Creative Research Publishers
https://www.crossref.org/06members/50go-live.html

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 5, March 2025 Series

Analyzing User Sentiments About Chatgpt Using Hybrid Deep Learning Model

[1]Agwi, Celestine Uche & [2]Akpojaro, Jackson
[1&2]Department of Mathematics & Computer Science
Faculty of Basic and Applied Sciences
University of Africa, Toru-Orua
Bayelsa State, Nigeria
E-mail: [1]celestine.agwi@uat.ed.ng; [2]jackson.akpojaro@uat.edu.ng
Phone Nos: [1]+2348065235687; [2]+2348138803299

ABSTRACT

Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". ChatGPT is an advanced language model that has emerged as one of the prominent AI language models which has greatly advanced chatbot technology in recent years. ChatGPT was developed by the OpenAI, and released to internet users on November 30, 2022. Since its release, it has attracted a lot of interest and great deal of attention. This study, analyzes twitter audience tweets about ChatGPT to determine whether the sentiment are positive, or negative. The model consisted of CNN-LSTM model constructed using Keras and was trained on a dataset of labeled tweets consisting of 217,874 rows and two columns and evaluated using metrics such as accuracy, precision, recall and F1-score. The model achieved performance accuracy of 88.45%. These results demonstrated the effectiveness of the CNN-LSTM architecture in capturing both local and global features within the text data, leading to an accurate sentiment prediction.

Keywords: Artificial intelligence, ChatGPT; Sentiment analysis; Twitter; CNN-LSTM; ChatBot, OpenAI Gated Recurrent Unit (GRU).

Journal Reference Format:
Agwi, C.U. & Akpojaro, J. (2025): Analyzing User Sentiments About Chatgpt Using Hybrid Deep Learning Model. Journal of Behavioural Informatics, Digital Humanities and Development Rese Vol. 11 No. 1. Pp 1-24. https://www.isteams.net/behavioralinformaticsjournal
dx.doi.org/10.22624/AIMS/BHI/V11N1P2

1. INTRODUCTION

The rapid advancement of artificial intelligence and natural language processing have led to the development of conversational AI models like ChatGPT, which has revolutionized the way humans interact with machines (Koonchanok *et al*., 2024). However, understanding the sentiment and emotions behind user-generated content on social media platforms like Twitter is crucial for businesses, researchers, and developers to gauge public perception and opinions about their products, services, or technologies (Rafique Yasir *et al*., 2024). In recent years, sentiment analysis has become increasingly important in the field of natural language processing, Various techniques and models are also being explored to improve its accuracy and effectiveness in accurately capturing the nuances of human emotion and opinion (Sabir et al., 2024). One such approach is the use of hybrid deep learning models that combine the strengths of traditional machine learning algorithms with the power of deep learning techniques, such as combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to analyze and extract meaningful patterns and features from large volumes of text data (Rafique *et al.*, 2024). A hybrid approach to sentiment analysis involves combining multiple machine learning and deep learning techniques to improve the accuracy and reliability of sentiment analysis models. This can involve combining different architectures, such as using a CNN to extract local features and an RNN to capture sequential dependencies, which can then be combined using techniques such as feature fusion or model stacking to produce a more accurate and robust sentiment analysis model that can effectively handle the nuances and complexities of natural language (Gogineni *et al.*, 2023).

By integrating multiple techniques, researchers and developers can create more sophisticated and effective sentiment analysis models that can better understand and capture the subtleties of user sentiment in tweets, such as detecting sarcasm, idioms, and figurative language, which are often challenging to detect using traditional machine learning approaches (Salur & Aydin, 2020). The hybrid deep learning model proposed in this study combines the strengths of both CNN and LSTM architectures to develop a more comprehensive and accurate sentiment analysis framework that can effectively handle the variability and diversity of user opinions expressed in tweets about ChatGPT. This study aims to analyze the opinion of tweeter audience about ChatGPT.

