ISSN 2723-6471

Opinion Mining in Text Short by Using Word Embedding and Deep Learning

Shaima Mahdi Orebi[1,*], Asmaa Mohsin Naser[2]

[1]*Department of Information Technology, College of Computer Science and Mathematics, University of Thi Qar, Iraq*

[2]*Department of Computer Science, College of Computer Science and Mathematics, University of Thi Qar, Iraq*

(Received: August 25, 2024; Revised: October 6, 2024; Accepted: November 21, 2024; Available online: December 30, 2024)

Abstract

Recently, with the increasing use of the Internet by people, millions use social media sites on a daily basis to express their opinions, suggestions and reactions about a new product or a specific topic. Through these views, the principle or topic of sentiment analysis. especially for text data (tweets), where classification techniques are used for the purpose of classifying these text tweets. Sentiment classification is a common and important in the field of natural language processing. Our study aims to utilize word embedding model. Word embedding is used to convert text words into vectors for word representation, capturing the semantic and syntactic relationships between words. It contributes by presenting a comparison and analysis of word embedding model and deep learning techniques. In this research, we propose to analyze sentiments or opinions using word embedding Global Vectors for Word Representation (GLOVE) with Bidirectional LSTM neural networks and Long Short-Term Memory (LSTM). Where we relied on a deep learning model that combines the power of word representations in (GLOVE) and (LSTM)'s ability to understand linguistic context. This model showed good performance in sentiment classification, which indicates its effectiveness of combining the two models. Here we used tweet dataset regarding (Generative Pre-trainer Transformer), which is one of the tools of generative artificial intelligence, Dataset :(CHATGPT sentiment analysis) CHATGPT Tweets first month of launch. We analyzed the data or tweets about the opinions and sentiments of tweeters. The use of the word embedding model with short-term memory (BILSTM and LSTM) achieved good results about 89% and 90%. According to the performance metrics used (confusion matrix, accuracy, precision, recall, F1 score), compared with the results of the (WORD2VEC) model. These metrics are vital tools for evaluating sentiment analysis models and measuring the model's ability to correctly classify tweets into good, bad, or neutral sentiments.

Keywords: Text Mining, NLP, Sentiment Analysis, Word Embedding, Deep Learning

1. Introduction

Sentiment analysis, or opinion mining, is the process of analyzing texts or studying people's opinions, comments, and emotions to determine the underlying opinion behind a particular text and to distinguish linguistic patterns associated with particular feelings, whether positive, negative, or neutral. The study of sentiment analysis is an NLP application also known as opinion mining, examines people's attitudes and feelings about many types of things, including goods, services, organizations, people, issues, events, subjects, and their qualities. It stands for a significant issue area. The major focus of sentiment analysis and opinion mining is on opinions that explicitly or implicitly indicate positive or negative feelings [1], [2], [3], [4].

Twitter may be a fantastic medium for opinion formation and presentation, but it also poses new and unusual challenges. such as Twitter faces a number of challenges, including the presence of many irrelevant or sarcastic tweets, as well as the complexity of the language and the use of slang and abbreviations that are difficult to understand. Without effective tools for evaluating those opinions to accelerate their consumption, the process would be ineffective. Over time, it has become clear that employing sentiment analysis technologies to pinpoint certain attitudes and emotions is the ideal strategy [5].With the increase in activities on the Internet, including booking tickets, e-commerce, chatting, communicating through social networking sites, conferences, microblogs, etc., these opinions and responses can be used in various applications in real life, including political and social events,

***Corresponding author: Shaima Mahdi Orebi (shiama_mahdi@utq.edu.iq)**
DOI: https://doi.org/10.47738/jads.v6i1.438

 ISSN 2723-6471

marketing movement, and customer preferences for products ,business, social sciences. Sentiment analysis applications go beyond simply determining whether text is positive or negative. Sentiment analysis techniques can help companies better understand their customers, improve user experience, and make informed decisions. Social media comments can be used to identify common issues, allowing them to improve their product or service [6]. Which increases its importance in a reality full of daily increasing data.

