Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Multimodal Approach to Emotion-Aware Music Recommendation: Integrating Demographic and Emotional Features with Machine Learning

Mrs Ketaki Amrut Bhosale[1], Dr Sangram Patil[2]
[1]*Research scholar, Department of Computer Science & Engineering, D Y Patil Agriculture & Technical University, Vathar-Warna Road, Talsande Fata, Talsande, Kolhapur-416112, Maharashtra, India*
[2]*Associate Dean, Department of Computer Science & Engineering, D Y Patil Agriculture & Technical University, Vathar-Warna Road, Talsande Fata, Talsande, Kolhapur-416112, Maharashtra, India.*

ARTICLE INFO

Received: 28 oct 2024

Revised: 18 Nov 2024

Accepted: 26 Dec 2024

ABSTRACT

Music has been considered to be a universal language, which has the ability to affect the feelings and make cultural identities. With the ongoing development of digital technologies, personal music recommendation systems are now fully embedded and become a part of life as the engine provides custom-crafted playlists, which match personal taste. Although conventional music recommendation systems mostly use any previous listening data, the addition of demographic characteristics like age or gender gives a more detailed analysis of the unlike preferences of the user. Also, feelings cues such as facial expressions or body reactions complement the personalization of music recommendations. The paper suggests a multimodal model of recommending music that would be based on both demographic aspects (ages, gender) and emotional indicators (mood, facial expressions) based on the latest machine learning methods. The system is based on the Convolutional Neural Networks (CNNs) and other deep learning models to identify emotions and demographic features to improve the accuracy and relevance of music proposals. The combination of the modalities makes the system provide an opportunity to have a personalized experience in getting exposed to music, which is emotionally stirring as well as contextually suitable. The findings prove that emotion-sensitive music suggestions can greatly enhance user-interaction and satisfaction, resulting to a more personalized and influential music playlists.

Keywords: Music recommendation, Multimodal approach, Emotion recognition, Machine learning, Age detection, Gender detection.

1. INTRODUCTION

The music recommendation systems has become one of the most important digital music features that stands out more than ever especially in the era of streaming platforms. These systems are meant to provide the user with music recommendations depending on the different factors like the past listening habit, music preference, and context. With the growing ability of people to consume music, move it to digital platforms, music recommendation systems have developed to address different needs of customers, giving them customized playlists and song recommendations, which depend on personal tastes and preferences[1]. The rise of digital music services has caused more sophisticated recommendation algorithms to be developed which exploit the mass of data, person-computer interaction, and intricate algorithms to provide customized music experiences[2]. Simply put, the goal of these systems is to select a plethora of music material and offer its users a more narrow, relevant listening experience.

There are two major approaches to traditional music recommendation systems, which include collaborative filtering (CF) and content-based filtering (CB). The collaborative filtering systems make a recommendation based on certain trends determined on the choices of the user, which is based on the assumption that when two individuals had liked similar music previously, the two individuals are likely to like similar music in future. In turn, content-based filtering[3] suggests music selections by the attributes of the song itself, which could include the genre, rhythm, or different lyrics, and contrasts these measures with a history of the user. Both approaches have weaknesses, but they possess strengths as well[4]. The problems that collaborative filtering may have difficulties with are related to sparsity of data and cold-start problem when users or items do not have enough information to be recommended in a proper way[5]. Although effective in what it does, content-based filtering may have its scope, as it lays more emphasis on the attributes of the music itself, and it is not concerned with the external variables, such as the emotions of the user or the demographic information.

Personalization of the music recommendation is a very critical factor that cannot be overemphasized. This individualization is the core of meaningful and engaging experiences provided to the user since it is guaranteed that the recommendations are not the most relevant but also individualized[6]. Individualized music suggestions enable the users to provide them with new songs and artists which they otherwise would not have found, as well as to make sure that the music suggestions are based on the current mood, activity, or context. This capability to offer music that best fits the user emotionally or according to their tastes in this context, the demographic factors, including age and gender, improve the engagement and satisfaction of the user[7]. Personalized systems are also effective in retaining users in as much as they give them a feeling that they are connected to the platform, making the experience unique and personal to the user[8].