2. LITERATURE REVIEW

Sentiment analysis is focused on automating mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It is also referred to as subjectivity analysis, opinion mining, and appraisal extraction (Kharde & Sonawane, 2016).

Sentiment analysis tells product owners and users whether the information about the product is satisfactory or not before they proceed with it. Marketers and firms use results from analysis of sentiment data to understand how users feel about their products or services in such a way that it can be offered to meet user’s requirements.

ChatGPT, since its release has gained widespread attention, reaching an estimate of over 100 million monthly active users in January 2023, after just two months of its release (Koonchanok *et al*., 2024). ChatGPT has wide range of applications ranging from academic writing, language translations, to coding which has in no doubt captured people’s imagination about a revolutionized future of academic research writing (Mirowski *et al*., 2022; Lee *et al*., 2022) and programming (Chen *et al*. 2021, Li *et al*.,2022).

However, despite the several good features possessed by ChatGPT, it is not without flaws as noted by various authors in their articles (Borji, 2023). It was noted that ChatGPT has the propensity to generate falsehood or 'hallucinate' with high confidence (Lin *et al*. 2022; Zhou *et al*., 2023) such as making up fictitious citations (Nature Machine Intelligence, 2023). It is also noted that ChatGPT is weak at handling mathematics and logic problems (Han *et al*., 2022), and from a societal perspective, the bias and discrimination inherent with LLMs from their training datasets as noted by (Jakesch *et al*., 2023; Lin *et al*. 2021; Weidinger *et al.*, 2021; Zhou *et al.*, 2023; Abid *et al*., 2021) have become alarmingly disturbing. These has contributed to making ChatGPT a subject of public discourse across various social and academic communities.

Considering this massive data that may have been generated on twitter owing to the release of ChatGPT, classic approach may not be feasible to handling such analysis as Twitter provides a unique challenge for sentiment analysis due to the informal, colloquial, and often ambiguous language that are used. Tweets can include abbreviations, slang, emoticons, hashtags, and URLs, all of which contribute to the complexity of sentiment detection (Koonchanok *et al*., 2024).

Several techniques are being developed and adopted to analyze sentiments from social media contents due to the vast volume of data and the need for prediction accuracy, which typically complicates the classic procedure for sentiment analysis (Sabir *et al*., 2024). Gogineni *et al*. (2023) noted that several studies have employed deep learning techniques for sentiment analysis, with some focusing on the use of specific architectures, such as Convolutional Neural Networks (CNNs), Last short Term Memory (LSTM) networks, gated recurrent unit (GRU), and bidirectionally long short term memory (BiLSTM) to analyze sentiment in text data in recent years.

However, these architectures have their limitations, for instance, CNNs often struggle with capturing long-range dependencies in text, while LSTMs can suffer from vanishing gradients, which can impact their performance, especially when dealing with long sequences of text (Abdullah *et al*., 2024). To address these limitations, some studies have proposed a hybrid deep learning model that leverages the strengths of various deep learning models to capture both local and global features in text data, and to effectively handle long-range dependencies and vanishing gradients in text data (Gogineni *et al*., 2023).

Ali *et al*. (2022) proposed a deep learning model to analyze sentiment from campaign data from social media in order to predict possible winner in a general election. He was of the thought that since social media are now been used extensively for election campaign, voters can express their personal interests about their preferred party or candidate. Therefore, using a machine learning technique, they provided a five-step process to analyze whether the overall election results were fair or unfair.

Salur & Aydin (2020) proposed a novel hybrid deep learning model that tactically combines different word embedding (Word2Vec, FastText, character-level embedding) and a deep learning algorithm. The proposed model used deep learning methods to extracts features of different word embedding, then combines these features to classify texts based on the sentiment conveyed. (Gogineni *et al.*, 2023) proposed a hybrid deep learning framework for efficient sentiment analysis.

