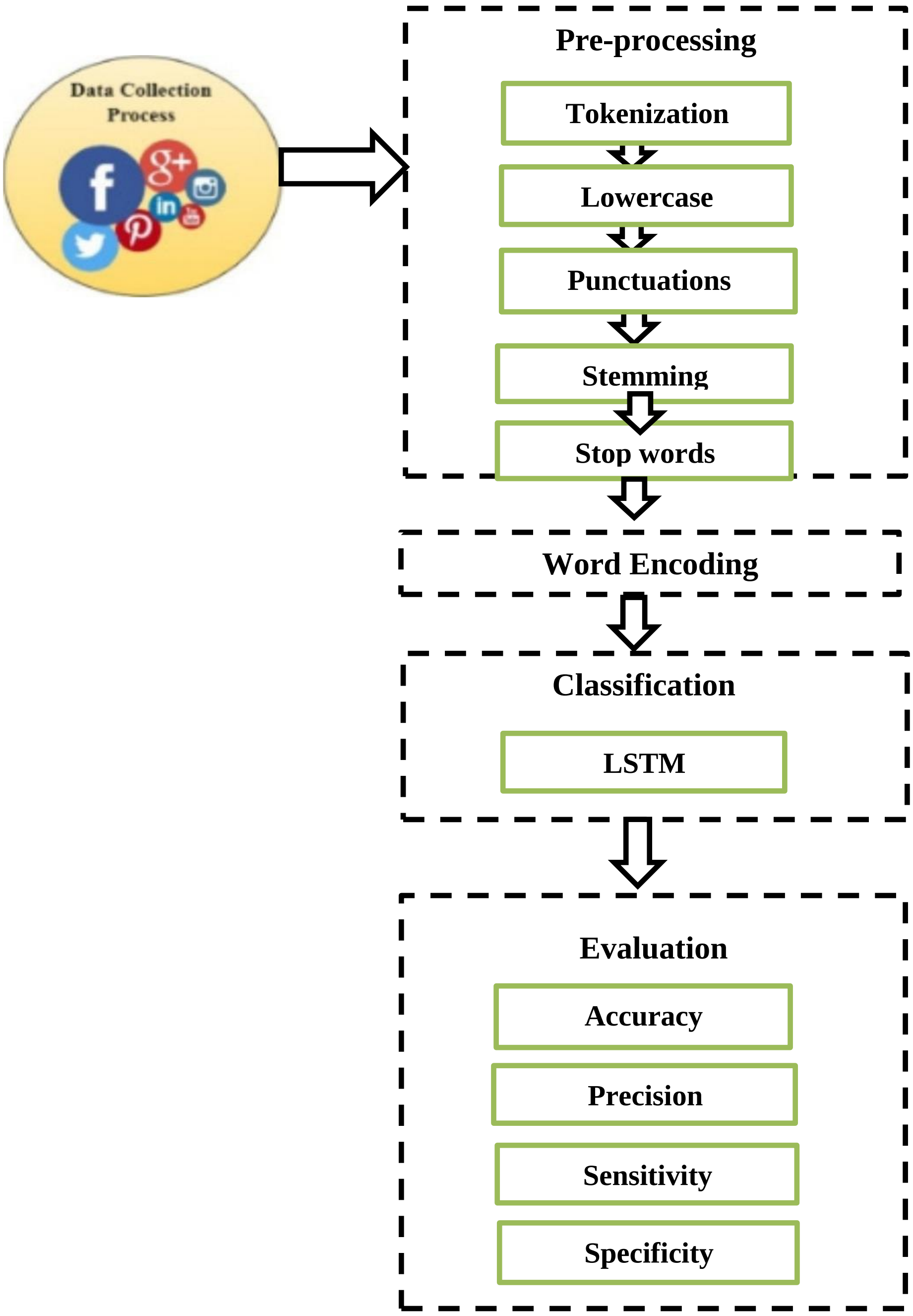


**Figure 1. Classification framework.**

### 3.1 Data Description

To classify comments as positive, negative, or neutral. The "twitter-airline-sentiment" dataset was obtained from Kaggle and used in this study [12]. There were a variety of neutral, negative, and positive remarks in the tweets. The opinions of travelers are represented in the dataset through tweets. A total of 14640 entries contain tweets from six US airlines. Figure 2 shows the distribution of tweets in the dataset, according to the class and frequency.

