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

Smart Chatbot for Emotion Detection and Adaptive Music Recommendation

Potlapally Ushaswi[1], Pandi Parimala[2], Pally Siddhartha[3], K Challuri Sai Krishna[4], Dr. K. Naveen Kumar[5]

[1,2,3,4]Students, Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

[5]Associate Professor, Department of Computer Science Engineering (Artificial Intelligence and Machine Learning),

Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, Telangana, India

***Abstract*— In recent years, the development of emotionally intelligent systems has gained significant attention due to their ability to enhance human-computer interaction. This project presents a multimodal chatbot emotion recognition and music recommendation system that integrates advanced Deep Learning techniques to understand and respond to user emotions effectively. The system utilizes models such as Convolutional Neural Networks , Long Short-Term Memory , Bidirectional LSTM and a hybrid LSTM-GRU architecture, combined with BERT embeddings for accurate textual feature extraction. The proposed system supports three modes of user interaction: text input, voice input, and webcam-based facial expression analysis, enabling comprehensive emotion detection. The models are trained and evaluated using a Kaggle dataset in a Jupyter Notebook environment, with performance assessed through metrics such as accuracy, precision, recall, and F1-score. Among the implemented models, BiLSTM demonstrates superior performance in emotion classification. Based on the detected emotion, the system recommends appropriate music that aligns with the user's mood through an interactive Flask-based web application. The chatbot enables real-time communication and provides personalized responses, allowing users to express their emotions naturally. This approach not only improves user engagement but also highlights the potential of integrating emotion recognition with recommendation systems to create more adaptive and human-centric AI applications.**

***Keywords*— Emotion Recognition, Chatbot, Deep Learning, BERT Embedding, Music Recommendation, CNN, LSTM, BiLSTM, Flask Web Interface, Real-time Emotion Detection.**

I. INTRODUCTION

In today's digital landscape, fostering user engagement is crucial, particularly for music streaming services. This paper introduces a Chatbot Music Recommendation System that leverages user emotions, time of day, and weather conditions to deliver context-aware music suggestions, thereby enhancing overall user satisfaction. The primary objective of this project is to develop an intelligent chatbot that provides personalized song recommendations based on user interactions, emotional context, and real-time environmental factors. By integrating advanced Natural Language Processing (NLP) techniques and real-time data from external APIs, the chatbot analyses user sentiment during conversations to generate relevant music suggestions, improving engagement and user experience [1], [2].

In an era where music serves as a universal language, personalized music experiences have become increasingly important. Traditional recommendation systems primarily rely on content-based and collaborative filtering techniques, which often fail to consider dynamic human emotions and contextual factors such as time and weather [3], [4]. This limitation creates a gap where recommendations may not align with users' real-time preferences, leading to reduced satisfaction and engagement.

To address these challenges, the proposed system incorporates emotional intelligence and contextual awareness into its recommendation engine. By utilizing machine learning and deep learning approaches, including emotion detection through text and facial expressions, the system can better understand user moods and adapt recommendations accordingly [5], [10], [11]. Additionally, chatbot-based interaction provides a more engaging and conversational interface compared to traditional systems, enabling users to explore music in a more natural and intuitive manner.

Recent research highlights the effectiveness of emotion-based music recommendation systems in enhancing user experience and personalization [6], [8], [14]. The integration of multimodal emotion recognition and contextual data allows the system to dynamically respond to user needs, offering music that resonates with their emotional state and environment. This approach not only improves recommendation accuracy but also transforms music discovery into an interactive and enjoyable process.

In conclusion, the proposed Chatbot Music Recommendation System aims to bridge the gap between traditional recommendation techniques and modern user expectations by combining emotional analysis, contextual awareness, and conversational AI. This innovative system enhances user engagement and contributes to the advancement of intelligent, human-centric music recommendation technologies. This approach not only improves user engagement but also highlights the potential of integrating emotion recognition with recommendation systems to create more adaptive and human-centric AI applications.