On November 30, 2022, CHATGPT, an advanced language model powered by artificial intelligence created by OPENAI, was made available to internet users. It has since garnered a lot of interest. The emotions and ideas of persons who first used CHATGPT are crucial input for assessing the usefulness and advantages and disadvantages of this technology [7], [8]. It is considered a qualitative leap in the field of natural language processing, as it has the ability to generate texts that resemble human texts, as well as conduct conversations, as it has the ability to write codes, translate languages, and write academically, as well as its ability to understand complex linguistic context [9].

In this paper, the analyze how social networking site users feel when they tweet about their opinions and the newest tools and techniques in artificial intelligence, such as the emergence of the generative chat bot, which now responds to every question posed to it in order to gauge how positive or negative a topic is, using machine learning and deep learning techniques. In order to extract of importance characteristics from the raw data, deep learning is a subset of machine learning techniques based on neural networks. Many other industries, including signal and image processing, have employed it. As a result of holding several neural networks, deep neural networks allow the output of one network to be used as an input for the next network, and so on. Deep learning researches the characteristics of text data independently, as well as the many layers of features that go into predicting that data. Deep learning networks have wide used in the field of natural language processing [10], [11].

In this work, 219,294 tweets were collected, taken from Kaggle [16] in CSV format. The percentage of bad tweets was higher than other tweets.

Our study aims to: Utilize word embedding model in addition to deep learning techniques. For the purpose of analyzing text data tweets (CHATGPT sentiment analysis). Where it contributes by presenting a comparison and analysis of word embedding model and deep learning techniques. And also provides new data, as the topic under discussion is modern and its data are modern in the field of sentiment analysis and the use of embedding models with bidirectional networks.

Despite the great development in the field of sentiment analysis using machine learning techniques, there is a need for more research to develop a method capable of in-depth analysis of users' sentiments towards interactive tools. The previous study that used machine learning in its study achieved good results, but it was limited in sentiment analysis to two categories: positive and negative. In this study, we seek to use deep learning tools and word embedding models to represent tweets to analyze users' sentiments towards one of the artificial intelligence tools (CHATGPT sentiment analysis) with three categories: bad, good and neutral.

The rest of the paper, section two deals with Review of the literature, the section three deals with the Methodology, and the section four deals The Results and Discussion, and section five is the Conclusions.

2. Literature Review

Sentiment classification for sentiment analysis is one of the important fields for knowing people's thoughts and opinions using various classifiers. This part presents some related works. This research used a method that combines convolutional neural network (CNN) and long short-term memory (LSTM) models to predict the sentiment of Arabic tweets. The Arabic Sentiment Tweets Dataset (ASTD) was used [12]. In this research, convolutional neural networks were used for sentiment classification of text data. CNN is a deep learning model that extracts features of interest from input data by running a “convolutional” filter (kernel) through the data. Three types of movie review (MR) datasets from the Kaggle data site were used [13]. This paper used English text data for sentiment analysis. Specifically, they analyzed sentiment classification methods that have been focused on deep learning over the past five years. In addition, sentiment analysis of COVID-19 tweets, where two datasets were used. One dataset contained all tweets issued from December 2019 to May 2020, and the second set prioritized tweets that were more

 ISSN 2723-6471

retweeted2021.It used deep learning techniques of CNN and LSTM to analyze sentiment on Twitter data, and uses Twitter data for sentiment analysis, as well as the IMDB dataset of 50,000 movie reviews to train deep learning models [2].This work on Twitter data analysis proposed an adaptive deep recurrent neural network method (ADRNN) [14].The use of LSTM during the classification process. The LSTM algorithm is applied to each test vector for class classification. This paper used the data used in these user comments related to CHATGPT on YouTube [15]. The current study is unique from previous studies in that it provides new data, as the topic under discussion is modern and its data are modern in the field of sentiment analysis and the use of embedding models with bidirectional networks.

3. Methodology

The methodology is described used for the purpose of analyzing the tweet texts of the dataset is presented in this section. All the steps followed in the research are below. The first step begins with collecting data and the steps continue by classifying tweets into good, bad and neutral using deep learning algorithms until reaching the model evaluation.

