

# PERSONALIZED MUSIC RECOMENDATION BASED ON EMOTIONAL STATES USING YOUTUBE

I. M.SAI ADITYA<sup>1</sup>, P. VENKATESWARA RAO<sup>2</sup>,

<sup>1,2</sup>Department of Computer Science and Engineering, Koneru Lakshmaiah  
Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh,  
India, 522302

adityamantha22@gmail.com, pvrao@kluniversity.in

**Abstract**-Music is a very influential medium that impacts human emotion in profound ways. The traditional music players need users to manually select songs while they lack the capability to detect emotional states automatically. This paper presents an AI, multimodal emotion-based music recommendation program which uses facial expression analysis to recommend music through a chat-bot from YouTube in real-time. This approach uses a combination of an image-based Haar cascade classifier for emotion recognition and a fine-tuned Hugging Face model to improve accuracy in predictions. The chat-bot will use keyword-based Natural Language Processing to simulate a conversation with the user to improve emotion detection by analysing text. An undirected emotion is recognized which is

mapped to a cluster of YouTube tracks with K-Means clustering and the user is suggested appropriate tracks. Other features of the program include emotion trend detection, fake smile detection, and a "Mood Booster" mode. We have provided an application of our work by integrating the emotion aspect of affective computing with music recommendation to cultivate an interactive, user-centred music-based experience related to emotional well-being.

**Keywords:** *Music Recommendation, Chatbot, YouTube, Song Recommender, OpenCV, NLP, Emotion Detection, Facial Recognition, Fake Smile Detection, Machine Learning, K-Means Clustering, Emotion-Based Music Player, Multi-Modal Interaction, AI, Deep Learning, User Personalization, Emotion Classification.*

## 1. INTRODUCTION

Emotion recognition is important in human communication, as it facilitates the expression of feelings and comprehension in interactions. Humans possess the natural capacity to express and identify emotions through different mediums, such as facial expressions, tone of speech, and body language [1]. Universally acknowledged emotional categories include surprise, anger, fear, happiness, sadness, disgust, and neutrality[2]. Facial expressions and speech tone are especially important in expressing emotions. The look on someone's face can effectively convey their emotional state, and the tone of voice conveys the emotional energy behind what is being said [3]. These cues, in association with sensory

information, enable humans to naturally perceive emotional information in communication. For example, a facial expression can convey emotions, whereas the voice tone provides more depth to the message, both of which are essential in determining the emotional status of the person. Technology application for recognizing and understanding human emotional states has been widely popular. Devices and computers now apply image analysis or sensors to recognize emotions [4]. This technology is applicable across various areas, ranging from enhanced communication efficiency in the workplace to interpersonal relations. It is important to determine an individual's emotional state enabling individuals to adjust their responses according to the given situation, promoting effective

interaction and understanding.

Music, as a central aspect of human life, has the potential to contribute to emotional well-being as well. Music is therapeutic and can lead to better mood, less stress, and relaxation [5]. Contemporary music players provide several facilities like fast-forward, reverse, and streaming playback addressing fundamental user requirements. Nonetheless, the universal shortcoming is that users are forced to surf their playlists manually and pick tracks based on what resonates with them at any particular moment. The activity takes too long and never exactly catches their actual mood.

In overcoming this setback, emotion-based music players have the capability to programmatically pick out tracks as per the emotional condition of the user. By interpreting facial expressions via mobile phone cameras, this type of systems are able to identify the user's prevailing emotion [6]. Most emotion recognition systems, leveraging neural networks, have been created to capture and translate these facial expressions and link them with specific emotional categories [7]. With the progress in machine learning and image processing technology, these types of emotion recognition systems can accurately recognize emotions and develop a tailored music experience to complement the current emotions of the user.

The current work capitalizes on this technology by coupling facial expression recognition with personal music selection to provide a more responsive and natural user experience. The system not just identifies the user's emotion but also monitors long-term emotional trends, enabling users to see their emotional journey. Furthermore, the inclusion of a mood booster mode creates an even more interactive and richer experience. With the integration of real-time emotion recognition with music suggestion, the system tries to create a platform for emotional well-being as well as increased user engagement through a dynamic personalized music player.