They examined how deep learning models like LSTM, GRU, CNN, can be combined with BOW and TF-IDF integration to capture complex sentiment patterns in text data to improve the predictive power of the model. This study proposed a hybrid CNN-LSTM model that consists of two main components: a CNN-based feature extractor and an LSTM-based sequence encoder. The CNN-based feature extractor uses multiple convolutional layers with different filter sizes to extract local features from the text data, such as sentiment-bearing phrases and contexts, and are then passed through a pooling layer to reduce spatial dimensions and extract the most relevant features that are used to train the LSTM-based sequence encoder.

The LSTM-based sequence encoder uses multiple LSTM layers to learn long-range dependencies and contextual information in the text data, and the output from the LSTM-based sequence encoder is then passed through a fully connected neural network to make the final sentiment classification. The fully connected network consists of multiple layers with ReLU activation, dropout to prevent overfitting, and a softmax output layer for predicting the sentiment of the text as positive, negative, or neutral. This hybrid approach will enable the model to effectively capture both local and global features of the text data, resulting in improved accuracy and robustness in sentiment classification tasks. Furthermore, the use of a hybrid approach allows the model to adapt to varying lengths of text data, making it a versatile and effective tool for analyzing user sentiment in tweets about ChatGPT.

LSTM

Long Short-Term Memory (LSTM) is a type of sophisticated recurrent neural network (RNN) crafted for proficient sequence modeling and prediction, mitigating the challenges of vanishing gradients in extended sequences (Divate, 2021). LSTMs are able to process and analyze sequential data, such as time series, text, and speech. They use a memory cell and gates to control the flow of information, allowing them to selectively retain or discard information as needed and thus avoid the vanishing gradient problem that plagues traditional RNNs. LSTMs are widely used in various applications such as natural language processing, speech recognition, and time series forecasting. The structure of an LSTM network consists of a series of LSTM cells, each of which has a set of gates (input, output, and forget gates) that control the flow of information into and out of the cell. The gates are used to selectively forget or retain information from the previous time steps, allowing the LSTM to maintain long-term dependencies in the input data.

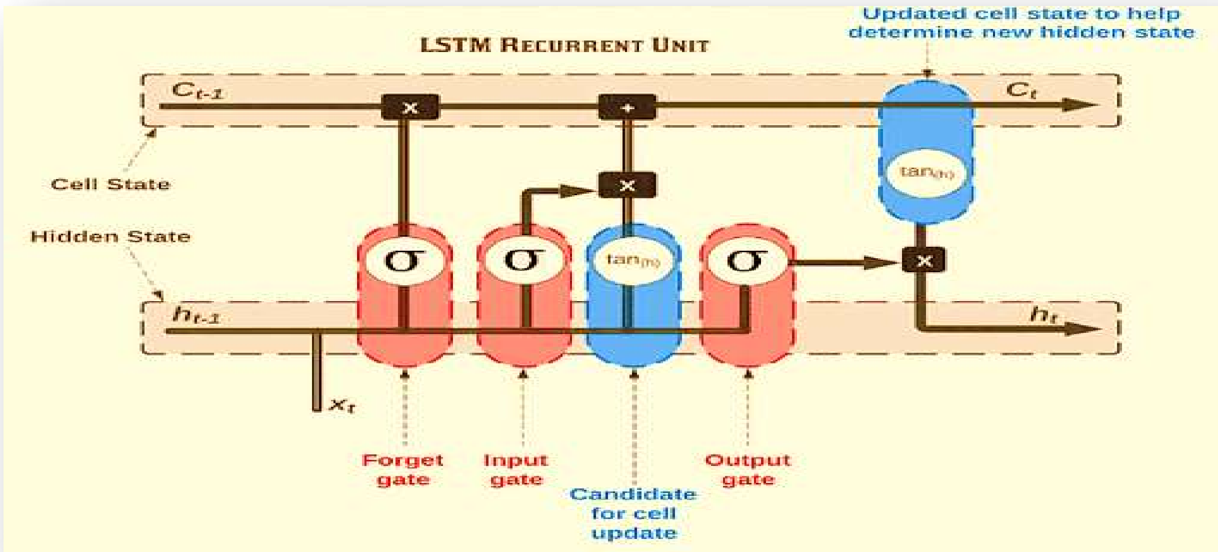

Figure 1. The structure of LSTM model

The LSTM cell also has a memory cell that stores information from previous time steps and uses it to influence the output of the cell at the current time step. The output of each LSTM cell is passed to the next cell in the network, allowing the LSTM to process and analyze sequential data over multiple time steps.