3.1. Data Collection

It has been (CHATGPT) a major topic of conversation about the latest developments in the world of technology. All tweets about CHATGPT from 11/30/2022 to 12/31/2022. This dataset contains tweets about CHATGPT in the first month of its launch and includes a range of information, Tweet content: This includes the actual text of the tweet, which can be anything from users asking CHATGPT questions, expressing their opinions about it, or sharing their experiences using the AI model. Sentiment analysis: The dataset also includes enough information to perform sentiment analysis scores for each tweet, indicating whether the tweet expresses positive, negative, or neutral sentiment towards CHATGPT. The geographic location or country of the tweeters or commenters is not mentioned. In this study, we used the tweet dataset to analyze the sentiment about (CHATGPT sentiment analysis) as this dataset is available on the Kaggle [16] community. Selected dataset in English. The data set under study consists of about (219294) Text tweets. Three categories of this data, including (107796) bad tweets, positive tweets (56011), and (55487) neutral tweets, shown in the figure 1(a). It contains some repetitive and unbalanced data. In the data set used in the study, the bad class data was higher than the other classes. In order to adjust the distribution of the data to be more balanced, fewer samples were taken from the largest category, which is the category of bad tweets, and a random sample was taken from it to maintain a fair representation of the class. In order to make the distribution of the three classes close or close to equality. Because when there is no balance in the data, the model tends to be biased towards the largest class, which leads to a decrease in the classification accuracy of the smaller classes. It has been organized and balanced as in the figure 2 (b). This dataset contains three columns (id, tweet, and labels).

3.2. Preprocessing

Because the data preprocessing process is considered an important step for all classification tasks, we will perform the preprocessing steps on this text dataset as this stage includes several steps, including: converting texts to lower case letters, deleting stop words, question marks, removing links, removing hash tags and removing @ with removing numeric, as well as stemming words using the famous Porter algorithm. The pre-processing steps are done because removing links and tags do not carry any meaning in sentiment analysis and are considered noise. Converting capital letters to lowercase to ensure that these words are similar but in lowercase and uppercase letters they are treated as the same word. Tokenization where it divides the text into unique words and is the first step in any text processing process. Stemming returns words to their roots, i.e. returns the word to its origin, where the number of unique words in the data is reduced and the efficiency of the model is improved. Removing stop words as they do not carry much contextual meaning and are noise in the sentiment analysis process.

The effect of removing unwanted words and other elements improves accuracy by focusing on words that are important in determining sentiment, as well as reducing the amount of data that the model processes. After that, the process of balancing the data is completed, since negative tweets are higher than other tweets. Therefore, the data was balanced, duplicate data was deleted, and then the data was coded into (0, 1, and 2). Where the first is bad, the

second is good, and the third is neutral. Figure 2 shows us the three categories as a word cloud.no.of data ((219294 rows and 3 column).

3.3. Word Embedding

Word representation or word embedding is one of the important and basic steps for the purpose of classifying textual data and in the field of natural language processing. To apply any type of algorithms, whether in machine learning or neural networks in deep learning, words must be converted into vectors to represent the word. Word embedding model provides the semantic relationships of the words. It is a vector representation of the word. Embedding models capture the similarity of the grammatical and semantic context of a word in the same document as well as the relationship of that word to other words. The location of the word is known based on the meaning and context [17]. In this research, we use one of the embedding models. The GlOVE is a representation of words in vectors that has several a pre-trained words embedding files [18]. After pre-processing and cleaning the texts, they are converted to numbers using the Keras library, and then we use the pre-trained word embedding (GLOVE) with dimensions (100) to find the semantic context of the word. Another embedding model is used, which is (WORD2VEC) for the purpose of comparing the results with our model (GLOVE). (WORD2VEC) is considered one of the common methods for learning word inclusion, especially in the field of sentiment analysis or classification. Where (WORD2VEC) maintains the meaning of grammatical words, so that words that are similar to each other remain close to each other, and it also works to arrange these words according to the similarity between them. It works to represent a vector for each word [19]. In addition, word embedding models faces some limitations. These include the difficulty of understanding the meanings of words, as they depend on the representation of each word and do not depend on the general context. Also, for complex or compound human emotions, it is difficult to understand the interactions between them. Such limitations include linguistic ambiguity and conflicting sentiments. In such cases, the model may have difficulty recognizing opinions or sentiments in a sentence or text [20], [21].