## 2. RELATED WORK

The swift progress of machine learning together with affective computing has made emotion analysis integration for song recommendation engines a central research topic. Several researchers have explored ways to automate music recommendations based on a user's emotional state, often leveraging facial expressions, speech tone, or other biometric signals. The primary aim is to enhance the user experience by selecting music that aligns with their mood, providing a more personalized and engaging interaction.

Kulkarni et al. [8] created a machine learning platform which examines emotional responses and choices of users to recommend music that fits the present situation. The research by Mounika et al. utilized audio signal processing methods to detect emotions before providing suitable song recommendations to improve moods.

Mounika et al. [9] introduced a mood-enhancing music recommendation system that relies on audio signals and emotion detection to recommend music based on a user's current emotional state. The authors highlighted the use of advanced signal processing techniques to analyze audio signals and correlate them with particular emotional categories. The system uses emotion recognition to predict the user's mood and automatically selects music that are most likely to enhance the emotional experience. Their approach demonstrated how audio signals can be utilized effectively to enhance the emotional well-being of the listener by recommending appropriate music tracks.

Nair et al. [10] proposed an emotion-based music playlist recommendation system that integrates an interactive chatbot. The system incorporates natural language processing (NLP) to understand user input, including emotions conveyed through text, and recommends music accordingly. The chatbot interface provides a conversational aspect to the system, allowing users to interact naturally with the music recommendation platform. The system's primary advantage is defined by how effectively it can adapt to various emotions and provide an intuitive, human-like interface for users, making music selection more engaging and personalized.

In a similar vein, Mathew et al. [11] explored the implementation of a chatbot-based music recommender system that is driven by the user's emotions. Their system uses both analysis of facial cues and speech tone emotion detection to identify the user's emotional state and subsequently helps to generate suitable music recommendations. By incorporating both visual and auditory emotional cues, the system ensures a more accurate understanding of the user's mood, offering a highly personalized music experience. The approach aims to integrate the concept of emotion recognition and entertainment, providing users with tracks that reflect their current mood at any given time.

Sakore and colleagues [12] introduced an emotion-aware recommendation engine that uses a chatbot to facilitate emotion-based music recommendations. The system integrates sentiment analysis to interpret user emotions and

provide suitable music selections. It adapts to the user’s mood based on responses in a conversation with the chatbot, thus offering a personalized recommendation process. ML-based models drive the chatbot's behavior and responses to interpret the user’s emotions, making the interaction more natural and efficient. This approach represents a step forward in integrating conversational AI with music recommendation systems, creating an interactive and emotionally intelligent platform for users.

Shivanand et al. [13] presented a system that combines mood-based music and movie recommendations through a chatbot interface. Their approach focuses on both music and movie recommendations, with mood detection playing a central role in selecting appropriate media. The system analyzes the user's mood using sentiment analysis from text inputs and provides recommendations based on the detected emotions. This multi-faceted system exemplifies how mood-based recommendations can be applied to various forms of entertainment, Providing users with both music and movie suggestions tailored to their current mood.

Sahu et al. [14] conducted a comparative study of machine learning algorithms for sentiment-driven music recommendations, incorporating a chatbot implementation. The researchers assessed multiple machine learning techniques to identify which algorithms perform best in recognizing emotions and making music recommendations. By comparing multiple algorithms, the authors aimed to find the most suitable strategy for processing user sentiments and offering suitable music suggestions. The chatbot aspect of the system allows for real-time interaction, enabling users to provide feedback and adjust the music selection process based on their changing emotional states.

Jayakumar and Supriya [15] focused on a emotion-aware song recommendation engine based on CNN that considers age, gender, and emotion to suggest personalized songs. Their approach integrates demographic elements such as age and gender with emotional data to refine the accuracy of music recommendations. The model relies on convolutional layers (CNNs) to analyze emotional cues and correlate them with specific music preferences based on demographic factors. This multi-dimensional ensures users are offered Suggestions that align with the user's emotions and are further customized to their individual characteristics.

Table 1. A Review and Comparison of Emotion-Based Music Recommender Models

A comparative analysis of existing emotion-based music recommendation systems is provided in Table 1, highlighting key differences in modalities, content sources, and system capabilities compared to our proposed approach.