II. RELATED WORK

The work by N. Mathew et al. presents a chatbot-based music recommendation system that utilizes user emotions to deliver personalized song suggestions. The system integrates Natural Language Processing techniques to analyse user input and identify emotional states such as happiness, sadness, or anger. Based on the detected emotion, the chatbot recommends music that aligns with the user's mood, thereby enhancing user engagement and satisfaction. The study emphasizes the importance of conversational interfaces in improving user interaction compared to traditional recommendation systems. Additionally, the system demonstrates how emotion-aware computing can be effectively applied in real-world applications like music streaming. The authors highlight that incorporating emotional intelligence into recommendation systems leads to more relevant and meaningful suggestions. Overall, this research contributes to the development of intelligent, user-centric music recommendation systems by combining chatbot technology with emotion recognition. [1]

H. Zhang explore how emotional preferences influence music recommendation in daily activities. Their study focuses on integrating emotional context into recommendation systems to improve personalization. The authors analyse how users emotions vary across different situations and how these emotional states impact music choices. By incorporating emotional data into the recommendation process, the system can provide more context-aware and relevant suggestions. The research demonstrates that music preferences are not static but change depending on users' moods and activities. The proposed approach enhances traditional recommendation methods by considering psychological and contextual factors. Experimental results show that emotion-based recommendations significantly improve user satisfaction compared to conventional systems. This work highlights the importance of understanding human emotions in designing intelligent systems and provides a foundation for developing more adaptive and personalized music recommendation technologies.[2]

The study by Y. H. Chin et al. introduces an emotion profile-based music recommendation system that categorizes users' emotional states to generate suitable music suggestions. The system builds emotion profiles by analyzing user behavior and preferences, allowing it to recommend songs that match specific emotional conditions. The authors emphasize the role of emotional modeling in enhancing recommendation accuracy. By mapping music features to emotional dimensions, the system can better understand the relationship between user feelings and music choices. The research demonstrates that emotion-aware systems can outperform traditional recommendation techniques by delivering more personalized results. Additionally, the study highlights the importance of integrating user feedback to continuously improve recommendation quality. This approach contributes to the advancement of intelligent music systems by combining emotional analysis with user profiling, ultimately improving user satisfaction and engagement. [3]

K. Yoon et al. propose a music recommendation system that utilizes emotion-triggering low-level audio features to enhance recommendation accuracy. The system analyses musical characteristics such as tempo, pitch, and rhythm to identify the emotional content of songs. By mapping these features to emotional categories, the system can recommend music that aligns with users' moods.

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

The study focuses on bridging the gap between audio signal processing and emotional perception in music. The authors demonstrate that low-level features can effectively capture the emotional essence of songs, enabling more accurate recommendations. This approach differs from traditional methods by focusing on the intrinsic properties of music rather than user behaviour alone. Experimental results show improved performance in emotion-based recommendation tasks. The research highlights the potential of combining signal processing techniques with emotion recognition to create more intelligent and context-aware music recommendation systems. [4]

S. Gilda et al. present a smart music player that integrates facial emotion recognition with music recommendation. The system uses computer vision techniques to detect user emotions through facial expressions captured via a camera. Based on the identified emotion, the system suggests music that matches the user's mood, creating a personalized listening experience. The study demonstrates the effectiveness of combining image processing and machine learning techniques for emotion detection. The authors highlight that facial expressions provide a reliable and real-time method for understanding user emotions. The system enhances user interaction by eliminating the need for manual input, making the recommendation process more seamless. Experimental results indicate that the proposed approach improves user satisfaction and engagement. This research contributes to the development of multimodal emotion-aware systems by integrating facial recognition with music recommendation technologies. [5]

III. DATASET DETAILS

The proposed system utilizes an emotion-labelled dataset sourced from Kaggle, which consists of textual data annotated with corresponding emotion labels. The dataset is structured into two primary columns: the first column contains textual inputs (user sentences), while the second column represents the associated emotion categories such as happy, sad, angry, surprise, and neutral. Initially, all required libraries and packages are loaded in the Jupyter Notebook environment, followed by the initialization of the BERT model for advanced text representation.

To better understand the dataset distribution, a graphical visualization is generated where the x-axis represents different emotion classes and the y-axis indicates the frequency of text samples in each category. This helps in identifying class balance and data distribution. A preprocessing function is then applied to clean the text data by removing noise such as special characters, stopwords, and unnecessary symbols. The cleaned text is transformed into numerical vectors using BERT embeddings, enabling the model to capture contextual meaning effectively. The dataset is then split into training and testing sets, with 80% used for training and 20% for evaluation. Various deep learning models, including CNN, LSTM, BiLSTM, and LSTM+GRU, are trained on this data.

Performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score. Among all models, BiLSTM achieved the highest accuracy of 91.84%, followed by CNN with 91%, LSTM+GRU with 84%, and LSTM with 79%. Confusion matrix visualization further confirms that most predictions fall along the diagonal, indicating correct classifications. This dataset plays a crucial role in enabling accurate emotion detection, which is later used to recommend personalized music through a real-time Flask-based web interface.