3.4.Sentiment Analysis Using BILSTM and LSTM

After creating an inclusion matrix using (GLOVE). Then, two types of recurrent neural networks are used, namely BILSTM is used and LSTM. We used a bidirectional layer LSTM, which deals with the sequence from left to right and the other from right to left. Using it allows the model to process information in both directions, which improves its ability to understand the context. The layer(LSTM with (64) of hidden units, which in turn is responsible for storing information in the long term. Also, the layer (Dropout) was used to prevent overgeneralization. Also, the layer (Dense) with (3) units with the activation function (softmax) was used. Recently, this type of network has been widely used in the field of natural language processing, especially in sentiment analysis, to analyze and classify textual data. where it is fed with the required information: the length of the text, the number of units, the external layers, the output layer, and other parameters that are listed in the table 2.

4. Results and Discussion

4.1. Performance Measurement

In this study, we used the four well-known performance evaluation metrics in both machine learning and deep learning. They depend on the set of correct classes with the set of classes predicted by the classifier. These metrics are described Precision, Recall, F1-score, Accuracy) in the equations (1,2,3,4) respectively.

Precision It represents the percentage of correct positive predictions or forecasts among all positive predictions or forecasts. The concept of Precision expresses the extent of the model's reliability in classifying emotions, i.e. the higher the Precision, the more accurate the model is, which means that it expects certain emotions to be correct precision is measure shows the total number of positive predictions. It is determined by dividing the total number for positives that are expected based on the total number of positives that are classified, the accuracy is defined as follows:

$$\text{Precision} = \text{True Positive}/\text{True Positive} + \text{False Positive} \quad (1)$$

Recall represents the percentage of correct positive predictions or predictions among all actual positive samples. It expresses the extent to which the model can identify all positive examples. If it is high, this means that the model

 ISSN 2723-6471

does not miss many positive cases. Recall is a measure of the ratio of all positively classified and correctly identified labels to all positively classified labels. The call is defined as follows:

$$\text{Recall} = \text{True Positive}/\text{True Positive} + \text{False Negative} \quad (2)$$

F-score is the harmonic mean between precision and recall, which is considered a comprehensive measure of performance as it combines precision and recall and is especially useful when the false positive is equally high. The F1 score scale uses information (precision and recall) and is defined as follows:

$$F1 = 2 \times \text{precision} \times \text{recall}/\text{precision} + \text{recall} \quad (3)$$

As well accuracy refers to the accuracy of the prediction and is calculated as follows:

$$\text{Accuracy}=\text{TruePositive} + \text{TrueNegative}/\text{TruePositive} + \text{TrueNegative} + \text{FalsePositive} + \text{FalseNegative} \quad (4)$$

One of the measures that was also used in the research for the purpose of evaluating the model's performance is the confusion matrix, In the confusion matrix, the rows represent the class instances predicted by the classifier, while the columns represent the true class instances, shown in table 1.

Table 1. Confusion Matrix

Predicted Values	**Actual Values**	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Negative	True Negative

4.2. Results

In order to classify text data into bad, good and neutral, an experimental study was conducted on the Twitter dataset (CHATGPT sentiment analysis) using deep learning and word embedding methods. The word embedding model (GLOVE) was tested against the other type of embedding (WORD2VEC) using the same parameters for both models and shown in the table 2. The study was conducted entirely using Python language with many private libraries in the work environment.