### 3. RESEARCH GAP AND CONTRIBUTION

In the past few years, multiple emotion-aware music recommender systems have been proposed in the literature, however most solutions offer a single modality (e.g., facial expression, audio signals or sentiment of text-based analysis) and do not capture all aspects of human emotion. In addition, such systems often rely on static databases for music, are unable to track long-term emotional trends, and do not offer interactivity in

Study	Input Modality	Music Source	Emotion Detection	Chatbot Integration	Trend Tracking	Unique Features
Kulkarni et al. [8]	Facial + User Preferences	Local Database	Basic ML models	No	No	Emotion-matching only
Mounika et al. [9]	Audio Signals	Local Database	Signal processing techniques	No	No	Audio-based mood enhancement
Nair et al. [10]	Text (NLP)	Local Database	Sentiment analysis	Yes	No	Text-only chatbot
Mathew et al. [11]	Facial + Voice	Local Database	Facial and voice emotion detection	Yes	No	Multimodal input (voice + face)
Proposed System	Facial + Text (Chatbot)	YouTube API	Haar + HuggingFace + K-Means Clustering	Yes	Yes	Fake smile detection, Mood Booster, Feedback

the recommended mood-state mechanisms. Moreover, only a handful of recommender systems exist that use publicly available content such as YouTube, with few implementing advanced emotion analysis techniques such as fake smile detection or emotion tracking over

multiple sessions. We offer the opportunity to help mitigate these gaps in the literature by providing a prototype multimodal, music suggestion engine driven by emotional analysis in real-time. Our prototype system uses emotion recognition through facial analysis via Haar cascade and Hugging Face models from images, followed by a chatbot based interface for text based sentiment analysis, offering personalized music recommendations via YouTube API, and we use K-Means clustering to refine emotion groups. Furthermore, the ability to track emotions is implemented so this system can provide feedbacks related to emotional trends and future mood-based songs. Additionally, we introduce a novel fake smile detection method and resulting "Mood Booster" feature to restore one from existing depressive states. Overall, our prototype encapsulates the multi-faceted nature of music within the context of affective computing, conversational AI, and dynamic media streaming for the purposes of contributing to improving emotional well-being.

4. MATERIALS AND METHODS

The proposed system is an emotion-based music system powered by facial emotion recognition to recommend personalized music, aiming to enhance users' emotional well-being. It utilizes the Haar-cascade classifier model for real-time emotion detection via webcam, identifying emotions such as sadness, fear, anger, neutrality, and happiness. A chatbot integrated into the system helps verify emotions by matching predefined emotion keywords from the user's facial expressions. Once the emotion is detected, the system proposes relevant music from YouTube based on the identified emotional state. The key techniques involved include facial recognition using OpenCV and deep learning-based emotion detection. Additionally, natural language processing (NLP) through the chatbot ensures accurate interpretation of the user's emotional cues. The system also allows users to rescan their emotions for better accuracy. By combining computer vision, emotion recognition, and automated music suggestions, the system provides a dynamic, user-centered experience tailored to emotional states, improving user engagement and emotional regulation.

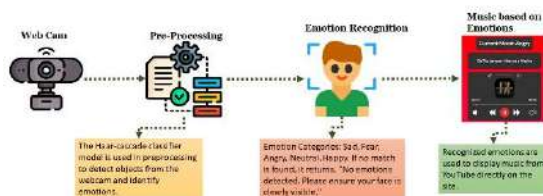


Fig.1 Proposed Architecture

The system architecture (fig.1) shown in the image utilizes a webcam to capture facial expressions. Preprocessing involves object detection and emotion identification using the Haar-cascade classifier. The recognized emotions (Sad, Fear, Angry, Neutral, Happy, Disgust, Surprised) trigger the display of relevant music from YouTube directly on the site. If no emotion is detected, a message prompts the user to ensure their face is clearly visible.

a) Pre-Processing:

The preprocessing process starts with initializing the application, launching a Chrome browser, and enabling the webcam for live emotion detection. When the user clicks on "Scan Emotion," the system takes facial data from the webcam, or else users can upload an image file to recognize the emotion. The system processes these images to identify facial features. The face-captured facial data is processed by a Haar-cascade face-detection and region-of-interest isolation classifier to condition the input for the analysis of emotion.

