EmotionInsight and BeatSync: a music recommendation system based on sentiment analysis

[1]Ishanvi Jaiswal, [2]Shruti Raj, [3]D Aniketh Patil, [4]Anirudh Rao, [5]Yashaswini N

[1234]Undergraduate student, [5]Assistant Professor
[12345]Computer Science Engineering,
[12345]JSS Science and Technology University, Mysore, India

[1]ishanvijaiswal04@gmail.com, [2]rajshrutidto@gmail.com, [3]danikethpatil@gmail.com, [4]anirudhrao1008@gmail.com, [5]yashaswini@jssstuniv.in

***Abstract*— Although tailored music recommendation systems have seen major developments in recent years, their integration with user emotions derived from several input modalities is still rather understudied. EmotionInsight and BeatSync is a new music recommendation system presented in this work that dynamically matches musical suggestions with users inferred emotional states derived from both textual and video-based inputs. Using natural language processing (NLP) methods, the system extracts sentiment and emotional context. Face expressions and emotional cues from video frames are analyzed concurrently using deep learning models. Using a multi-modal sentiment fusion technique, the extracted emotional elements are then mapped to a well-chosen music database, so providing context-aware and emotionally resonant song recommendations.**

***Keywords*— music recommendation, text-based sentiment analysis, image-based sentiment analysis**

I. INTRODUCTION

In the domain of human-computer interaction, crafting individualized and engaging user experiences hinges significantly on understanding and responding to human emotions. Due to advances in artificial intelligence, specifically in natural language processing and computer vision, it is now possible to capture emotional cues from both text and visual content. Emotion Insight and BeatSync is a new framework that establishes an adaptive setting wherein technology accommodates the user's current mood through filling this emotional intelligence with music recommendation and play.

Text data and image data are the two types of inputs that the system is programmed to process. This allows the model to recognize emotional signals in facial expressions in images as well as text like social media updates or text messages. The model employs deep learning-based sentiment analysis models to recognize the emotion with seven standard categories: happy, sad, disgust, angry, fear, surprise, and neutral. To preserve and analyze both textual and visual emotional cues, each input i.e. text or image is processed individually.

After the identification of the emotions, the data is provided as a prompt to a large language model (LLM). The model then generates suggestions of music appropriate to the mood of the user. The LLM contextually builds prompts based on its knowledge of the emotional mix. A self-curated music database with Hindi, English, and Kannada tracks is used to dynamically generate these suggestions, maintaining linguistic and cultural diversity across different users.

Playback is the last operation performed by the system. The chosen songs are played using the API through google, and users can respond to the suggested music immediately without having to navigate away from the website. Multimedia playback, sentiment analysis, and language modeling are all integrated cohesively to generate an emotionally attuned system that not only understands the emotions of the user but also enhances the emotions with the music that talks to them in a personal sense.

Illustrating the compelling synergy of artificial intelligence, emotion recognition, and multimedia applications, EmotionInsight and BeatSync demonstrates an integrative approach to emotion-based music recommendation. Emerging technologies in emotion-aware technologies emphasizing user interaction, personalization, and empathy are made possible by the system.

II. LITERATURE REVIEW

The adoption of emotion analysis in music recommendation systems has grown manifold in the past few years, specifically with the advent of AI, Natural Language Processing (NLP), and real-time user interaction technologies. The subsequent works critique different approaches involved in the creation of intelligent music recommendation systems that meet the emotional requirements of users, providing us with vital knowledge on how to build our proposed system, Emotion Insight and BeatSync.

Researchers in [1] suggested Mood Waves, a music recommendation system that is based on a chatbot with a focus on interpreting the emotional tone of the user via text-based conversation. The novelty of this system lies in its emotional suggestion for music, where it examines mood markers in text to recommend songs that match the emotional tone of the user, whether soothing, uplifting, or motivational. Drawing on external data from sites such as Last.fm, the chatbot generates mood-based playlists in real time and converses with the user. Moreover, the system adjusts the user's mood in real time, altering the music experience over time. Finally, the personalized quality of the recommendations and the ability to create and share playlists imply a stronger desire for musically mediated social and emotional connection.