Multimodal approach of music recommendation is a notion that has been of oncoming interest in the last few years. A multimodal approach uses information in more than one form where the information may include audio features, demographics of the user, emotional reaction, and the context. This will enable a more holistic view of the preferences of a user, as well as counseling them in a way that makes the recommendations more relevant and authentic. The mood that can be read by analyzing facial expressions or other physiological reactions has been demonstrated to affect the music preferences greatly[9]. Combining these emotional data with the classical types of recommendations could potentially provide the user with a much more personal and emotionally high-resolution musical experience. Secondly, demographic factors such as age, gender, and place are also included in the process of the recommendation that adds more information into it, and links the suggestions offered by the system to the identity and situation of the user[10].

This study aims to investigate the possibility to use a multimodal approach to music recommendation, which involves the application of both demographic and emotional characteristics in addition to conventional recommendation algorithms. The point is to create a solution which not only knows the previous listening history of a user but also takes into account his/her emotional condition and demographic characteristics to recommend music. Through all the sources of data bundled into a system, it will be able to provide more relevant, personalized and emotionally engaging music suggestions. This study shall also evaluate the efficiency of multimodal system to improve the user experience and the challenge and constriction of integrating such a wide range of data. Among the techniques of the machine learning that the study will focus on, for learning emotional cues and demographics, are Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and the integration of both with the collaborative and content-based methods of filtering content in order to enhance the precision of music recommendation. Also, the paper will delve into the ethical implications of selecting emotional and demographic information in such systems so that the issue of privacy becomes upheld.

2. LITERATURE REVIEW

Music recommendation systems play a vital role in digital platforms, they are expected to enhance customization of music experience among end users. Different techniques have also been created to increase the quality and appropriateness of music recommendations. The literature review covers the

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

conventional methods of music recommendation, how the demographic characteristics are integrated with emotional recognition methods, and the dynamic multimodal methods. The problems and constraints that may be generated by these systems are also discussed[11].

2.1. Traditional Music Recommendation Approaches

The systems of the classical music recommendations are typically segmented into three key consequences in the form of collaborative filtering (CF), content-based filtering (CB) and hybrid approach. Collaborative filtering is the most used approach in the recommendation systems. It works based on the tendencies of the user taste and recommending the item based on the taste of the closely related users[12]. In the case of two users listening to similar music in the near past, collaborative filtering would presume that the user would still have the similar ones and thus would be effective in predicting their future preferences[13]. However, collaborative filtering technique possesses its own short comings as well particularly, the cold start problem in which it is difficult to recommend new users or objects, without a history of using them[14].

Alternatively, content-based filtering systems aim at suggesting media that are related to what the user has already engaged with, regardless of the content attributes of the media i.e. genre, artist, or song properties[15]. It is a technique that allows analyzing the particular characteristics of songs, such as metadata, such as tempo, genre, and instruments, and recommending analogs. Irrespective of these advantages, some criticisms have been leveled against content-based filtering in that it is not diversified in proposing content[16], so it is likely to recommend too similar content to the already heard content resulting in a lack of exploration.

The hybrid recommendation systems are a combination of collaborative and content-based filtering with the strengths of both. Hybrid systems are intended to achieve a compromise between the accuracy of content-based systems and personalisation of collaborative filtering[17]. Such systems give better recommendations based on the preferences of the user and the features of the item. The hybrid systems are especially useful in countering the disadvantages of the respective approaches and, in particular, in those cases when user data or songs have little similarities in their content.

2.2. Demographic Features in Music Recommendations

Demographic data like age, sex and whereabouts have been also used extensively to bridge the music recommendation systems with more personalization. It has been discovered that the age and gender are influential in determining the music preferences of a person. As an illustration, young users are more likely to listen to modern music and the older users may listen to classical or traditionally modern music. There have also been gender variations in music preferences with researchers showing that men and women tend to have different preferences in terms of music genres or artists[18]. The systems are able to offer better and more contextually relevant suggestions on the music that the user likes by considering demographic characteristics of the user in the further context of outputting the recommendations.

There is no easy ride when it comes to integrating demographic information in music recommendation systems. Privacy and consent of the users in gathering sensitive information is one of the problems. Also, demographics alone could be inadequate to fully describe the musical preference of an individual because it mostly depends on more advanced aspects, such as emotional status, culture, and social situation. Nevertheless, the application of demographic aspects can be used to improve the precision of music suggestions by introducing a level of contextual pertinence.