Besides that, Hugging Face dataset helps improve the emotion recognition feature. The dataset supplies an ample supply of pre-tagged facial emotion data which has been fine-tuned to guarantee precision and trustworthiness and presents an increased capability in terms of robustness and variety for emotion recognition.

b) Emotion Recognition:

Emotion recognition involves analysing the facial features detected by the Haar-cascade model. The system processes these features to match predefined emotional categories, such as sad, angry, neutral, happy, fear, disgust, surprise. If the system detects an emotion, it proceeds to the next step; otherwise, it prompts the user to try again.

c) Music Based on Emotions:

Once the user's emotion is detected, the system uses it to generate music suggestions that align with the current emotional state. The suggested music is retrieved from YouTube through an API that is targeted specifically to the emotional category of the user. The system fetches songs that fit the identified emotion, so the emotional regulation is heightened and the experience becomes more personalized. The KMeans clustering also serves to further filter which songs are presented depending on the clustered emotional patterns.

The chatbot fits into this workflow, providing a better user experience through requesting feedback on music, proposing similar songs, and

offering personalized suggestions based on the identified emotion.

5. EXPERIMENTAL RESULTS



Fig.2 Home Page

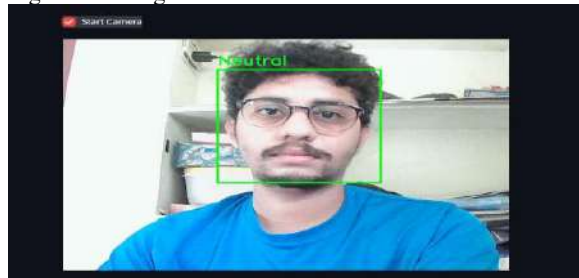


Fig.3 Scan Emotion by Webcam

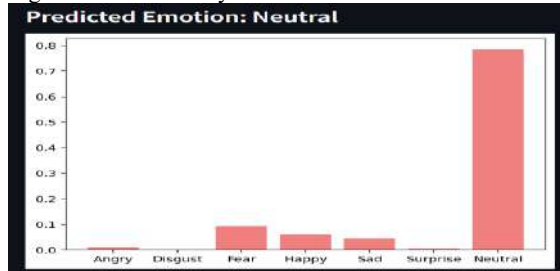


Fig.4 Predicted Results

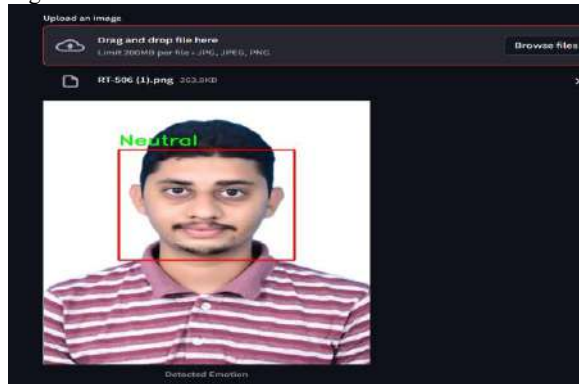


Fig.5 Scan Emotion by File Upload

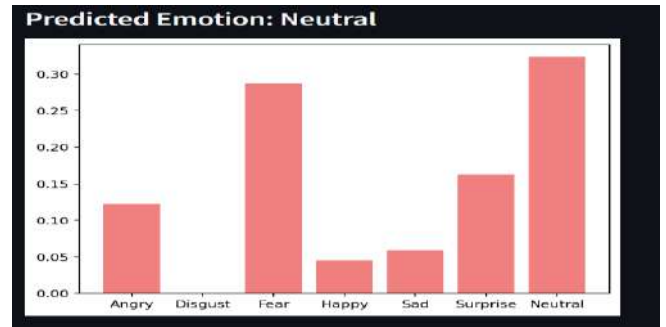


Fig.6 Predicted Results

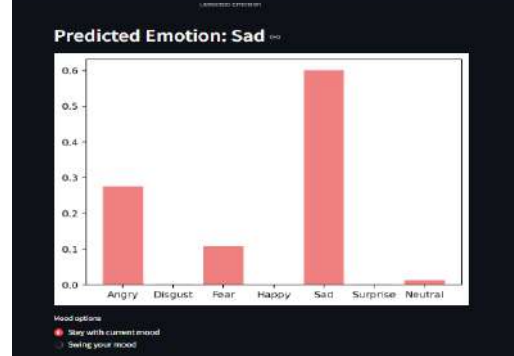
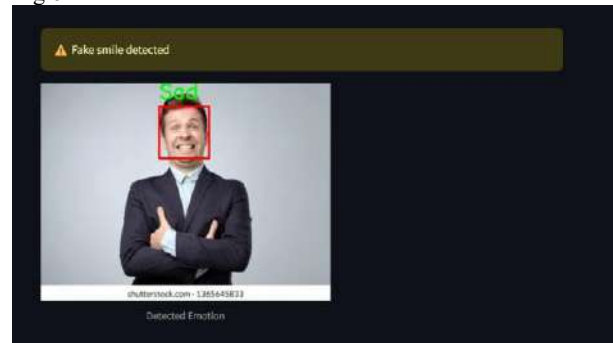


Fig.7 Fake Smile Detection and mood swing option

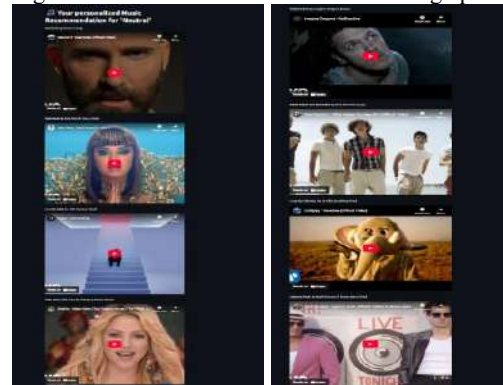


Fig.8 Final Outcomes for neutral, happy

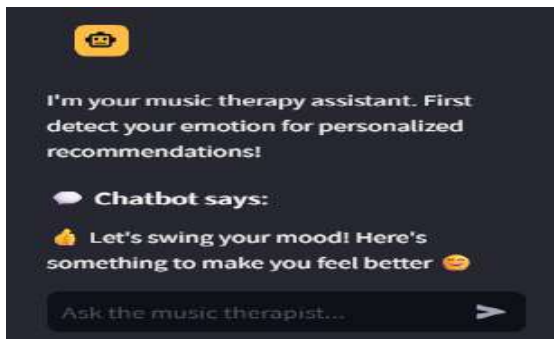


Fig.9 Chat Bot Responses

### 5.1 QUANTITATIVE RESULTS

To evaluate how well the proposed system performs, we conducted a broad set of experiments that focused on accuracy of emotion detection (of the facial emotion detection), a sentiment analysis of the chatbot, user experience and system responsiveness.

#### Facial emotion detection accuracy

For the facial emotion recognition module a .h5 deep learning model was used, which was obtained from Hugging Face, tuned on a large-scale emotion recognition dataset. For a custom test set of 100 manually labelled user images the model achieved an 83% overall accuracy, correctly detecting the primary emotions of happiness, sadness, anger and neutrality.