CNN

Convolutional neural networks (CNNs), are a kind of deep learning neural network designed and frequently used for sentiment analysis, picture categorization, and other related applications. CNNs are known to be composed of several layers, including the input, convolutional, pooling, and dense layers. While the pooling layer reduces the dimensions to lessen computational overhead, the convolutional layer uses filters to extract features from the input. Ultimately, the fully connected layer is responsible for generating the final prediction (Gogineni *et al.*, 2023)

CNN-LSTM

The CNN-LSTM is the hybrid deep learning (HDL) model is adopted in this study. The model combines two Deep Learning algorithms, Convolutional Neural Network (CNN) and Last Short-Term Memory (LSTM) for the sentiment analysis task to classify public sentiment in tweets about ChatGPT. The structure of the two HDL (architecture of CNN-LSTM) models used in this study is depicted in Figure 2. The model was developed with CNN layers at the front end so that it could extract the features from the input dataset. The outputs of the CNN layers was passed to the LSTM layers and a dense layer at the output to support sequence prediction.

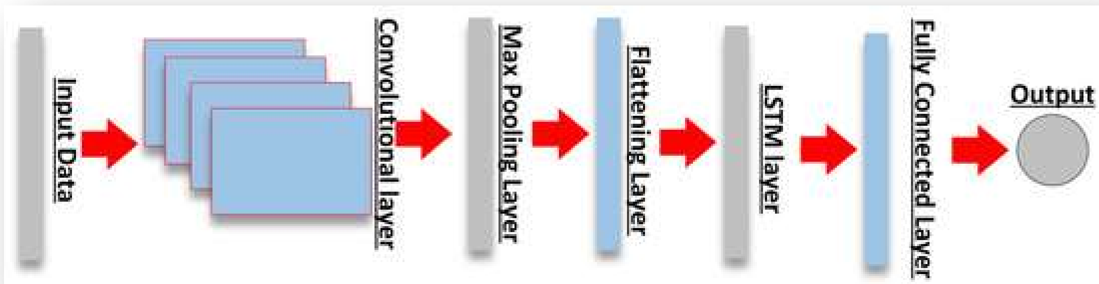

Figure 2. The Structure of CNN-LSTM model.

3. METHODOLOGY

3.1 Data Acquisition and Description

The dataset of public opinions on ChatGPT used for implementation in this study was obtained from Kaggle, which is a public repository and a subsidiary of Google, and an online community of data scientists and machine learning engineers. Kaggle allows users to publish datasets and find datasets that can be used in building AI models, and collaborate with other data scientists and machine learning engineers to solve data science challenges. The dataset consiste of 217,874 rows and two columns. The data consist of tweets and their corresponding labels. The code loads a dataset named "GPTDataset.csv" containing tweets and their corresponding sentiment labels.