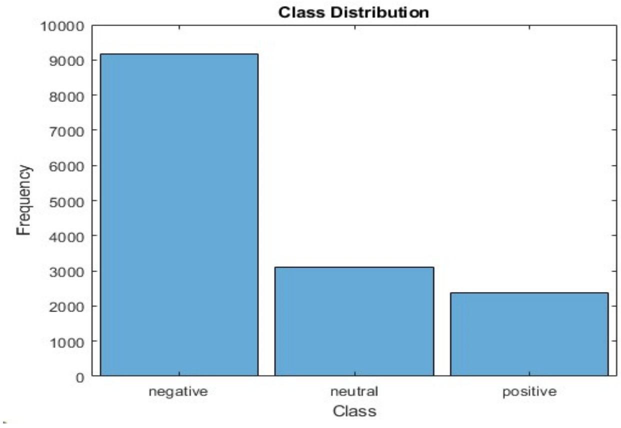


**Figure 2. Positive, negative, and neutral tweet frequencies in a dataset.**

### 3.2 Pre-processing

The suggested system will use natural language processing methods to preprocess the gathered data. The actions listed below will be taken [13]:

1. **Tokenization:** The process of tokenization divides text into distinct words or tokens.
2. **Lowercase text conversion:** Converting all text to lowercase will lessen the dimensionality of the data and enhance model performance.
3. **Eliminating punctuation:** You can omit punctuation since it adds nothing to the text's meaning.
4. **Stemming:** Stemming is the process of breaking down words into their stem or basic form. This will reduce the dimensionality of the data and improve the performance of the model [14].
5. **Eliminating stop words:** Words are instances of common terms that have no additional meaning in the text. Getting rid of stop phrases can improve the model performance by lowering the dimensionality of the data. Table 1 displays most of the stop words that were included in this study.

**Table 1. Samples of stop words**

| |
|---|
| The, a, an, they, there, this, these, what, where, who, that, with, why, about, is, are, has, have, was, were, in, on, of, under, before, after, might, be, been, do, most, up, from |

### 3.3 Word Encoding

Coding is a fundamental process in natural language processing (NLP) that aims to convert words from textual form to digital form. This is achieved by assigning a unique number to each word in the dataset, allowing textual data to be converted into a digital representation that can be understood by deep models such as neural networks. The encoding process allows words and sentences to be represented directly by numbers, making them easier to use as input for deep models. All sequences must be the same length so that they can be entered into the form regularly. If there is a discrepancy in the length of sentences, the length of the sequences should be standardized using padding, i.e. filling short sequences with a certain value (such as zero) until all sequences are the same length. The sequences are then compiled into a dataset that can be used to train the deep model.

### 3.4 Classification Model LSTM

An improved kind of recurrent neural network (RNN) called an LSTM network was created to address the vanishing gradient issue that conventional RNNs have. It is regarded as one of the most significant advancements in the recurrent neural network discipline. Because of its long-term information preservation capabilities, LSTM is perfect for jobs involving the processing of text, audio, and video that call for long-term context understanding. An LSTM cell consists of several major components that work together to control the flow of information within the cell. Among these components are the input gate that controls the extent to which new information is entered into memory, the forget gate that decides which part of the stored information will be forgotten, and the output gate that determines what will be produced from the cell at the current time [15].

This complex structure enables LSTM to precisely control what information is stored and discarded, enhancing its ability to handle sequential data. At each time step, the LSTM cell updates its state based on the current input and the previous memory state. The process begins with a forget gate that identifies unnecessary information that should be deleted from memory. Next, the input gate decides which new information to store in memory. Finally, the output gate determines what information will be output based on the current memory. This integrated process allows LSTM to retain important information for long periods and deal effectively with sequential data [16]. The LSTM model steps can be explained in algorithm 1.

