International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

CHATBOT EMOTION RECOGNITION AND MUSIC RECOMMENDATION

[1]MANTHENA SAHITYA, [2]Y SRINIVAS RAJU

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

In recent years, emotion-aware systems have gained significant importance in improving human-computer interaction. This project proposes a Chatbot Emotion Recognition and Music Recommendation System that detects user emotions through multiple input modes such as text, voice, and facial expressions using deep learning techniques. The system utilizes advanced models including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and hybrid LSTM+GRU models combined with BERT embeddings for accurate emotion classification. The models are trained on a publicly available dataset from Kaggle Emotion Dataset, which contains labeled text data for different emotional states. The system evaluates model performance using metrics such as accuracy, precision, recall, and F1-score. Experimental results show that the BiLSTM model achieves the highest accuracy of 91.84%, outperforming other models. The system is implemented using Jupyter Notebook for model training and a web-based interface for real-time prediction. Based on the detected emotion, the chatbot recommends appropriate music to enhance user mood. The system supports multimodal interaction, allowing users to input text, voice, or facial expressions. This approach improves user experience and provides personalized recommendations. The proposed system demonstrates the effectiveness of deep learning in emotion recognition and its application in intelligent recommendation systems.

Keywords: Emotion Recognition, Chatbot, Deep Learning, CNN, LSTM, BiLSTM, BERT, Music Recommendation, NLP

I.INTRODUCTION

Emotion recognition plays a vital role in enhancing human-computer interaction by enabling systems to understand user feelings and respond accordingly. Traditional chatbots are limited to basic text-based responses and lack the ability to interpret user emotions. This

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

limitation reduces their effectiveness in providing personalized and empathetic responses. With advancements in artificial intelligence and deep learning, it is now possible to develop systems that can detect emotions from multiple sources such as text, voice, and facial expressions. These systems can be used in various applications, including virtual assistants, mental health support, and entertainment platforms.

Deep learning models such as CNN, LSTM, and BiLSTM have shown significant success in analyzing sequential and visual data. LSTM models are particularly effective in processing text data due to their ability to capture long-term dependencies, while CNN models are widely used for image-based emotion detection. Additionally, BERT embeddings enhance text representation by capturing contextual meaning, improving classification accuracy. By combining these techniques, the system can accurately detect emotions from different input modalities.

This project focuses on developing a multimodal chatbot system that detects emotions and recommends music accordingly. The system processes user input through text, voice, or webcam and applies trained deep learning models to classify emotions such as happy, sad, angry, and neutral. Based on the detected emotion, the system suggests relevant songs to improve user experience. This integrated approach enhances personalization and demonstrates the practical application of AI in real-world interactive systems.

II SURVEY OF RESEARCH

[1] The research by Jacob Devlin et al. (2019) introduced Bidirectional Encoder Representations from Transformers (BERT) for natural language processing tasks. The methodology uses bidirectional training to understand the context of words in a sentence more effectively. BERT generates contextual embeddings that significantly improve text classification and sentiment analysis tasks. The results showed state-of-the-art performance in various NLP benchmarks. However, BERT requires high computational resources and large datasets for training. This research is highly relevant to emotion recognition systems, as it enhances text representation and improves the accuracy of emotion classification models in chatbot applications.

[2] The study by Sepp Hochreiter and Jürgen Schmidhuber (1997) introduced Long Short-Term Memory (LSTM) networks for sequence

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

learning tasks. The methodology uses memory cells and gating mechanisms to capture long-term dependencies in sequential data such as text and speech. The results demonstrated improved performance over traditional recurrent neural networks in handling time-series and language data. However, LSTM models can be computationally intensive and may suffer from slow training. This research supports the use of LSTM for emotion detection from text and voice inputs in chatbot systems.

[3] The research by Yann LeCun et al. (1998) introduced Convolutional Neural Networks (CNNs) for pattern recognition and image analysis. The methodology uses convolutional layers to extract spatial features from images, making it effective for facial emotion recognition. The results showed high accuracy in image classification tasks. However, CNNs require large labeled datasets for optimal performance. This research is relevant for detecting emotions from facial expressions using webcam input in multimodal chatbot systems.

[4] The study by Kyunghyun Cho et al. (2014) introduced Gated Recurrent Units (GRU) as an alternative to LSTM. The methodology simplifies the gating mechanism while maintaining the ability to capture dependencies in sequential data. The results demonstrated comparable performance to LSTM with reduced computational complexity. However, GRU may not always outperform LSTM in complex tasks. This research supports the use of hybrid LSTM+GRU models for efficient emotion recognition.

[5] The research by Paul Ekman (1992) focused on the classification of basic human emotions such as happiness, sadness, anger, fear, surprise, and disgust. The methodology identifies universal facial expressions associated with emotions. The results demonstrated that emotions can be reliably detected across cultures. However, subtle emotional variations can be difficult to classify. This research provides a theoretical foundation for emotion recognition systems.

[6] The study by Rosalind Picard (1997) introduced the concept of affective computing, which enables machines to recognize and respond to human emotions. The methodology integrates AI, signal processing, and psychology to detect emotional states. The results showed improved human-computer interaction through emotion-aware systems.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

However, challenges such as data variability and accuracy remain. This research supports the development of intelligent chatbot systems capable of emotion recognition and personalized responses.