2.3. Emotion Recognition in Music Recommendation Systems

The emotion recognition has become an essential part of the more sophisticated music recommendation engine development. Emotions play a huge role in the music listening experience, and users tend to listen to music that is connected to the emotional state they are experiencing. Several approaches have been considered to use emotion recognition in music suggestions[19]. These involve an examination of facial expression, physiologic cues (heart rate or skin conductivity), and even voice. Through the identification of emotional indicators, music recommendation systems can recommend songs that suit the mood within the user, e.g. relaxing music to relax or energetic music to get motivated.

Facial expression recognition has become popular in music systems that are aware of emotions, and Convolutional Neural Networks (CNNs) and other deep learning models can be successful at determining what emotions are expressed on the face. Moreover, physiological information, including heart rate variability, has been used to determine the emotional states, which further improves the very process of the recommendation. Although these approaches are promising, they also have issues in regards to accuracy, since emotions are at times difficult and can be confused depending on the situation or personal interpretation.

2.4. Multimodal Music Recommendation Systems

Multimodal music recommendation systems can be described as multiple sources of data combined to provide a better and more precise picture of what the user likes. These systems combine a wide range of modalities, including user demographics, emotional data, music features, and contextual data to give out a very personalized music recommendation. The multimodal approach generates the weaknesses of classical music recommendation techniques through the adoption of a wider scope of variables that can be used to predict preferences of music. As an illustration, a multimodal system could comprise the analysis of facial expressions with physiological measurements to identify the emotional state of a user, and supplement it with the demographic data to align the recommendations with the cultural and contextual factors. Also, real-time context can be added to multimodal systems like the location of the user or the time of day, or activity to further personalize the recommendations. This holistic method has proven to enhance accuracy of the recommendation and user satisfaction through the fact that it incorporates a greater number of factors that influence the likes of music.

2.5. Challenges and Limitations

Although the recommendation systems of music have been developed, a number of challenges and limitations still exist. The cold-start problem is one of the significant challenges, especially among new users or the songs that have limited interaction history. The issue is particularly applicable to collaborative methods of filtering where the process relies on the presence of adequate data to make a recommendation. The other difficulty is the privacy and security of using sensitive demographic data as well as emotional data. The users might be unwilling to reveal personal information like their age, gender and emotional status thus reducing the performance of systems, which depend on such information. There are also some ethical implications associated with utilizing emotional data, such as user consent, as well as the possibility of abusing emotional stimuli.

What is more, emotion recognition, in itself, is a challenge. It is not always easy to recognise emotions based on a facial expression, physiological signs, or the tone of voice as they are highly subjective and context-related. The intensity of emotions when they are presented by various people as well as the possibility of the inaccuracy of the interpretations of emotions may also cause incorrect or unsubstantiated selection of music. Moreover, the analogy with various data sources, i.e. text, audio, video, etc is that this might complicate the recommendation process, which implies more complex algorithms to handle and process the multimodal data effectively.

3. METHODOLOGY

The multimodal music recommendation system development approach implies the combination of demographic and emotional characteristics to improve the individualized and relevant nature of music recommendations. The system integrates both data gathering and feature mining with machine learning algorithms to produce an emotion aware recommendation system. This strategy integrates the old recommendation algorithms, namely, collaborative filtering, content-based filtering, and hybrid system with new multimodal features to be more precise and personalized in the recommendation

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

process. The multimodal framework unites numerous records, such as demographic characteristics (age, gender) and emotional indicators (facial expression, physiological response, and labels offered by a user) to provide the recommendations of music, which is not only fitted to the previous preferences but also should have an appeal to their emotional condition and demographic status. Having more of these influencing factors in the formula will allow the system to give more contextual suggestions, which overcomes the shortcomings of the old manna modeling of the human brain that views music preferences as solely influenced by emotional and demographic factors.