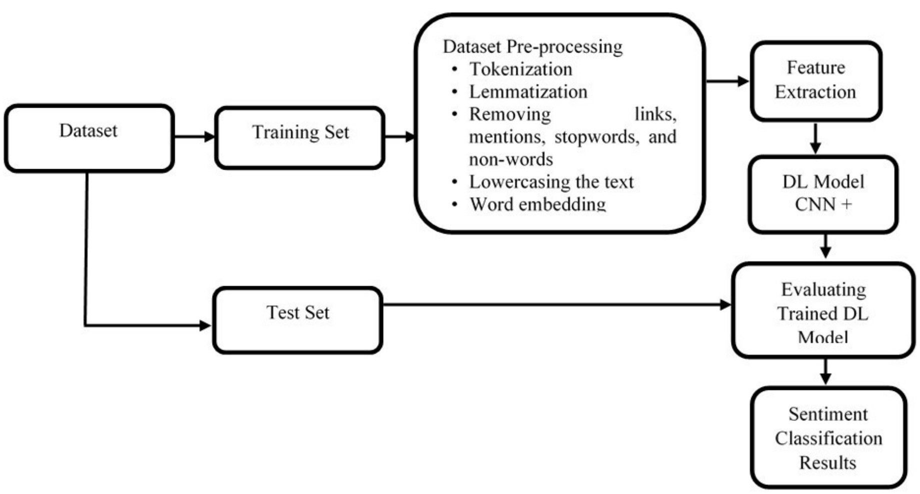

Fig 3. General Methodology of Sentiment Analysis.

Deep learning techniques follow a typical sequence of steps referred to as the pipeline. Figure 3 depicts a framework that shows the methodology adopted in this study. The process began with importing necessary libraries and loading the dataset into the implementation environment. The framework consists of modules which include pre-processing, features extraction, Model development and model evaluation.

3.2 Preprocessing

The tweets were cleaned by removing URLs, mentions, punctuations, numbers, and stop words. Lemmatization was applied to reduce words to their base forms; thereafter, texts were converted to lowercase. This dataset included tweets labeled with their corresponding sentiment (Good, Bad, or Neutral).

```
# text length
df['text_length'] = df['tweet'].astype(str).apply(len) # Convert the 'tweet' column to string type
df[['label','text_length','tweet']].head()
```

	label	text_length	Tweet
0	labels	6	Tweets
1	neutral	80	ChatGPT: Optimizing Language Models for Dialog...
2	good	139	Try talking with ChatGPT, our new AI system wh...
3	neutral	264	ChatGPT: Optimizing Language Models for Dialog...
4	good	188	THRILLED to share that ChatGPT, our new model ...

The following libraries were imported to enable download of stopwords.

```
import re
import pandas as pd
import nltk # Import nltk
from nltk. corpus import stopwords
from wordcloud import WordCloud
import matplotlib. pyplot as plt
```

Thereafter, stopwords were downloaded using: nltk. Download('stopwords') from the dataset from NLTK's data repository. This will open a pop-up window where you can select the "stopwords" package and download it. If you don't see a pop-up, you can manually specify the download directory using nltk.download('stopwords', download_dir='/path/to/your/download/directory').