In this section, the experimental results of the dataset are analyzed. Also, as we notice in the table 3 which Shows the performance of each category with the four measures, we notice that using (GLOVE and BILSTM) It gives good performance compared to the other embedding model. Although both models use the same parameters, (GLOVE and BILSTM) gave good performance because it takes into account the common frequencies between words more accurately. Also, the vectors produced with (GLOVE) better represent the meaning and connotations inherent in the words. Its ability to deal with rare words, in addition to better agreement between (GLOVE) with (BILSTM) which improves the performance of the model. In figure 3 is the confusion matrix for the GLOVE and BILSTM, where (11,075) out of 11,963 bad emotions were classified correctly, while (10,087) out of 10,937 good emotions were classified correctly, and (9,066) out of 10,647 neutral emotions were classified correctly. The figure shows (4) a visualization of the confusion matrix for the GLOVE and LSTM, where (16,648) out of 18,040 bad emotions were classified correctly, while (15,189) out of 16,426 good emotions were classified correctly, and (13,208) out of 15,855 neutral emotions were classified correctly. While the figure shows us (5) is the confusion matrix for the WORD2VEC and BILSTM, where (18,581) out of 20,948 bad emotions were classified correctly, while (12,531) out of 16,264 good emotions were classified correctly, and (9282) out of 15,667 neutral emotions were classified correctly.

Figure 6 shows us the confusion matrix for (WORD2VEC and LSTM), where (18,718) out of 20,948 bad emotions were classified correctly, while (11,233) out of 16,264 good emotions were classified correctly, and (7,159) out of 15,667 neutral emotions were classified correctly. Observing the confusion matrix in figure 3 and figure 4, which we can observe the classification evaluation when using (GLOVE) with each of (BILSTM and LSTM) compared to the confusion matrix in figure 5 and figure 6, which explains or depicts the classification evaluation when using

ISSN 2723-6471

(WORD2VEC) with (BILSTM and LSTM). Where we notice the high performance of the models when use (GLOVE).

Table 2. Shows the parameters used in training a deep learning model.

Parameter	Value
Num.Embedding	100
Dropout rate:	0.3
Num.epoch:	10
Num. batch-size:	128
Optimizer:	Adam
Activation	softmax
loss	categorical crossentropy
Max-len	140

Table 2 shows the parameters that were used with neural networks and the word embedding model. As these parameters play a crucial role in determining the performance of the model and how it learns from the data. After several experiments, these parameters were adjusted to obtain the results. Here is a description of each parameter and what it means. Num.Embedding:this parameter specifies the size of the vector representing each word in the dictionary. Dropout rate: this technique is used to prevent the model from over-memorizing the training data.Num.epoch:Num. batch-size: It is a complete cycle of feeding all training data to the model. Optimizer: It specifies how to update the model's parameters to minimize loss. Activation: This function is used to add an element of non-linearity to the model. Loss: It measures the difference between the expected output and the actual output. Max-len: It specifies the maximum sentence length that the model can process.

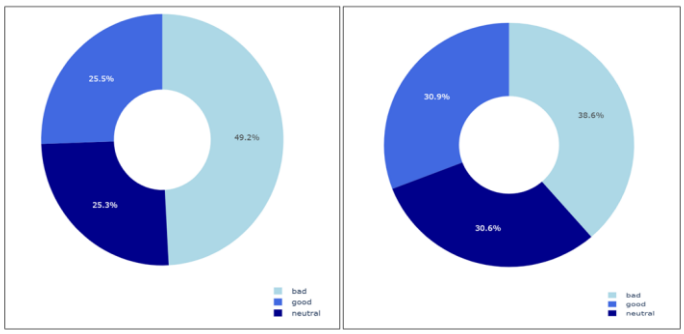
Table 3. Comparing the performance of different sentiment classification models

Algorithms	Accuracy	Precision			Recall			F1-Score		
		bad	good	neutral	bad	good	neutral	bad	good	Neutral
WORD2VEC and LSTM	0.70	0.70	0.77	0.62	0.89	0.69	0.46	0.78	0.73	0.53
WORD2VEC and BILSTM	0.76	0.78	0.82	0.67	0.89	0.74	0.62	0.83	0.78	0.64
GLOVE and LSTM	0.89	0.93	0.90	0.86	0.92	0.92	0.83	0.93	0.91	0.84
GLOVE and BILSTM	0.90	0.94	0.90	0.86	0.92	0.93	0.86	0.93	0.91	0.86

Table 3 presents a comparative analysis of four different models, namely (WORD2VEC and LSTM, WORD2VEC and BILSTM, GLOVE and LSTM, GLOVE and BILSTM). The models are evaluated through four main criteria (Accuracy, Precision, Recall, F1-Score). Where we notice that both precision and recall achieve good performance in general in the scales. This indicates that the model is able to accurately classify sentiment and found most of the positive cases. We also notice that the (F-score) scale gives a high value, which confirms the balanced performance of the model.