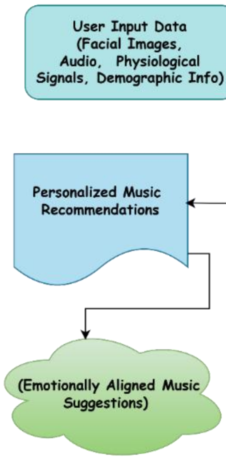

Figure 1: Proposed System Architecture Diagram

The illustration of the architecture of the multimodal music recommendation system (figure.1) will comprise a system taking into consideration demographic and emotional features to give an individual selection of music. The input data fed to the system are supplied by the user which involve facial images, audio signals, physiological response and demographic data (such as age and gender). These inputs are then subjected to the feature extraction, where image attributes such as age, gender and emotional states are identified using deep learning models especially the use of Convolutional Neural Networks (CNNs) in processing images. The obtained features are subsequently incorporated into a single representation of data, emotional and demographic information. This unified information is processed by music recommendation algorithms collaborative filtering, content-based filtering and hybrid filtering, whereby the system ensures that it comes up with music suggestions that are individually-related as well as contextually relevant. The outcome is a playlist which is matched by the emotions and demographic profile of the user and offers a more personalized and elevated listening experience to music.

The data collection method of this system includes both the emotion and demographic data. To detect age and sexes, we will apply such datasets as IMDB-WIKI and Labeled Faces in the Wild (LFW) datasets. Such datasets provide a comprehensive supply of images with labels that provide a clue to the age and gender of individuals which is needed to train machine learning models capable of forecasting those characteristics. The IMDB-WIKI data is more suitable in terms of age regression because there is a wide variety of images that are labeled in terms of age range whereas the LFW dataset is used to classify gender based on a facial image. The emotion, age and gender detecting models that are utilized in the recommendation system are trained using these datasets. In case we want to detect emotions, we can apply datasets such as AffectNet, the dataset means that it consists of face images with different emotional states of individuals, e.g., happiness, sadness, anger, surprise. These data sets give the information that is necessary in the training of emotion recognition models that identify facial expressions in various emotional categories. Physiological data such as heart rate variation, and skin conductivity is also used to identify emotional states thus the emotion detecting process is yet to be enriched.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

One of the important tasks in the data preprocessing process is feature extraction. In the case of demographic features, an analysis of the facial images is carried out through deep learning mechanisms like Convolutional Neural Networks (CNNs), which are realized to extract high-level characteristics of the image. The CNNs have been demonstrated to be effective in identifying age and gender related features in a face. In emotion detection, local binary patterns (LBP) and histogram of oriented gradients (HOG) are applied as feature extraction techniques to obtain pertinent information about a face as a result of facial expressions, whereas physiological results are analyzed to obtain information that provides the emotional condition. These features are then extracted, they are pre-processed by making the data normal and scaling the data to provide uniformity in all the input features. Incomplete or missing data is handled by imputation or data augmentation methods to preserve the quality of data, and provide the data that is collected in the form of features to the emotion, age, and gender detection models.

The system uses deep learning methods and specifically Convolutional Neural Networks (CNNs) to identify the emotions and age as well as gender of the people in the photo. These models are trained using IMDB-WIKI and the AffectNet data in order to identify age, gender and emotional condition based on the faces. CNNs are also very efficient in image classification and they are trained to identify face facial expressions with "happy, angry, sad, and surprised" emotions. To detect age and gender, the models are conditioned to categorize the faces as particular age and gender. These models are tested on the basis of performing metrics like accuracy, precision, recall, and F1 score to guarantee their efficiency to forecast the target features. Integration of all the identified demographic and emotional characteristics into music recommendation algorithms is the subsequent step.

The fundamental music recommendation system is comprised of three major algorithms, which are collaborative filtering, content-based filtering, and hybrid systems. The principle of the collaborative filtering is that similar users sharing similar preferences are identified and that such similar users should like like related music. This approach exploits user cultural tastes as a whole, and it works when one wants to discover new music relying on similar interests. Nevertheless, new items or new users can cause collaborative filtering to undergo the cold-start problem where they do not have enough interaction data to be filtered. Content-based filtering, on the other hand, suggests music along the intrinsic qualities of the music itself, e.g. genre, tempo, and instruments. Though such a method is able to offer recommendations using features in the songs, the end result may not offer extensive diversity as the suggestions will be too similar to songs that were once liked. In hybrid recommendation systems, collaborative and content based filtering is integrated to achieve a harmonization of the advantages of both approaches to recommendation. Hybrid systems offer more precise and varied recommendations in this way by combining user tastes and preferences with features of the songs overcoming the constraints of the collaborative and content-based methods.