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 5, March 2025 Series

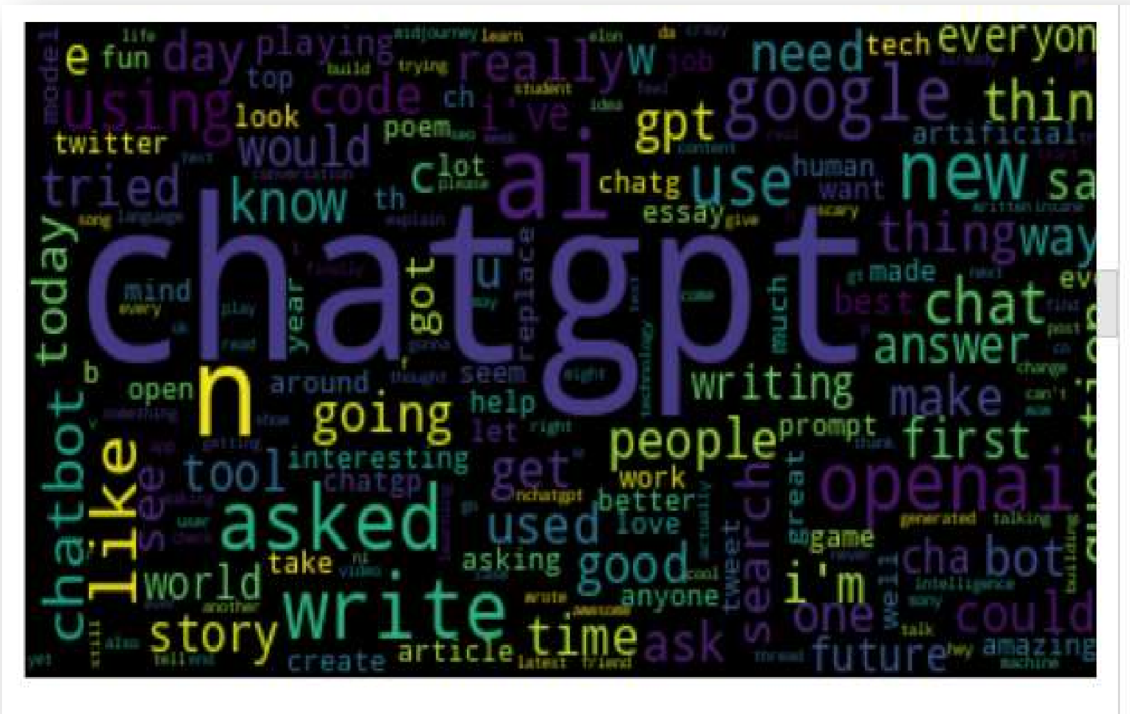

Fig 4: Stopwords Obtained from the Dataset using NLTK's framework

3.3 Feature Engineering

The preprocessed tweets are tokenized, which means breaking them down into individual words or sub-words. A tokenizer is used to create a vocabulary of unique tokens. Each tweet is then converted into a sequence of numerical tokens representing the words in the tweet. At this stage, a tokenizer is used to create a vocabulary of unique tokens from all unique words in the processed tweets. Then a word-to-index mapping was generated to represent words numerically. Thereafter, tweets were converted into sequences of numerical indices using the word-to-index mapping. Since tweets have varying lengths, padding is applied to make them uniform in length. This involves adding special padding tokens to shorter tweets to make them equal in length to the longest tweet in the dataset, say a fixed length (max_length = 100).

3.4 Model Building, Training and Evaluation

In this stage, a CNN-LSTM model was constructed using Keras. An embedding layer maps words to dense vectors. A convolutional layer (Conv1D) extracts local features from the word embeddings. A max-pooling layer (MaxPooling1D) reduces the dimensionality of the convolutional output. An LSTM layer captures long-term dependencies in the sequence. A dense layer with softmax activation predicts the sentiment (3 classes: Bad, Good, Neutral).

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 5, March 2025 Series

In the Training and Evaluation stage, the model was trained using the categorical crossentropy loss function and the adam optimizer. Data was split into training and testing sets (80% and 20%, respectively). The model was trained for 7 epochs with a batch size of 64. The model was **evaluated using accuracy, precision, recall and F1-score as the evaluations parameters.** A confusion matrix (Figure 5) was generated to visualize the model's performance.

4. DISCUSSION AND ANALYSIS OF RESULTS

The model processed 1371 batches of data during validation reported as 1371/1371. The fact that both numbers are the same means that the entire validation dataset was seen. The total time taken for the validation process was 6 seconds, and an average of 5 milliseconds was used to process each batch of data. The model performance parameters recorded as follows: accuracy: 88.45%, precision: 87.6%, recall: 87.4%, and F1-score: 87.3%. In the light of this, it can be conluded that the model's performance shows that it has good ability to predict sentiment.