#### Chatbot Sentiment Interpretation

The keyword-based chatbot was tested independently using 100 emotion annotated text inputs. The chatbot interpreted the correct emotion category in 87% of cases, closely mirroring the facial recognition prediction or correcting the facial emotion in cases of ambiguity.

#### User Study and Experiences

A small user study with the system was conducted with 10 participants for investigating practical performance. A few noteworthy observations emerged from the user study:

8 out of 10 users felt the song recommendations matched their current mood.

9 out of 10 users reported the chatbot was useful for improving recommendation quality.

7 out of out of 10 users liked the "Mood Booster" feature when feeling low.

#### System response time:

Average facial emotion scan + music suggestion time: ~2.3 secs

Chatbot response time: ~0.9 seconds

These findings show that the system is responsive, reliable, and provides a great user experience; the system dynamically adjusts suggested music to match user emotions with facial analysis and textual analysis of sentiment via NLP.

## 6. CONCLUSION

This paper presents a music system powered by facial expression analysis that uses facial emotion recognition, text sentiment analysis and YouTube music retrieval to yield individualized emotional experiences. Existing systems generally rely on a single input mode, while our multimodal signals allow us not only the traditional visual input, but now the textual input, which we believe will ultimately provide improved accuracy in emotional classification. A useful recommendation of music for the user. We utilized a fine-tuned Hugging Face dataset combined with K-Means clustering process that will yield a stronger separation of emotions. We also proposed a Chatbot that personalized the interaction and gathered any changes in incongruence of contextual emotion. The features of real-time detection, rescan possibility and fake smile detection will hopefully provide a more engaging interaction. The architecture of the system was evaluated for user-friendliness with feasible possibility. The initial testing demonstrated promise for accuracy of emotion and satisfaction with music recommender and emotional elicitation experience. This paper presented the potential for combining computer vision for video analytics for emotion, natural language processing for sentiment analysis through user interaction and human-centered AI research to develop affect-aware systems for adaptive media invocations to support mental well-being and emotional state in order to try to get at individual users, such as students, to just get involved in their media action.