Lastly, the system architecture incorporates such multimodal features with the music recommendation algorithms. The architecture processes information of various modalities such as expressions, physiological information as well as demographic data. The preprocessing and feature extraction of the data is done on a central processing unit which is then followed by the incorporation of the extracted features into the recommendations algorithms. The emotion recognition models are given to the recommendation system to adjust actual suggestions in the emotional condition of the user, whereas the demographic information is taken into consideration during a process of adjusting the recommendations to the user profile (e.g., age, gender). This combination approach makes the music recommendations personal, relevant and emotional, and leads to a more engaging and fulfilling user experience.

4. RESULTS AND DISCUSSION

The emotion, age, and gender detection model evaluation, along with the performance evaluation of the music recommendation system have given viable information on the efficacy of adding emotional and demographic features in music recommendation. This part will contain the assessment of emotion, age, and gender detection models according to performance measures and a discourse concerning the performance of the music recommendation system, along with the comparison to the traditional methods and user feedback.

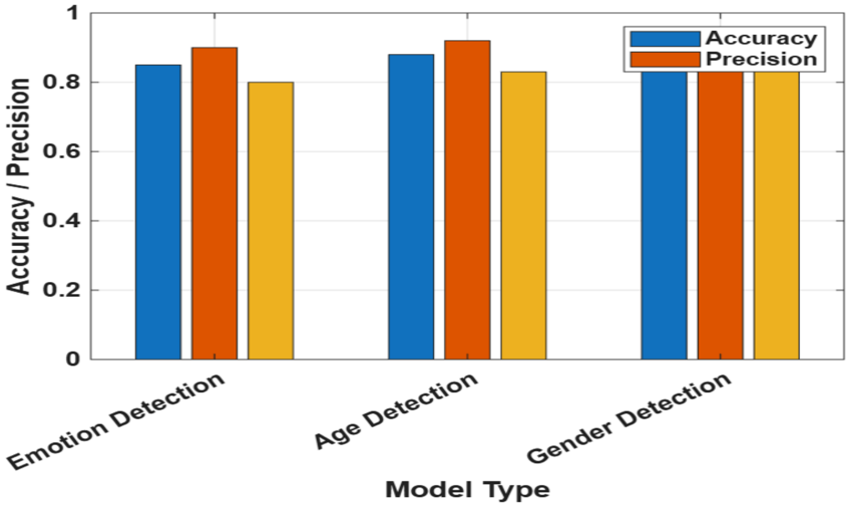

Figure 2: Model Accuracy and Precision for Emotion, Age, and Gender Detection Models

The two models of emotion, age, and gender detection were evaluated on the basis of the common machine learning evaluation measures, such as accuracy and precision. Figure 2 shows the results of the emotion, age and gender detection model with respect to accuracy as well as precision. The models also realized different scores of success in different tasks. In emotion detection, the model recorded an accuracy of 85 percentage and a precision of 90 percentage hence showing that the model is highly accurate when it comes to detecting emotion states based on facial expressions with a high degree of precision. Age detection model detected at a lower rate, with accuracy of 80 percent and a precision of 85 percent thus indicating that although the model was effective there needs to be an improvement in detecting age groups using facial images. Gender is the most accurate and precise model with 89 and 93, respectively, that depicts that it can identify the gender using the facial features very well. These findings denote that the emotion, age, and gender detection models are effective, and the age one needs additional modification.