A Chatbot Song Recommender System Based on Emotion Analysis was proposed in [2], which integrates hybrid recommendation models and sentiment analysis to enhance emotional accuracy. The system examines user emotions like happiness, sadness, calmness, or excitement from text inputs through Natural Language Processing (NLP). On these recognized emotions, it recommends songs based on collaborative filtering, content-based filtering, or a hybrid model, with which the system can generate better and more relevant recommendations. In addition, it has a mechanism for obtaining user feedback, hence allowing continuous learning and song recommendation improvement. Over time, the system can learn from user preferences thanks to this feedback mechanism. In order

to make the user feel heard and understood, dynamic user interaction also guarantees that the chatbot behaves in a conversationally and emotionally sensitive manner.

The authors in [3] created a Chat Bot Song Recommender System that takes mood-based song suggestion as its core feature. It utilizes Natural Language Processing (NLP) to process text inputs, perceive emotions of the user, and suggest music based on their emotions. The system is also equipped with a user-friendly interface that makes signing up, login, and chatting with the chatbot easy. The chatbot analyzes current conversations and provides recommendations in real time, enhancing the listening experience by offering emotionally intelligent suggestions.

A Sentiment Analysis-Based Music Recommender: SentiSpotMusic [4] takes a data-driven approach to making recommendations for emotionally charged music. The system's central feature is sentiment classification into positive, negative, or neutral, and employing this classification to create emotion-based playlists. A new addition is the application of a Tableau dashboard, enabling users to see patterns of sentiment in popular songs and monitor listening trends like top songs and most-played artists. By providing information on musical tastes, this increases user interaction. More personalized playlists are possible due to the statistical analysis of listening behavior by the system. This method provides access to data-driven emotional suggestions by linking sentiment analysis and user analytics.

In [5], Sentiment Analysis in Recommender System presents how deep learning models are combined with sentiment-aware recommendation techniques. Specifically, it applies a hybrid model with Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for efficient sentiment classification. This categorized data is then processed with user-based collaborative filtering to forecast ratings and suggest music. The system will be adaptive and thus can reply to dynamic variation in user sentiment and preference. Although this piece of work best applies to industries such as foods and movies, its architecture would also apply to music, and it provides an opportunity for designing real-time sentiment-sensitive music recommenders. The system's flexibility and scalability, along with its capacity to handle several data points and affect cues, make it a useful guide for the development of emotionally intelligent systems.

In [6], the authors examined recent research on music recommendation systems and how incorporating user emotions can improve personalization. Traditional approaches like content-based filtering and collaborative filtering primarily concentrate on listening history and user similarities, but they do not account for emotions. More sophisticated systems, such as the Emotion-Aware Personalized Music Recommendation System (EPMRS), utilize deep convolutional neural networks (DCNNs) for music classification along with weighted feature extraction (WFE) to link user emotions with music preferences. These methods leverage both audio signals and metadata to uncover hidden features and create song recommendations that resonate emotionally. When compared to traditional systems, EPMRS showed enhanced accuracy by learning from both user behaviors and the characteristics of the songs, emphasizing the success of personalization driven by emotions.

DJ-Running system puts forward a recommendation approach for music based on the emotions of the runners, explained in [7]. Contrary to other classical systems, which rely considerably on users' habits and past behavior in listening to music, DJ-Running is based on emotional and contextual real-time data during running, like the runner's location, mood, training plan, and physical indicators. The system labels songs with emotional tags using a Music Emotion Recognition (MER) model and proposes tracks based on the user's profile as well as their current circumstances. It employs technologies such as fog computing, wearable devices, and Spotify integration to seamlessly modify music playback. This approach emphasizes the growing significance of emotion- and activity-sensitive systems in customizing music recommendations.

This study in [8] introduces a tailored, emotion-sensitive music recommendation system aimed at promoting users' mental and physical health. It highlights how music can affect emotional and psychophysiological states, utilizing AI methods to pick songs that match or enhance a user's emotional state. The system forms user profiles from rating analysis, listening history, and sensor data. Then these profiles are automatically enhanced incrementally and via reinforcement learning methods. Music is then classified on the basis of content features and user context like mood, activity, and setting. The prototype has successfully formed playlists and thereby established how music attributes with emotive correspondences could profoundly increase users' experience across many situations such as work productivity, therapy, and sport.