Figure 1 indicates the percentage of data: Figure 1(a) shows the dataset for the three categories (good, bad, and neutral) before balancing the data, where the percentage of bad data is the highest, which is about (49%). Figure 1(b) represents the percentage of the dataset for the three categories (good, bad, and neutral) after balancing the categories with close percentages to prevent the system from being biased towards one category compared to the other

 ISSN 2723-6471



(a) (b)

Figure 1. Percentage for the dataset: (a) before balanced (b) after balanced)

Figure 2 shows the word cloud for the three categories, where we notice that the frequency of the most frequently repeated words appears to be large, as figure 2(a) shows (the bad category) some words such as (thing ,bad, asked chatgpt ,using chatgpt) while figure 2(b) represents the good category some words such as (answer ,better,love,good) and as for figure (1), it indicates the neutral category (using ,request, work) .

(a) (b) (c)

Figure 2. The Word Cloud For dataset after preprocessing :(a) Bad (b) Good (c) Neutral

Figure 3 and figure 4: Confusion matrix of the model (GLOVE with BILSTM and LSTM) where the confusion matrix shows the model's performance in classifying emotions. On the test set, the rows represent the actual classes (bad, good, neutral), while the columns represent the predicted classes. The values on the main diagonal indicate the correct predictions, while the other values represent the errors that the model made. We notice that the model was more accurate in classifying bad and good sentiment compared to neutral sentiment).

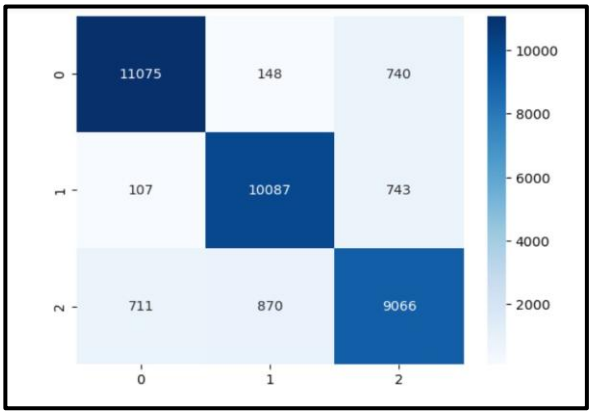


Figure 3. Confusion Matrix (GLOVE and BILSTM)

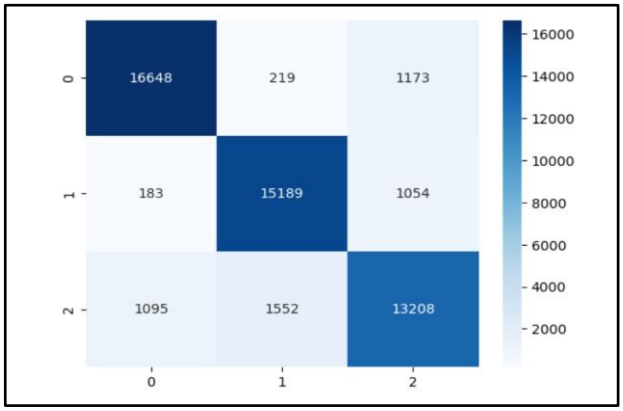


Figure 4. Confusion Matrix (GLOVE and LSTM)

Figure 5 and figure 6: Confusion matrix of the model (WORD2VEC with BILSTM and LSTM) where the confusion matrix shows the model's performance in classifying emotions. On the test set, the rows represent the actual classes (bad, good, neutral), while the columns represent the predicted classes. The values on the main diagonal indicate the

 ISSN 2723-6471

correct predictions, while the other values represent the errors that the model made. We notice that the model was more accurate in classifying bad and good sentiment compared to neutral sentiment.