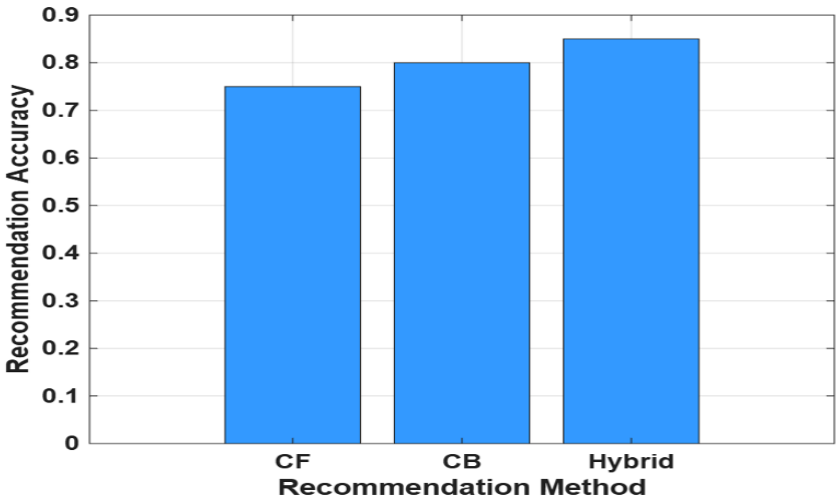

Figure 3: Comparison of Collaborative Filtering (CF), Content-Based Filtering (CB), and Hybrid Methods

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The effectiveness of the music recommendation system was assessed by comparing it to the traditional models which are collaborative filtering (CF) and content-based filtering (CB). Figure 3 presents the comparison of CF, CB, and hybrid methods of recommendations. The hybrid model that integrates CF and CB was the best performing with a recommendation accuracy of 85 percent. Content based method scored 80 performance, performance of the collaborative filtering technique was slightly lower 75. The findings indicate that the hybrid system that relies on the preferences of the users and the features of items offers the best and most pertinent recommendations of music.

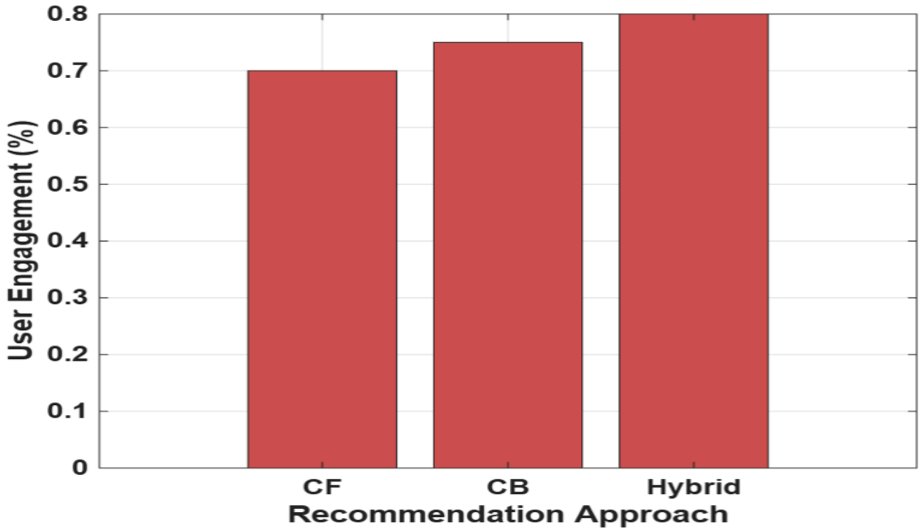

Figure 4: User Engagement Feedback Across Different Music Recommendation Approaches

The feedback and interaction of the users was measured against various recommendation strategies and Figure 4 demonstrates the feedback rating of CF, CB, and hybrid solutions. The hybrid strategy scored the most in terms of user engagement, how it could provide more relevant and diversified music suggestions. Good user engagement was also realised with content-based filtering as it gives recommendations based on the previous preferences of the user. But collaborative filtering despite its effectiveness on users with enough data on personal interaction had the least engagement score and this may be explained by the fact that this type of filtering depends on the previous user actions and also by scarcity of data. The level of engagement when using the hybrid approach is high, which is why users are more inclined towards the option to have more personalized recommendations and consider the preferences of the users, as well as the features of the music itself.