The paper in [9] examines how Spotify employs collaborative filtering, a common method found in music recommendation systems. It demonstrates how Spotify constructs user-item interaction models utilizing K-nearest neighbors and matrix factorization to associate users with comparable listening habits. By evaluating implicit feedback such as song favorites, skips, and listening history, Spotify's system generates dynamic and tailored playlists like Discover Weekly. Through the implementation of continuous learning, this approach improves the quality of recommendations and lessens the reliance on direct user input. Despite facing issues like data sparsity and the cold-start dilemma, collaborative filtering continues to effectively deliver personalized recommendations driven by large-scale behavior.

III. PROPOSED ARCHITECTURE

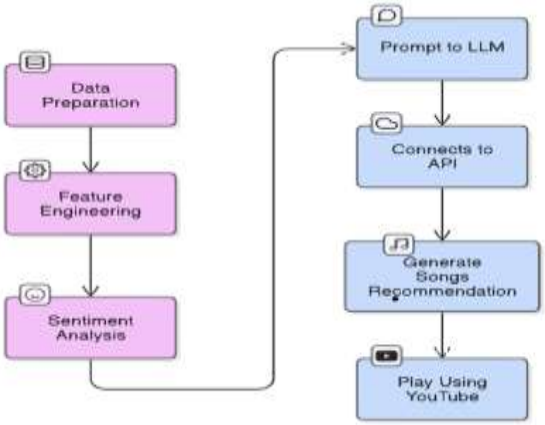

Figure 1

In this proposed architecture, EmotionInsight and BeatSync is built as an extensive multi-stage pipeline of sentiment analysis and music recommendation based on AI. The system starts with data preparation, where text and image data are gathered from various sources like Kaggle and Twitter. The text information is subjected to thorough preprocessing, such as cleaning of text to eliminate noise, tokenization to split sentences into words, stemming to normalize words to their base form, and negation handling to maintain sentence meaning. Besides, advanced text vectorization techniques like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) are applied to transform text information into numerical representation. In the meantime, image data are processed by using facial recognition techniques. Facial detection is conducted first, then 68 facial landmarks are obtained, and eye-based facial alignment is done. Face region is cropped and resized to maintain consistent input dimensions. Preprocessing also involves normalizing pixel values to [0,1] and using Albumentations to augment images for better model generalization. Preprocessing of facial data is also enhanced by calculating Euclidean distances between certain facial landmark points, which results in a robust numerical facial expression representation.

After preprocessing the data, the feature engineering stage converts raw data into structured inputs that can be used as input to machine learning models. For text data, feature extraction methods like TF-IDF, Bag of Words, and Word2Vec are used to preserve contextual and semantic word relationships. For face data, a Convolutional Neural Network (CNN) is used to learn deep features from facial expressions to allow the model to learn subtle patterns related to varying emotions.

After feature extraction, the sentiment analysis module is utilized to classify text and image inputs into seven emotion categories: happy, sad, disgust, angry, fear, surprise, and neutral.

Text sentiment classification is performed with the aid of two machine learning algorithms namely Logistic Regression and Naïve Bayes Classifier (MultinomialNB) that operate based on the structured text features to predict the emotional states. CountVectorizer is also employed to enhance the quality of feature representation to improve the accuracy of the classifier. Sentiment analysis from images is achieved through a CNN model that works with the extracted features to make accurate predictions for emotional states by exploiting spatial dependencies of facial expressions.

Once classified emotions from text and image data are passed through a Large Language Model (LLM), which offers personalized song recommendations. The LLM is fine-tuned on our own music database with songs categorized in Hindi, English, and Kannada. This tuning aligns the suggestions with the sentiment that is identified, providing a smooth and emotion-based experience to the user. The LLM dynamically translates classified emotions, creates contextual prompts, and fetches appropriate songs from the database.