IV. PROPOSED METHODOLOGY

The proposed system introduces a multimodal emotion recognition and music recommendation chatbot that leverages advanced deep learning techniques to enhance user interaction and personalization. Unlike traditional systems that rely on single-modal inputs, this approach integrates three input modes: text, voice, and facial expressions, enabling a comprehensive understanding of user emotions. The system begins with data acquisition using an emotion-labelled dataset obtained from Kaggle. For text-based emotion detection, the input is first pre-processed through cleaning techniques such as removal of stop words, punctuation, and irrelevant symbols. The processed text is then passed through a Bidirectional Encoder Representations from Transformers (BERT) model to generate contextual embeddings. These embeddings capture semantic relationships and improve the accuracy of emotion classification.

Multiple deep learning models, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and a hybrid LSTM+GRU model, are trained using the generated embeddings. Each model learns different aspects of the data: CNN captures spatial patterns, LSTM handles sequential dependencies, BiLSTM processes information in both forward and backward directions, and the hybrid model combines the strengths of LSTM and GRU for improved performance. Among these, BiLSTM achieves the highest accuracy, making it the most effective model for emotion classification.

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

For voice-based interaction, user speech is converted into text using speech recognition techniques, after which the same text-based emotion detection pipeline is applied. In the case of facial emotion recognition, images captured through a webcam are processed using computer vision techniques and CNN-based models to identify facial expressions and classify emotions in real time.

Once the emotion is detected through any of the input modalities, the system maps the identified emotion to a corresponding set of songs. The recommendation engine selects music that aligns with the user's emotional state, ensuring a personalized and engaging experience. The entire system is deployed using a Flask-based web application, providing an interactive interface where users can communicate with the chatbot in real time. The system also includes performance evaluation using metrics such as accuracy, precision, recall, and F1-score to ensure reliability. Overall, the proposed method combines multimodal emotion detection, deep learning models, and a real-time recommendation engine to deliver an intelligent, user-friendly, and emotionally aware music recommendation system.

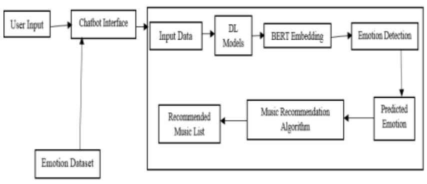


Figure [1] : System Architecture

Figure[1] illustrates User inputs interact with chatbot interface, processed using BERT and deep learning models for emotion detection, enabling personalized music recommendations.

V. RESULT AND DISCUSSION

The implemented AI-based music recommendation system integrates multiple emotion recognition approaches with interactive chatbots, providing a highly personalized user experience. The system first uses text input, applying BERT embeddings to convert sentences into numeric vectors, and trains multiple deep learning models including CNN, LSTM, BiLSTM, and LSTM+GRU. Among these, BiLSTM achieved the highest accuracy of 91.84%, while CNN also performed strongly with 91% accuracy. LSTM and LSTM+GRU recorded 79% and 84% accuracy, respectively, highlighting the effectiveness of bidirectional context learning for emotion detection.

The system allows users to interact via three modes: text, webcam, and voice. In the text-based chatbot, users type messages and receive predicted emotions along with curated song recommendations. The webcam-based interface detects facial expressions in real time, predicting emotions such as happy, sad, or neutral, and suggests relevant songs. The voice chatbot captures live speech through the microphone, analyzes emotional content, and recommends music accordingly. Each interface demonstrates accurate real-time emotion recognition, ensuring that recommendations align with the user's current emotional state.

The integration of emotion-aware chatbots in music streaming enhances user engagement, offering a more interactive and emotionally intelligent experience. By leveraging advanced deep learning models and multimodal inputs, this system provides a robust framework for personalized music discovery, supporting both user satisfaction and retention. Future enhancements could further improve accuracy and expand emotion categories.

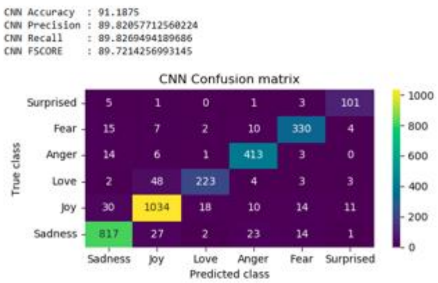


Figure [2] : CNN Performance

Figure [2] The CNN confusion matrix shows high accuracy for Joy, Anger, and Love, with most misclassifications in Sadness and Fear classes.