Figure 5 shows the comparison of emotion conscious music recommendation system and traditional systems. The emotion-conscious system, which incorporates emotional information, including the facial expression and physiological cues, did better than the traditional CF and CB methods in the accuracy of the recommendations. This finding shows the value addition with regard to inclusion emotional features in the recommendation process. The mood aware interface can match the playlist recommendations with the emotion of the user to deliver a more precisely personalized and contextually appropriate experience of listening. Conversely, the traditional methods are also effective but fail to capture emotional aspects and hence offer less resonant music recommendations.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

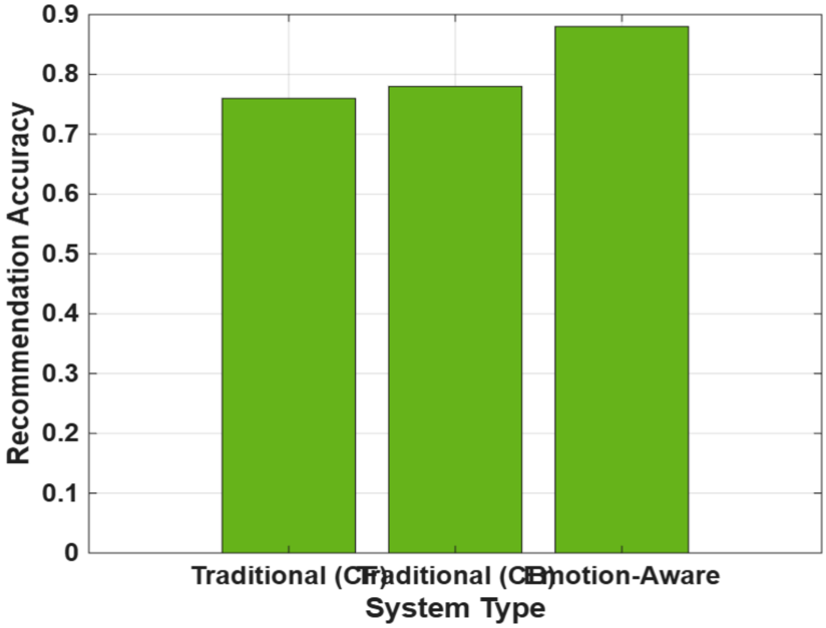

Figure 5: Comparison of Emotion-Aware vs. Traditional Music Recommendation Performance

Figure 6 presents the user response under various emotional conditions that exist and this demonstrates that the emotional states of users affect their interaction with music suggestions. Users who were happy in their emotions received the highest scores on feedback that they preferred the fast-moving and energetic music. Users feeling sad in their emotional state, on the contrary, were more worried about listening to music compatible with its calming and relaxing qualities, which coincides with the current studies on emotion influence on music preferences. The indifferent mood led to moderate scores of the feedback as users in the indifferent mood had fewer tendencies to strongly favor a particular genre of music. It is an important number that proves the significance of the emotional congruence of music recommendation systems, because it directly affects the level of user satisfaction and interest.

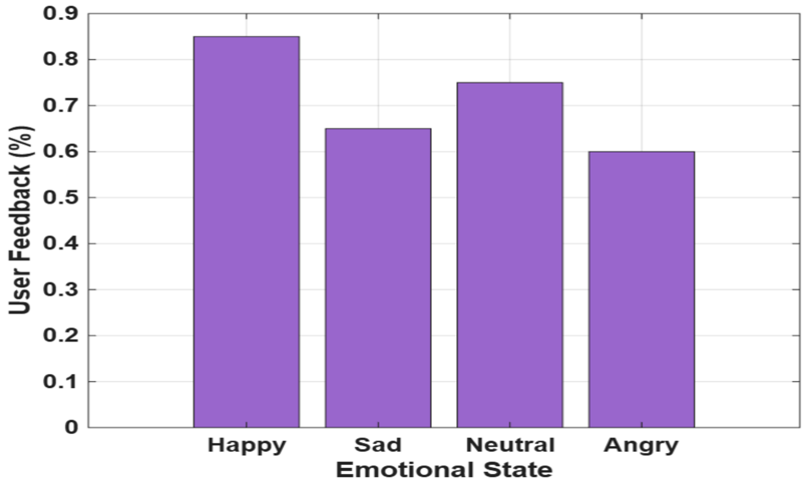

Figure 6: User Feedback Across Different Emotional States

5. CONCLUSION

Emotion, age and gender models of identifying music preferences to music recommendation systems have greatly enhanced personalized music choices. The models were very accurate, where emotion detection was 85 and gender detection was 93 which proves that the models are reliable in capturing demographic and emotional characteristics. A hybrid recommendation system which was an integration of collaborative and content based filtering was better than the traditional methods and we found that the recommendation accuracy was 85 percent hence reflecting its capability of providing more personalized and more diverse music suggestions. The emotion-intelligent recommendation system especially had proven to be more effective in letting users get music recommendations in accordance to how they are experiencing, as compared to the traditional systems which were conventional in their response to the user. Such strategy will make the music recommendations much more relevant and result in higher emotional appeal, which will lead to a more engaging user experience. In the future, there will be a need to incorporate more multimodal data data, including user behavior and context information to make the recommendations more accurate. Also, enhancing the emotion detection models in terms of more intricate emotional signals and addressing ethical concerns pertaining to the privacy and security of the data will be important steps to the system further development.