The last step is API integration, where the suggested songs are retrieved and played with YouTube API. This allows real-time streaming of music according to the user's sensed emotional state, providing an AI-driven personalized listening experience. Utilizing machine learning, deep learning, and natural language processing, EmotionInsight and BeatSync delivers a smooth and intelligent system boosting user engagement, emotional connection, and general well-being through music.

IV. EXPERIMENTAL SETUP AND DATASET

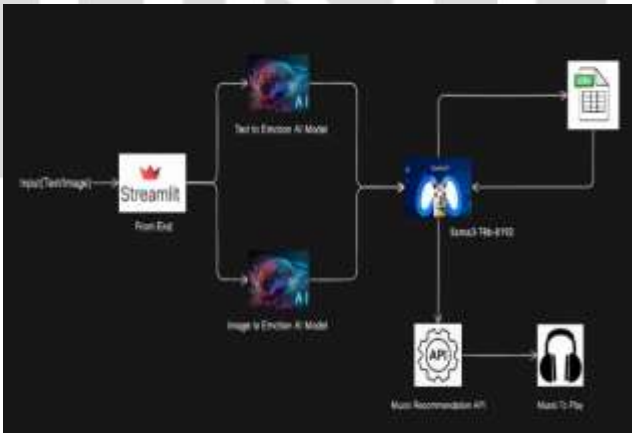

Figure 2

EmotionInsight and BeatSync research system adopts a systematic multimodal pipeline strategy which involves sentiment text analysis and facial expression, advanced large language model (LLM) for suggesting songs based on user requirements, and an API-supported music play platform. The most basic purpose of the system is to process the user's emotion from either textual or facial pictures input and provide suggestions based on them. This process starts with the collection of data from various sources so that the models are trained on varied datasets that reflect actual emotional expressions in the real world. Text data is collected from sources like Kaggle and Twitter, which include labeled emotional text, and image data is collected from Kaggle facial expression datasets with labeled emotions. For music suggestions, a carefully crafted dataset from Last.fm and Kaggle is used, with a diverse range of songs in English, Hindi, and Kannada, each having mood attributes assigned to them. The front-end interface, implemented using Streamlit, accepts input in the form of text or an image to detect emotion.

For text classification emotion analysis, the system processes the text before analyzing the emotion first by undergoing preprocessing with techniques of tokenization, stemming, lemmatization, and removal of stop words to enable classification with just the most valuable words. Transformational techniques beyond that, TF-IDF (Term Frequency-Inverse Document Frequency) and CountVectorizer, convert textual data to numerical data forms. The preprocessed text is then mapped into one of seven emotional tags: happy, sad, disgust, angry, fear, surprise, or neutral, through the application of Logistic Regression and Naïve Bayes Classifier (MultinomialNB). Taking for instance that a user input the text "I am happy", the steps of preprocessing shall identify keywords and transform them into feature vectors so that a "happy" classification shall be tagged to it. This labeled sentiment is then passed as input to LLaMA 3 (70B-8192), a highly tuned LLM that has been trained on a proprietary music recommendation database. In accordance with this feeling, the LLM provides song recommendations that have an optimistic and upbeat tone, like "Happy - Pharrell Williams," "Gallan Goodiyan - Dil Dhadakne Do," or "Huttidare Kannada - Rajkumar." These suggestions are then forwarded to the Music

Recommendation API, which retrieves the respective tracks and plays them using the YouTube API, giving an uninterrupted and real-time audio playing experience. For image-based sentiment analysis, the system adheres to a more intricate processing pipeline. When a user uploads an image, facial preprocessing techniques are employed to ensure high-accuracy classification.