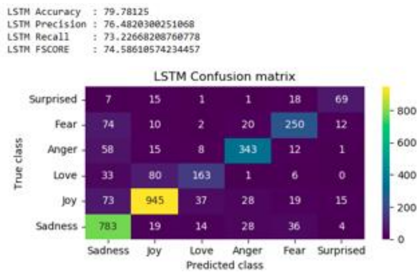
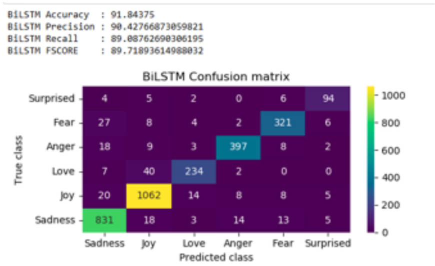


Figure [3] LSTM Performance

Figure [3] illustrates The LSTM model shows moderate accuracy, high misclassification for "Joy" and "Sadness," and strong performance on "Anger" and "Fear" classes.

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



Figure[4] : BILSTM Performance

The Figure[4] The BiLSTM model achieves higher accuracy than LSTM, with strong performance on "Anger" and "Fear," but notable confusion in "Joy" and "Sadness."

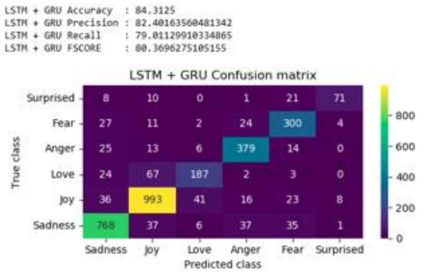


Figure [5] LSTM + GRU

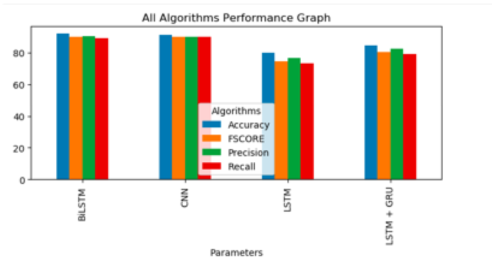
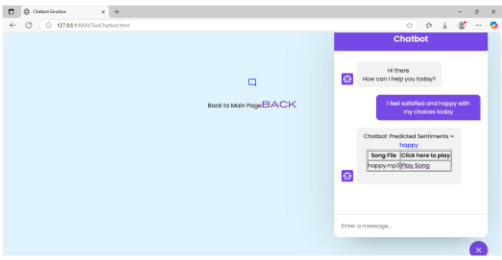
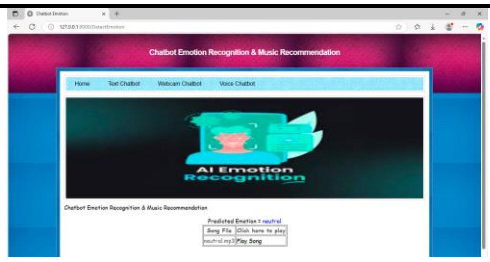


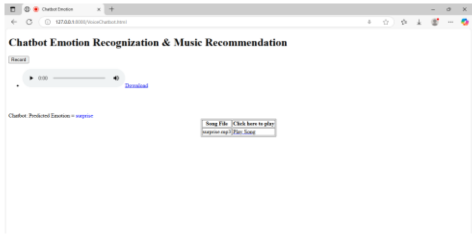
Figure [6] All Algorithm Comparison Graph



Figure[7] : Chatbot Emotion Recognition



Figure[8] : Image Emotion Recognition



Figure[9] : Speech Emotion Recognition

DISCUSSION

The AI-based music recommendation system demonstrates the significant potential of combining emotion recognition with interactive chatbots to enhance user engagement and personalization. By leveraging multiple deep learning models—CNN, LSTM, BiLSTM, and LSTM+GRU—the system effectively interprets emotional cues from text, facial expressions, and voice inputs. Among these models, BiLSTM emerged as the most effective for text-based emotion detection, achieving 91.84% accuracy. Its superior performance highlights the advantage of capturing bidirectional context, which allows the model to understand both preceding and succeeding words in a sentence. CNN also performed strongly, indicating that convolutional approaches can extract salient patterns from textual embeddings efficiently. While LSTM and LSTM+GRU showed lower accuracy, their results suggest that sequential learning remains valuable, particularly when combined with other modalities.

The multimodal integration—text, webcam, and voice—ensures a more holistic understanding of the user's emotional state. The text chatbot allows fast, precise emotion prediction for typed inputs, while the webcam interface provides real-time facial emotion detection, enabling non-verbal cues to influence recommendations. The voice chatbot further adds a natural interaction layer, interpreting tonal variations and speech patterns to determine emotions. This multimodal approach addresses the

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

limitations of single-input systems and ensures that recommendations remain contextually relevant.