**Algorithm 1: LSTM model.**

**Input:** Raw comments.

**Output:** LSTM Accuracy

**Begin**

**Step1.** Read the comment on the dataset.

**Step 2.** Comment pre-processing.

**Step 3.** Split data into 80% train and 20% test.

**Step 4.** Generate a word cloud to visualize frequently occurring words in the training data.

**Step 5.** Encoding words to convert documents into numerical format.

**Step 6.** Convert text documents into sequences of fixed length (sequence Length).

**Step 7.** Build LSTM model architecture with layers for sequence input, word embedding, LSTM, fully connected, softmax, and classification.

**Step 8.** Train the LSTM model.

**Step 9.** Predict comment labels for the test data

**Step 10.** Calculate accuracy and display the confusion matrix.

**End**

### 3.5 Evaluation Metrics

Several performance metrics, including accuracy, precision, sensitivity, and specificity, can be computed using the values in the confusion matrix. These metrics reveal the classifier's advantages and disadvantages .The metrics utilized in this thesis are determined using the following equations:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{2}$$

$$\text{Sensitivity}=\frac{TP}{TP+FN} \tag{3}$$

$$\text{Specificity}=\frac{TN}{TN+FP} \tag{4}$$

## 4. Experimental Result

The training phase of the dataset consists of 80%, whereas the test phase comprises 20%. After 30 training epochs, the LSTM-based model obtained an accuracy of 87% on the test set. Table 2 shows the parameters that were used to obtain these results, these parameters include the following:

**Table 2. The main parameter of LSTM**

| Parameter name | Value |
|---|---|
| Embedding dimension | 100 |
| Hidden units | 80 |
| Number of epochs | 30 |
| Batch size | 64 |
| Optimizer | Adam |
| Gradient threshold | 2 |
| Activation function | Softmax |
| Dropout | 0.02 |
| Initial learning rate | 0.001 |

Table 3 displays the model's evaluation of the test partition in classifying comments into three categories positive, negative, and neutral. The table shows three performance measures precision, sensitivity, and specificity for each category. Based on the results, we note that the positive category has the best overall performance due to the very high specificity of 0.98 and good precision of 0.84.

**Table 3. Results of LSTM**

| LSTM | | | |
|---|---|---|---|
| | **Negative** | **Neutral** | **Positive** |
| **Precision** | 0.92 | 0.72 | 0.84 |
| **Sensitivity** | 0.93 | 0.74 | 0.73 |
| **Specificity** | 0.82 | 0.94 | 0.98 |

Figure 3 shows generally good performance with little variation between different classes. The neutral category seems to be the most challenging for the model as the sensitivity is lower compared to the other categories. Species with high exemplar values indicate strong performance in correctly distinguishing specimens within those classes

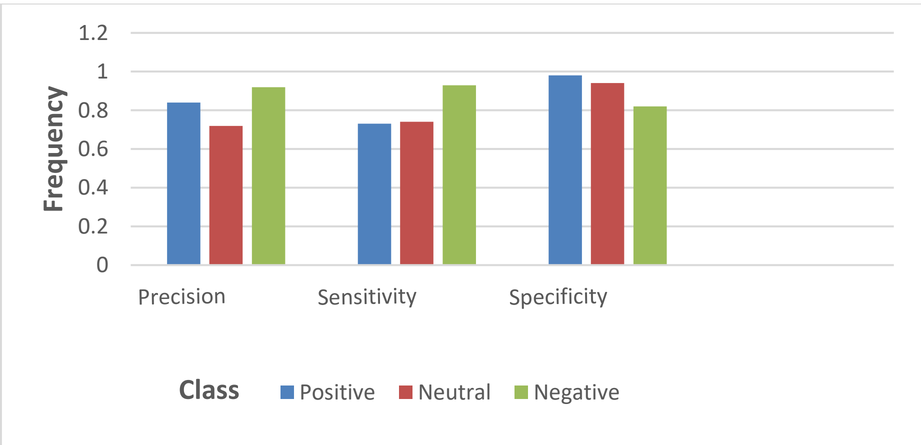


**Figure 3. The results of the LSTM model.**

After comparing the suggested system with earlier studies conducted on the same Twitter database, it was discovered that our suggested model outperformed the earlier models as shown in Table 4.