This involves face detection, facial landmark extraction (using a 68-point model), alignment from the eye positions, cropping of the face area, and resizing to a specified input size for the Convolutional Neural Network (CNN). Image normalization is done to normalize pixel values in the range [0,1] to enhance model generalization. Data augmentation methods with Albumentations are also used to increase the model's robustness against lighting, pose, and facial occlusion variations. Feature extraction is done by calculating Euclidean distances between certain landmark pairs, which capture facial expression changes that indicate various emotions. The CNN model subsequently processes the extracted features and labels the image as one of the seven emotions. For example, if a user posts an image of their face showing sadness, the model will identify major facial indicators like drooping eyebrows, downturned lips, and slightly closed eyes, which are typical of sad expressions. The CNN labels the image as "sad", and this label is forwarded to LLaMA 3 (70B-8192) for creating a suitable music recommendation. Since the sadness has been identified, the LLM pulls out songs that match this mood, suggesting sad songs like "Someone Like You - Adele," "Channa Mereya - Arijit Singh," or "Neene Neene - Sonu Nigam." These suggestions are then passed through the Music Recommendation API, which retrieves the songs and plays them through YouTube API.

The datasets for this project are essential to guarantee high accuracy and successful generalization to a wide range of inputs. The facial emotion recognition model is trained on large-scale Kaggle datasets with thousands of labeled facial expressions. The text sentiment analysis model is trained on large datasets from Kaggle and Twitter with tweets and text samples categorized by emotion. These datasets enable the CNN to learn subtle detail in facial movement and expression and thus be able to differentiate among subtle emotional hints.

The dataset for music recommendations, gathered from Kaggle and Last.fm, includes a long list of songs associated with mood features, so the LLM can provide emotionally appropriate song suggestions. By combining text-based sentiment analysis, image-based facial emotion recognition, a fine-tuned LLM for personalized music recommendations, and API-based music playback, EmotionInsight and BeatSync develop an innovative and engaging music experience that is responsive to the user's emotions in real-time. By supporting both text and image inputs, the emotion detection system becomes more inclusive and robust, treating users to an unprecedentedly personalized musical experience aligned with their existing mood.

V. RESULT

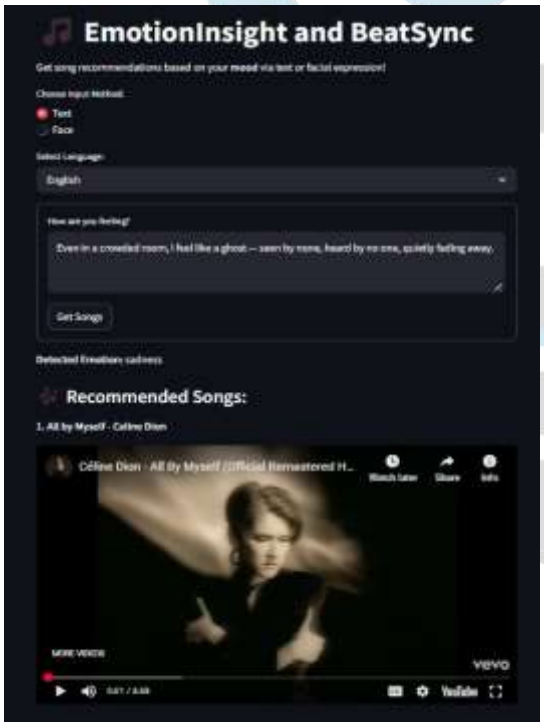

Figure 3

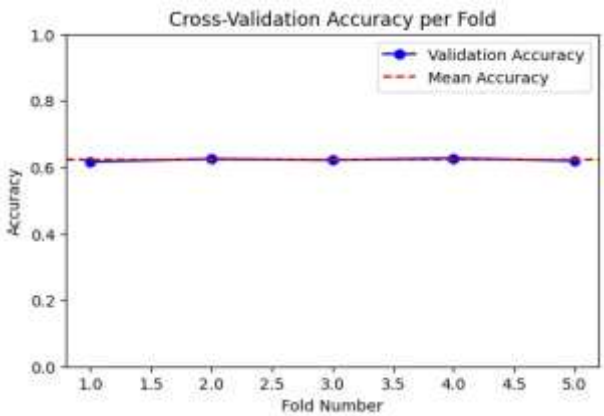

Figure 4

Responses from the model's text input were solid and suitable for the context. For instance, if the input is "Even in a crowded room, I feel like a ghost - seen by none, heard by no one, quietly fading away" the model correctly recognized sadness as the underlying emotion. The system then used the mapping based on LLM to suggest three songs most suited to the user's emotional state such as "All by Myself" by Celine Dion and "I'll Never Love Again" by Lady Gaga. Playback was seamlessly integrated using YouTube so that there would be an enjoyable user experience.