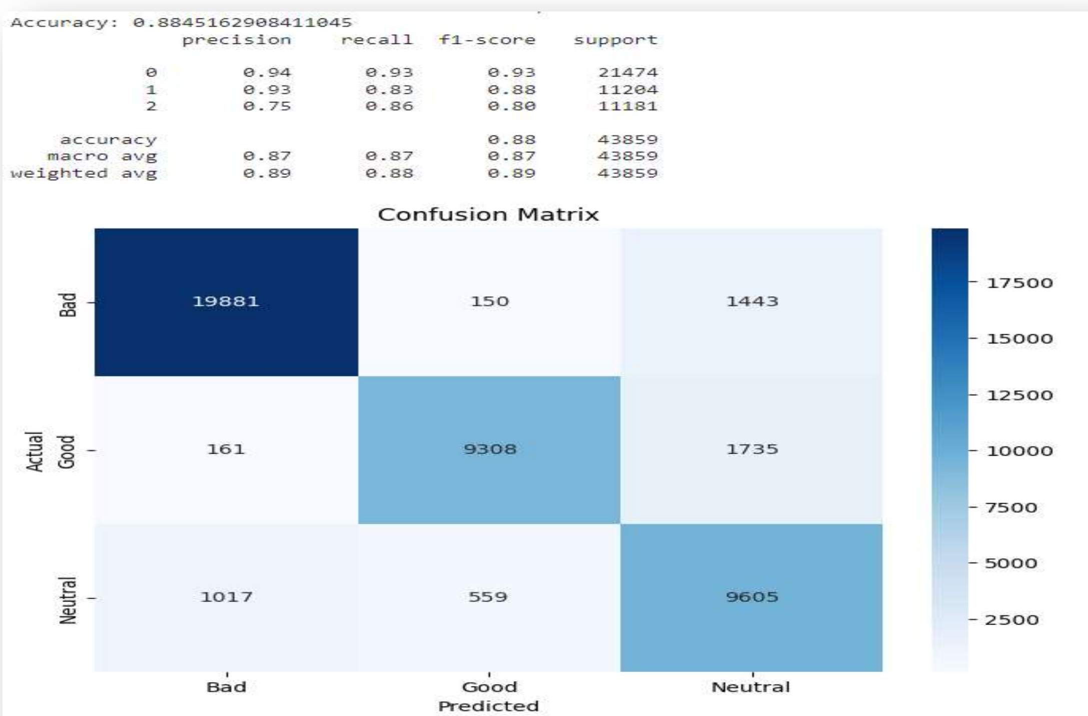

Figure 5: Confusion Matrix for Sentiment about ChatGPT Analysis Using CNN-LSTM Model

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 5, March 2025 Series

The confusion matrix (Figure 5) provides a detailed breakdown of the model's predictions, showing which classes are often confused with each others. Thus, this helps to identify areas where the model might be making mistakes. The combination of CNN and LSTM layers leverages the strengths of both architectures. CNNs excel at capturing local patterns (like phrases or n-grams), while LSTMs are effective in learning long-range dependencies in the text. The code uses default values for many hyperparameters (e.g., number of filters, kernel size, LSTM units). Further experimentation with these hyperparameters might improve performance.

5. CONCLUSION

This experiment explored the application of CNN-LSTM model for predicting the sentiment of tweets about ChatGPT. The model was trained on a dataset of labeled tweets and evaluated using metrics such as accuracy, precision, recall and F1-score. The model performance was visualized using a confusion matrix. The confusion matrix provided valuable insights into the model's performance, revealing its strengths and weaknesses in classifying different sentiment categories. This information can guide further improvements to the model's architecture or data preprocessing techniques. The CNN-LSTM model demonstrated a promising result for sentiment prediction, achieving an accuracy of 88.45% on the given dataset. The results demonstrate the effectiveness of the CNN-LSTM architecture in capturing both local and global features within the text data, leading to accurate sentiment predictions.

REFERENCES

Abdullah, S., Van Cauwenberge, P., Vander Bauwhede, H., & O'Connor, P. (2024). Review Ratings, Sentiment in Review Comments, and Restaurant Profitability: Firm-Level Evidence. *Cornell Hospitality Quarterly,* 65(3), 378-392.