III. WORKING METHODOLOGY

The proposed Chatbot Emotion Recognition and Music Recommendation System follows a structured pipeline consisting of data collection, preprocessing, model training, and real-time prediction. Initially, the system uses an emotion dataset obtained from Kaggle, which contains labeled text data corresponding to different emotional states such as happy, sad, angry, and neutral. The dataset is preprocessed by cleaning the text, removing stop words, and normalizing the input. BERT embeddings are then applied to convert textual data into numerical vectors that capture contextual meaning. The processed dataset is divided into training (80%) and testing (20%) sets to evaluate model performance effectively. This stage ensures that the data is prepared and suitable for training deep learning models.

In the next phase, multiple deep learning algorithms such as CNN, LSTM, BiLSTM, and LSTM+GRU are trained using the processed dataset. Each model learns patterns associated with different emotions and is evaluated based on performance metrics such as accuracy, precision, recall, and F1-score. CNN is used for extracting features from text representations, while LSTM and BiLSTM are used to capture sequential dependencies. The hybrid LSTM+GRU model combines the strengths of both architectures to improve efficiency. Among all models, BiLSTM achieves the highest accuracy due to its ability to process input data in both forward and backward directions. Confusion matrices are generated to visualize prediction performance and identify misclassifications.

In the final stage, the trained model is integrated into a web-based application that supports multimodal interaction. Users can provide input through text, voice, or webcam. For text input, the system directly processes the input using the trained model. For voice input, speech-to-text conversion is performed before emotion detection. For webcam input, facial features are extracted using CNN-based models. The system then predicts the user's emotion and recommends appropriate songs based on the detected emotional state. This real-time system enhances user engagement and provides a personalized experience.overall security in decentralized cloud systems.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

IV RESULTS EXPLANATIONS

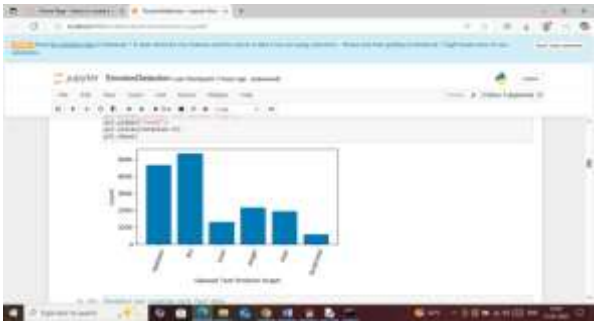
In propose work we have utilized different deep learning algorithms such as CNN, LSTM, BILSTM and LSTM + GRU along with BERT embedding to detect emotion using various Chatbot techniques such as Text, Voice or Webcam Face video.

To train above algorithms we have used emotion dataset available on KAGGLE which can be download from below URL

https://www.kaggle.com/datasets/parulpandey/emotion-dataset?select=training.csv

Above algorithms get trained on Emotion dataset and then each algorithm performance is evaluated in terms of accuracy, precision, recall and FSCORE.

For training and algorithm testing we have used JUPYTER notebook and for prediction we have utilized WEB framework.



In above screen visualizing graph of different class labels found in dataset where x-axis represents emotion and y-axis represents number of text found under that emotion

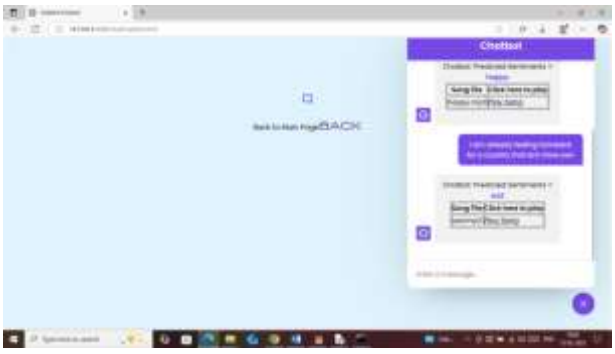
In above screen CNN got 91% accuracy and can see other metrics also and then in confusion matrix graph x-axis represents True Labels and y-axis represents predicted labels and then all different colour boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count which are very few

In above screen visualizing all algorithms performance where x-axis represents algorithm

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

names and y-axis represents accuracy and other metrics in different colour bars

In above screen giving another sentence and below is the Chatbot response



In above screen predicted emotion is 'SAD' and similarly give any text and get emotion and recommended song list. Now click on 'Back' link to get back to main page

In above screen in webcam you can show face and then click on 'Take Snapshot' button to capture face and then click on 'Detect Emotion' button to get list of songs

V. CONCLUSION

The proposed Chatbot Emotion Recognition and Music Recommendation System demonstrates the effective use of deep learning techniques for understanding human emotions and providing personalized responses. By integrating models such as CNN, LSTM, BiLSTM, and LSTM+GRU with BERT embeddings, the system achieves high accuracy in emotion detection, with BiLSTM performing the best. The multimodal approach, which supports text, voice, and facial input, enhances user interaction and makes the system more flexible and user-friendly. The experimental results show reliable performance in emotion classification and successful music recommendation based on detected emotions. Although challenges such as environmental noise, facial recognition limitations, and data variability exist, the system provides a scalable and efficient solution for real-world applications. Overall, this project highlights the potential of AI-driven emotion-aware systems

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

in improving user experience and enabling intelligent recommendation systems.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL*, 2019, pp. 4171–4186.

[2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[4] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. EMNLP*, 2014.

[5] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.

[6] R. Picard, *Affective Computing*, MIT Press, 1997.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[8] A. Vaswani et al., "Attention is all you need," in *Proc. NIPS*, 2017.

[9] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep CNNs," in *Proc. NIPS*, 2012.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[11] T. Brown et al., "Language models are few-shot learners," in *Proc. NeurIPS*, 2020.