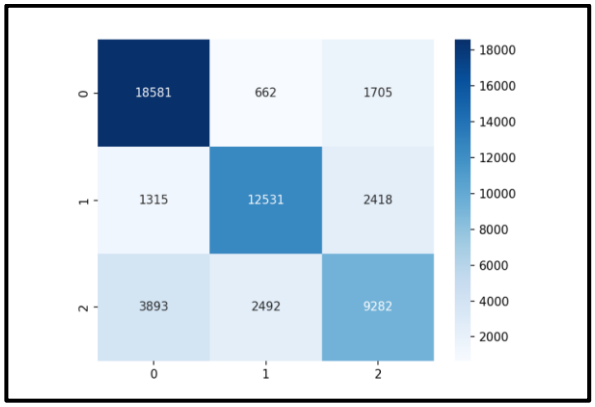


Figure 5. Confusion Matrix (WORD2VEC and BILSTM)

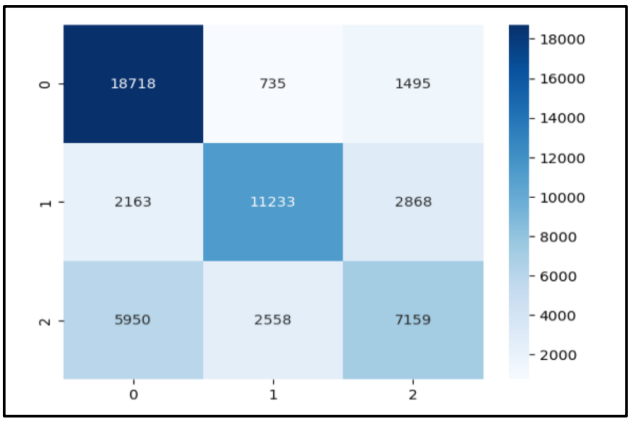


Figure 6. Confusion Matrix (WORD2VEC and BISTM)

5. Conclusion

Currently, with the spread of social media, sentiment analysis or analysis of user comments has become an important field, especially short texts, and for the purpose of analyzing and classifying these texts, such as blogs and tweets on Twitter. This paper presented two types of deep learning algorithms where they were combined with (GLOVE) pre-trained word embedding model (100 D). The algorithms were used to classify tweets. We used the topic of sentiment analysis. We analyzed the data or tweets about the opinions and sentiments of tweeters on Twitter about one of the artificial intelligence tools (Generative Pretrainer Transformer). This was done after pre-processing the data through several tools. From the experimental results, we notice that the system gives good results when combining (GLOVE) word embedding with (BILSTM and LSTM) compared to another embedding model which is (WORD2VEC). In our future work, the use of transformers such as the modified BERT transformer or other transformers. Transformers take into account the relationships between words in an entire sentence, not just the lexical meaning of individual words, and transformers are trained on massive amounts of data, giving them a deep understanding of texts. Instead of word embedding models, this modification is expected to improve the model's ability to understand the bidirectional context in texts, which in turn will improve the performance of tasks that require deep understanding in text summarization.

6. Declarations

6.1. Author Contributions

Conceptualization: S.M.O., A.M.N.; Methodology: S.M.O.; Software: S.M.O.; Validation: S.M.O., A.M.N.; Formal Analysis: S.M.O., A.M.N.; Investigation: S.M.O.; Resources: A.M.N.; Data Curation: A.M.N.; Writing Original Draft Preparation: S.M.O., A.M.N.; Writing Review and Editing: S.M.O., A.M.N.; Visualization: S.M.O. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

 ISSN 2723-6471

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Liu, "Sentiment Analysis and Opinion Mining", Morgan and Claypool Publishers, May 2012. https://www.cs.uic.edu/~liub/FBS/liub-SA-and-OM-book.pdf

[2] W. Etaiwi, D. Suleiman, and A. Awajan, "Deep Learning Based Techniques for Sentiment Analysis: A Survey," *IJCAI*, vol. 45, no. 7, pp. 89-95, Dec. 2021, doi: 10.31449/inf.v45i7.3674.

[3] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 81-97, Dec. 2021, doi: 10.1007/s13278-021-00776-6.