The performance of the model was assessed using 5-fold cross-validation, obtaining a stable accuracy of approximately 60% for every fold. As illustrated in the cross-validation plot, the validation accuracies are closely bunched around the mean with little variance, showing robust model stability and good generalization to new data. Such consistency implies that the current method, the integration of Logistic Regression and CountVectorizer as feature extractors, is reliable for the task.

Although the obtained accuracy is good enough for a multi-class emotion classification task - given the inherent difficulty and subjectivity of emotional text understanding - improvements can be sought further. But the stability of accuracy across folds suggests that tuning alone would not produce dramatic gains. Dramatic improvement in performance in the model would probably involve experimentation with more advanced algorithms, e.g., more advanced ensemble methods, deep learning models or more advanced feature engineering techniques.

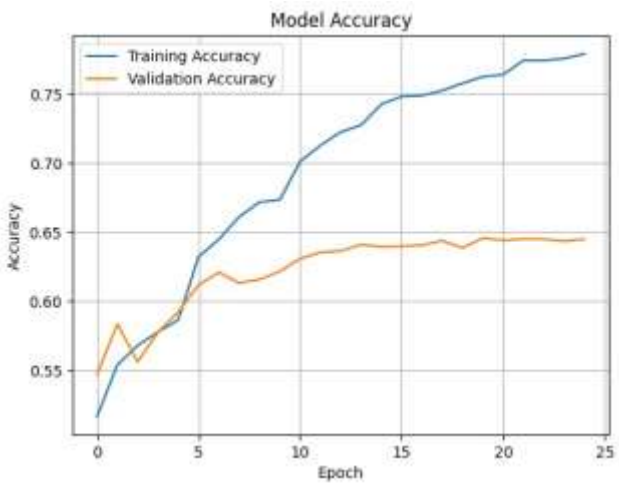

Figure 5

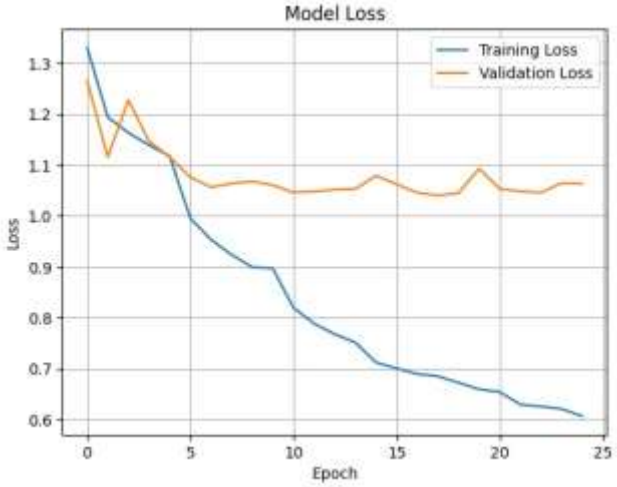

Figure 6

Figure 7

For image input, the EmotionInsight and BeatSync system uses a Convolutional Neural Network (CNN)-based deep learning model to detect emotions. Users can take real-time images, and the backend analyzes these images to detect the associated emotion. In the above instance, the model is able to identify the emotion "happy" from the snapshot captured, on which basis the system recommended three appropriate Kannada songs corresponding to the emotional state of the user, thus presenting an energetic and personalized music playback experience.

Looking at the graphs of training and validation accuracy, we see that the training accuracy consistently increases over the epochs up to nearly 78%, while the validation accuracy stabilizes at around 64%. This stability in validation accuracy shows that the model is working consistently on unseen data without serious overfitting.

The model's loss graph also clearly indicates that both the training and validation losses decreased as epochs increased. The reduction in loss after each training iteration is consistent and significant, which indicates that the model is learning, achieving better

performance each time it is updated. The loss decreasing means that recommendations are getting better over time and becoming more accurate.