**Table 4. A comparative study with related work**

| Paper | Accuracy |
|---|---|
| Rustam, et al [5] | 79 % |
| Tusar, et al [6] | 77% |
| Kumar , et a. [7] | 75.99% |
| Rane, et al [17] | 84.5%. |
| Saad [18] | 83.31%. |
| Patel, et al.[11] | 83 % |
| **Proposed Model** | **87%** |

## 5. Conclusion

Deep learning and natural language processing techniques are combined in this study to analyze a sizable, uneven, multi-class dataset that was obtained from Twitter datasets, namely the airline sentiment dataset. The goal was to test the proposed model for classifying comments. This collection includes multiple category labels and Tweets from real customers, which helps categorize comments into positive, negative, or neutral. To achieve this, an LSTM network was used to classify these tweets. The model achieved an accuracy of 87%, which shows the effectiveness of this model in text classification. In the future, efforts will focus on improving the model to include recognition of different ratings contained within a single comment. We will highlight parts of comments that are considered negative and alert the user to any offensive content.

## REFERENCES

[1] Ozoh, P. A., Adigun, A. A., & Olayiwola, M. O. (2019). Identification and classification of toxic comments on social media using machine learning techniques. International Journal of Research and Innovation in Applied Science (IJRIAS), 4(11), 142-147.

[2] Poojitha, K., Charish, A. S., Reddy, M., & Ayyasamy, S. (2023). Classification of social media Toxic comments using Machine learning models. arXiv preprint arXiv:2304.06934.

[3] Jalil, A. A., & Aliwy, A. H. (2023). Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model. Al-Bahir Journal for Engineering and Pure Sciences, 2(2), 12.

[4] AminiMotlagh, M., Shahhoseini, H., & Fatehi, N. (2022). Reliable sentiment analysis for classification of tweets in social networks. Social Network Analysis and Mining, 13(1), 7.

[5] Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets classification on the base of sentiments for US airline companies. Entropy, 21(11), 1078.

[6] Tusar, M. T. H. K., & Islam, M. T. (2021, September). A comparative study of sentiment analysis using NLP and different machine learning techniques on US airline Twitter data. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT) (pp. 1-4). IEEE.

[7] Ravi Kumar, G., Venkata Sheshanna, K., & Anjan Babu, G. (2021). Sentiment analysis for airline tweets utilizing machine learning techniques. In International Conference on Mobile Computing and Sustainable Informatics: ICMCSI 2020 (pp. 791-799). Springer International Publishing

[8] Al Taawab, A., Tasnia, L., Dhar, M., & Mehedi, M. H. K. (2022, September). Transliterated Bengali comment classification from social media. In 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 365-371). IEEE.

[9] Şengül, F., & Yılmaz, E. K. (2023). SENTIMENT ANALYSIS BASED ON MACHINE LEARNING METHODS ON TWITTER DATA USING oneAPI.

[10] Kalluri, S. (2023). Deep Learning Based Sentiment Analysis (Master's thesis). Blekinge Institute of Technology, Karlskrona, Sweden.

[11] Patel, A., Oza, P., & Agrawal, S. (2023). Sentiment analysis of customer feedback and reviews for airline services using language representation model. *Procedia Computer Science*, *218*, 2459-2467.

[12] https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment.

[13] Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, *25*, 319-335.

[14] Yang, S., & Zhang, H. (2018). Text mining of Twitter data using a latent Dirichlet allocation topic model and sentiment analysis. *International Journal of Computer and Information Engineering*, *12*(7), 525-529.

[15] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780

[16] Salehi, A. W., & Gupta, G. (2021). A prospective and comparative study of machine and deep learning techniques for smart healthcare applications. Mobile Health: Advances in Research and Applications, 163-189.

[17] Naresh, A., & Venkata Krishna, P. (2021). An efficient approach for sentiment analysis using a machine learning algorithm. *Evolutionary Intelligence*, *14*(2), 725-731.

[18] Saad, A. I. (2020, December). Opinion mining on US Airline Twitter data using machine learning techniques. In *2020 16th International Computer Engineering Conference (ICENCO)* (pp. 59-63). IEEE.