[4] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artif Intell Rev*, vol. 56, no. 8, pp. 8469–8510, Aug. 2023, doi: 10.1007/s10462-022-10386-z.

[5] A. Bello, S.-C. Ng, and M.-F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, pp. 506-517, Jan. 2023, doi: 10.3390/s23010506.

[6] Q. Li, X. Li, Y. Du, Y. Fan, and X. Chen, "A New Sentiment-Enhanced Word Embedding Method for Sentiment Analysis," *Applied Sciences*, vol. 12, no. 20, pp. 1-12, Oct. 2022, doi: 10.3390/app122010236.

[7] A. Korkmaz, C. Aktürk, and T. Talan, "Analyzing the User's Sentiments of ChatGPT Using Twitter Data," *ijcsm*, vol. 4, no. 2, pp. 202–214, May 2023, doi: 10.52866/ijcsm.2023.02.02.018.

[8] S. Soni, S. S. Chouhan, and S. S. Rathore, "TextConvoNet: a convolutional neural network based architecture for text classification," *Appl Intell*, vol. 53, no. 11, pp. 14249–14268, Jun. 2023, doi: 10.1007/s10489-022-04221-9.

[9] Y. Su and Z. J. Kabala, "Public Perception of ChatGPT and Transfer Learning for Tweets Sentiment Analysis Using Wolfram Mathematica," *Data*, vol. 8, no. 12, pp. 180-191, Nov. 2023, doi: 10.3390/data8120180.

[10] B. D. Lund and T. Wang, "Chatting about ChatGPT: how may AI and GPT impact academia and libraries?," *LHTN*, vol. 40, no. 3, pp. 26–29, May 2023, doi: 10.1108/LHTN-01-2023-0009.

[11] U. D. G, P. M. K, G. C. Babu, and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wireless Personal Communications*, vol. 2021, no. 1, pp. 1-10, Mar. 16, 2021, In Review. doi: 10.21203/rs.3.rs-247154/v1.

[12] M. Heikal, M. Torki, and N. El-Makky, "Sentiment Analysis of Arabic Tweets using Deep Learning," *Procedia Computer Science*, vol. 142, no. 1, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.10.466.

[13] H. Kim and Y.-S. Jeong, "Sentiment Classification Using Convolutional Neural Networks," *Applied Sciences*, vol. 9, no. 11, pp. 2347-2359, Jun. 2019, doi: 10.3390/app9112347.

[14] D. P. Kavitha, "Twitter Sentiment Analysis Based On Adaptive Deep Recurrent Neural Network," TURCOMAT, vol. 12, no. 9, pp. 2449-, 2021.

[15] T. H. Rochadiani, "Sentiment Analysis of YouTube Comments Toward Chat GPT," *JT*, vol. 21, no. 1, pp. 60-68, Aug. 2023, doi: 10.26623/transformatika.v21i2.7033.

[16] C. SA," chatgpt-sentiment-analysis," Kaggle datasets,2022. https://www.kaggle.com/datasets/charunisa/chatgpt-sentiment-analysis

[17] D. Dessì, D. R. Recupero, and H. Sack, "An Assessment of Deep Learning Models and Word Embeddings for Toxicity Detection within Online Textual Comments," Electronics, vol. 10, no. 7, pp. 779-799, Mar. 2021, doi: 10.3390/electronics10070779.

[18] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar: Association for Computational Linguistics, vol. 2014, no. 1, pp. 1532–1543, 2014. doi: 10.3115/v1/D14-1162.

 ISSN 2723-6471

[19] Md. Al-Amin, Md. S. Islam, and S. Das Uzzal, "Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words," in 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh: IEEE, vol. 2017, no. Feb., pp. 186–190, 2017. doi: 10.1109/ECACE.2017.7912903.

[20] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," Knowledge-Based Systems, vol. 89, no. 1, pp. 14–46, Nov. 2015, doi: 10.1016/j.knosys.2015.06.015.

[21] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artif Intell Rev*, vol. 56, no. 9, pp. 10345–10425, Sep. 2023, doi: 10.1007/s10462-023-10419-1.