From a practical perspective, the system significantly enhances music recommendation by aligning song suggestions with the user's current emotional state. It also promotes engagement by making the interaction more immersive and emotionally intelligent. However, challenges such as subtle emotion differentiation, background noise in voice inputs, and diverse facial expressions across users can affect prediction accuracy. Future work could explore advanced multimodal fusion techniques, additional emotion categories, and personalization strategies based on user preferences over time to further improve system performance.

VI. CONCLUSION

This study presents a personalized news recommendation system that integrates BERT-based semantic embeddings with a Graph Convolutional Neural Network (GCNN). The proposed approach effectively captures both the contextual meaning of news content and the complex relationships between users and articles. By modelling user interactions as a graph, the system is able to learn deeper patterns and provide more accurate recommendations compared to traditional methods such as CNN and LSTM. The experimental results demonstrate that the proposed model achieves improved performance across multiple evaluation metrics, confirming its effectiveness in delivering relevant and personalized news suggestions. The ability to handle sparse data and generalize across different user behaviours further strengthens its applicability in real-world scenarios. In conclusion, the integration of graph-based learning with advanced text representation techniques offers a promising direction for modern recommendation systems. Future work can focus on optimizing computational efficiency, incorporating real-time user feedback, and exploring advanced graph architectures such as attention-based GNNs. This approach has the potential to significantly enhance user experience by delivering more meaningful and timely news content..

REFERENCES

[1]. N. Mathew, N. Chooramun, and M. S. Sharif, "Implementing a Chatbot Music Recommender System Based on User Emotion," in Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), University of Bahrain, Bahrain, Nov. 2023, pp. 195-199. doi: 10.1109/3ICT60104.2023.10391771. "Chatbot Song Recommendation System," International Journal of Engineering Research and Technology (IJERT), vol. 9, no. 5, pp. 1-5, May 2022. [Online].

[2]. H. Zhang, K. Zhang and N. Bryan-Kinns, "Exploiting the emotional preference of music for music recommendation in daily activities," 2020 13th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2020, pp. 350–353, doi: 10.1109/IS-CID51228.2020.00085.

[3]. Y. H. Chin, S. H. Lin, C. H. Lin, E. Siahaan, A. Frisky and J. C. Wang, "Emotion Profile-Based Music Recommendation," 2014 7th Inter-national Conference on Ubi-Media Computing and Workshops, Ulaan-baatar, Mongolia, 2014, pp. 111–114, doi: 10.1109/U-MEDIA.2014.32.

[4]. K. Yoon, J. Lee and M.-U. Kim, "Music recommendation system using emotion triggering low-level features," in IEEE Transactions on Consumer Electronics, vol. 58, no. 2, pp. 612–618, May 2012, doi: 10.1109/TCE.2012.6227467.

[5]. S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player in-tegrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017,pp. 154–158,doi: 10.1109/WiSPNET.2017.8299738.

[6]. Nethravathi k.s, Priyanka, madhusree, "Emotion-Based Music Player," 2024 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), Luang Prabang, Laos, 2019, pp. 1–4, doi: 10.1109/ICEAST.2019.8802550.

[7]. S Metilda Florence and M Uma 2020 IOP Conf. Ser.: Mater. Sci. Eng. 912 062007.

[8]. Mahadik, Ankita Milgir, Shambhavi Patel, Janvi Kavathekar, Vaishali Jagan, Vijaya Bharathi. (2021). Mood based music recommendation system.

[9]. Kaur, Prabhjot Kapoor, Aditya Solanki, Yash Singh, Piyush Sehgal, Devansh (2022). Deep Learning Based Emotion Detection in an Online Class. 10.1109/DELCON54057.2022.9752940.

[10]. Rahul ravi, S.V Yadhukrishna, Rajalakshmi, Prithviraj, "A Face Expres-sion Recognition Using CNN and LBP ", 2020 IEEE.

[11]. F. Abdat, C. Maaoui and A. Pruski, "Human computer interaction using emotion recognition from facial expression ", 2011 UKSim 5th European Symposium on Computer.

[12]. Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats International Journal of Advanced Engineering Research and Science 3 236854.

[13]. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The extended

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

cohn-kanade dataset (ck+) A complete dataset for action unit and emotion-specified expression In 2010 ieee computer society conference on computer vision and pattern recognition-workshops 94–101 IEEE.
[14]. S. Deebika, K.A. Indira, Jesline, "A Machine Learning Based Music Player by Detecting Emotions ", 2019 IEEE.
[15]. S.L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," in IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1–12, 1 Jan.-March 2015.