Abid, A., Farooqi, M., & Zou, J. (2021). Large language models associate muslims with violence. *Nature Machine Intelligence*, 3(6), 461–463.

Ali, A., Khan, M., Khan, K., Khan, R. U., & Aloraini, A. (2024). Sentiment Analysis of Low-Resource Language Literature Using Data Processing and Deep Learning. *Computers, Materials & Continua*, 79(1), 713-733.

Borji, A. (2023). A categorical archive of ChatGPT failures. arXiv preprint arXiv:2302.03494. available at https://arxiv.org/abs/2302.03494

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H.P.d.O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., & Brockman, G. (2021) Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374

Divate, M. S. (2021). Sentiment analysis of Marathi news using LSTM. *International journal of Information Technology*, 13(5), 2069-2074.

Gogineni, A. K., Sai-Reddy, S. K., Kakarala, H., Gavini, Y. C., Venkat, M. P., Hajarathaiah, K., & Enduri, M. K. (2023). A Hybrid Deep Learning Framework for Efficient Sentiment Analysis. *International Journal of Advanced Computer Science & Applications*, 14(12). 1032-1038.

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., Naaman, M. (2023) Co-writing with opinionated language models affects users' views. arXiv preprint arXiv:2302.00560

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971.

Koonchanok, R., Pan, Y., & Jang, H. (2024). Public attitudes toward ChatGPT on Twitter: sentiments, topics, and occupations. *Social Network Analysis and Mining*, 14(1), 106.

Lee, M., Liang, P., Yang, Q. (2022). Coauthor: Designing a human-AI collaborative writing dataset for exploring language model capabilities. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–19 . https://doi.org/10.1145/3491102.3502030

Li, Y., Choi, D., Chung, J., Kushman, N., Schrittwieser, J., Leblond, R., Eccles, T., Keeling, J., Gimeno, F., & Dal-Lago, A. (2022). Competition-level code generation with alphacode. Science, 378(6624), 1092–1097 https://doi.org/10.1126/science. abq1158

Lin, S., Hilton, J., & Evans, O. (2022). Teaching models to express their uncertainty in words. arXiv preprint arXiv:2205.14334 https://doi.org/10.48550/ARXIV. 2205.14334

Mirowski, P., Mathewson, K.W., Pittman, J., & Evans, R. (2022) Co-writing screenplays and theatre scripts with language models: An evaluation by industry professionals. arXiv preprint arXiv:2209.14958

Nature Machine Intelligence (2023). The AI writing on the wall. *Nature Machine Intelligence,* 5(1), 1-1 (Editorial), https://doi.org/10.1038/s42256-023-00613-9

Rafique, Y., Wu, J., Muzaffar, A. W., & Rafique, B. (2024). An enhanced integrated fuzzy logic-based deep learning techniques (EIFL-DL) for the recommendation system on industrial applications. *PeerJ Computer Science*, 10, e2529.

Sabir, A., Ali, H. A., & Aljabery, M. A. (2024). Chatgpt tweets sentiment analysis using machine learning and data classification. *Informatica*, 48(7). 103-112.
DOI: 10.31449/inf.v48i7.5535

Salur, M. U., & Aydin, I. (2020). A novel hybrid deep learning model for sentiment classification. *IEEE Access*, 8, 58080-58093.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., & Kasirzadeh, A. (Webb 2021) Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359 https://doi.org/10.48550/ ARXIV.2112.04359

Zhou, K., Jurafsky, D., & Hashimoto, T. (2023). Navigating the grey area: Expressions of overconfidence and uncertainty in language models. arXiv preprint arXiv:2302.13439 https://doi.org/10.48550/ARXIV.2302.13439

Zhuo, T.Y., Huang, Y., Chen, C., & Xing, Z. (2023) Exploring ai ethics of chatgpt: A diagnostic analysis. arXiv preprint arXiv:2301.12867 https://doi.org/10.48550/ARXIV. 2301.12867