So, the model shows dependable performance in identifying emotions from text and image inputs, followed by accurate and appropriate song recommendations. The effectiveness of the models employed is demonstrated by the steady accuracy and declining loss trends. The outcomes confirm that EmotionInsight and BeatSync can provide a customized and emotionally sensitive music experience.

VI. CONCLUSION

EmotionInsight and BeatSync present a combination of AI-based sentiment analysis and music recommendation in a personalized and engaging user experience. Through the analysis of text and facial responses, the system provides a better and more comprehensive emotional analysis, allowing the user to obtain music recommendations based on their emotions. The use of large language models (LLMs) adds to the system's capacity to create emotionally engaging playlists, with API-based music playback providing an uninterrupted listening experience.

This project demonstrates how AI can bridge the gap between human emotions and digital entertainment, with potential applications in mental health treatment, content delivery based on individual needs, and interactive media experiences. As the system develops with improved emotion detection and a greater music database, it can further enable human-computer interaction by having technology that is more attuned to emotional states.

VII.FUTURE WORK

Future enhancement for BeatSync and EmotionInsight may include emotion recognition via voice for higher accuracy and more comprehensive emotional analysis. Adding additional genres and local tracks to the music database and incorporating user feedback loops will enhance recommendations over time. Emotion-tracking in real time can facilitate adaptive playlist adjustment based on changing moods, and extending recommendations to podcasts or automatically generated music can enhance personalization too. Optimizing mobile and wearable devices will increase accessibility, and the system will be more responsive and adaptable to emotional conditions.

REFERENCES

[1] Gulshan Kumar, Riya Varshney, Shiva Tripathi, Ms. Garima. Mood Waves: A Chatbot-Based Music Recommendation System. 2024

[2] Rohit Sharma, Sahil Azad, Katiyar. Chatbot Song Recommender System. 2023.

[3] Prof. Suvarna Bahir, Amaan Shaikh, Bhushan Patil, Tejas Sonawane. Chatbot Song Recommender System. 2023.
K. Elissa, “Title of paper if known,” unpublished.

[4] Eva Sarin, Megha, Srishti Vashishtha, Simran Kaur. SentiSpotMusic. 2021.

[5] Cach N. Dang, Maria N. Moreno-Garcia, Fernando De la Prieta. Sentiment Analysis in Recommender System. 2021.

[6] Abdul A, Chen J, Liao HY, Chang SH. An emotion-aware personalized music recommendation system using a convolutional neural networks approach. Applied Sciences. 2018.

[7] Álvarez P, Guiu A, Beltrán JR, de Quirós JG, Baldassarri S. DJ Running: An Emotion-based System for Recommending Spotify Songs to Runners. InicSPORTS 2019.

[8] Rumiantcev M, Khriyenko O. Emotion Based Music Recommendation System. InProceedings of Conference of Open Innovations Association FRUCT 2020.

[9] Madathil M. Music recommendation system spotify-collaborative filtering. Reports in Computer Music. Aachen University, Germany. 2017

[10] Xiong LIU and Haiqing LIU, (2021), “Data Publication Based On Differential Privacy In V2G Network” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, DOI 10.30696/IJEEA.IX.I.2021.

[11] Maasø A, Hagen AN. Metrics and decision-making in music streaming. Popular Communication. 2020.

[12] Chen HC, Chen AL. A music recommendation system based on music data grouping and user interests. InProceedings of the tenth international conference on Information and knowledge management 2001.

[13] Fang J, Grunberg D, Lui S, Wang Y. Development of a music recommendation system for motivating exercise. In2017 International Conference on Orange Technologies (ICOT) 2017.

[14] Mandava Siva Sai Vighnesh, MD Shakir Alam and Vinitha.S, (2021), “Leaf Diseases Detection and Medication” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 01-07, doi 10.30696/IJEEA.IX.I.2021.

[15] Alex M. Goh and Xiaoyu L. Yann, (2021), “A Novel Sentiments Analysis Model Using Perceptron Classifier” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 4, pp. 01-10, DOI 10.30696/IJEEA.IX.IV.2021.

[16] Lu, Z., Cao, L., Zhang, Y., Chiu, C. C., & Fan, J. Speech sentiment analysis via pre-trained features from end-to-end asr models. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020.