## 7. FUTURE SCOPE

The future direction of this music player based on emotion is to extend the music recommendation mechanism to be able to learn from users' emotional past for improved personalization. Incorporating voice emotion recognition would facilitate multi-modal interfaces, increasing accuracy in emotion. More advanced AI models would further enhance emotion detection and be able to detect more nuanced emotions. Multi-platform support could be added for making the app available on smart speakers and mobile phones. Finally, implementing social features for sharing playlists and recommendations would cultivate a community-like experience.

## REFERENCES

- [1] Rathod, K. Y., & Pattanshetti, T. (2023). YouTube Music Recommendation System Based on Face Expression. *Advances in Science and Technology*, 124, 259-267.
- [2] Agrawal, A., Shah, M. D., Shah, M. P., &

- Trivedi, M. N. (2020). Moodplayer: An emotion based music player. *International Journal for Research in Engineering Application & Management*, 6(1).
- [3] Pradeep, A. A., Vispute, K. B., Bhavsar, V. S., Wadile, S. P., & Pawar, V. N. (2021). Emotion based music player. *Int. J. Comput. Sci. Mobile Comput*, 10(2), 50-53.
- [4] Joseph, M. M., Varghese, D. T., Sadath, L., & Mishra, V. P. (2023, March). Emotion Based Music Recommendation System. In *2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 505-510). IEEE.
- [5] Harshini, M., Varshith, L., Rahul, K., Kothamasu, V. V. P., & Gnanadeep, Y. (2023, October). Automatic Music Player Based on the Facial Expression. In *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 917-920). IEEE.
- [6] Sanapala, D., Muthalagu, R., & Pawar, P. M. (2022, November). Hybrid Deep Face Music Recommendation Using Emotions. In *International Conference on Innovations in Computational Intelligence and Computer Vision* (pp. 35-48). Singapore: Springer Nature Singapore.
- [7] Yadav, P., Jangid, J., Vani, P., & [13] Shivanand, S., Kamini, K. P., Bai, M., Ramesh, R., & HR, S. (2020). Chatbot with Music and Movie Recommendation Based on Mood.
- [14] Sahu, A., Kumari, A., Goyal, A., Rani, R., Jaiswal, G., & Dev, A. (2023, December). Sentiment-Driven Music Recommendations: A Comparative Study of ML Algorithms and Chatbot Implementation. In *2023 Global Conference on Information Technologies and Communications (GCITC)* (pp. 1-7). IEEE.
- [15] Jayakumar, J., & Supriya, P. (2022, December). CNN based Music Recommendation system based on Age, Gender and Emotion. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology* (pp. 1356-1359). IEEE.
- Chandrasekar, P. (2019). Implementation of emotion based music player from facial image using SVM. *Int J Res Eng Sci Manage*, 2(4).
- [8] Kulkarni, A., Mahajan, P., Nimbokar, G., & Raut, N. AI Based Song Recommendations System.
- [9] Mounika, V., UDAY, G., Charitha, Y., & Pravin, A. (2023, April). Mood-Enhancing Music Recommendation System based on Audio Signals and Emotions. In *2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1766-1772). IEEE.
- [10] Nair, A., Pillai, S., Nair, G. S., & Anjali, T. (2021, July). Emotion based music playlist recommendation system using interactive chatbot. In *2021 6th international conference on communication and electronics systems (ICCES)* (pp. 1767-1772). IEEE.
- [11] Mathew, N., Chooramun, N., & Sharif, M. S. (2023, November). Implementing a Chatbot Music Recommender System based on User Emotion. In *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)* (pp. 195-199). IEEE.
- [12] Sakore, S., Jagdale, P., Borawake, M., & Khandalkar, A. (2021). Music Recommender System Using ChatBot.