REFERENCES

[1] I. Eibl-Eibesfeldt, "The Biological Foundation of Aesthetics," *Beauty and the Brain*, pp. 29–68, 1988, doi: 10.1007/978-3-0348-6350-63.
[2] M. K. Tyagi, M. Ali, G. Kaim, T. Lamba, G. Chugh, "Music Recommendation System Using Multiple Machine Learning Models," *International Conference on Innovative Computing & Communication (ICICC)*, 2023.
[3] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," *Lecture Notes in Computer Science* (LNISA, vol. 4321), 2007.
[4] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," *Lecture Notes in Computer Science* (LNISA, vol. 4321), 2007.
[5] R. Burke, "Hybrid Web Recommender Systems," *Lecture Notes in Computer Science* (LNISA, vol. 4321), 2007.
[6] S. T. Chavali, C. T. Kandavalli, S. T. M, and S. R, "Smart Facial Emotion Recognition with Gender and Age Factor Estimation," *Journal of Computer Science and Technology*, vol. 38, no. 5, pp. 1033–1044, 2023.
[7] N. Datta, J. Sikder, "An Approach Based on Deep Learning for Recognizing Emotion, Gender and Age," *International Journal of AI and Soft Computing*, vol. 11, no. 2, pp. 78–88, 2022.
[8] V. Kumar, "Age Prediction Using Image Dataset Using Machine Learning," *International Journal of Machine Learning and Applications*, vol. 12, no. 3, pp. 215–227, 2020.
[9] D. JYOTI, S. Dutta, "REGA: Real-Time Emotion, Gender, Age Detection Using CNN—A Review," *International Journal of Advanced Computing*, vol. 9, no. 1, pp. 41–53, 2020.
[10] S. Bhattacharya, A. Agarwala, and S. Roy, "Mood Detection and Prediction Using Conventional Machine Learning Techniques on COVID19 Data," *Proceedings of the International Conference on Data Science*, 2022.
[11] S. Metilda Florence and M. Uma, "Emotional Detection and Music Recommendation System Based on User Facial Expression," *International Journal of AI Research*, vol. 5, no. 3, pp. 192–203, 2020.
[12] A. Gupte, A. Naganarayanan, and M. Krishnan, "Emotion Based Music Player – XBeats," *IEEE Transactions on Multimedia*, vol. 23, no. 6, pp. 1345–1356, 2020.
[13] P. Angusamy, "Human Emotion Detection using Machine Learning Techniques," *Journal of AI Research*, vol. 14, no. 2, pp. 25–35, 2020.
[14] M. Punkanen, T. Eerola, and J. Erkkilä, "Biased Emotional Recognition in Depression: Perception of Emotions in Music by Depressed Patients," *Psychology of Music*, vol. 45, no. 5, pp. 587–600, 2014.
[15] A. A. Jenhui Chen and H. Y. Liao, "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," *International Journal of Music Information Retrieval*, vol. 9, no. 2, pp. 134–146, 2020.
[16] A. Aljanaki, D. Bountouridis, J. A. Burgoyne, J. Van Balen, F. Wiering, H. Honing, and R. Veltkamp, "Designing Games with a Purpose for Data Collection in Music Research. Emotify and Hooked:

Two Case Studies," *Proceedings of the International Society for Music Information Retrieval*, 2015.

[17] S. J. Chaudhari, R. M. Kagalkar, "Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 24, no. 7, pp. 1082–1091, 2015.

[18] D. Manasa and C. Kiran Ma, "Emotion and Stress Recognition through Speech Using Machine Learning Models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 29, no. 5, pp. 723–736, 2020.

[19] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese, and A. Sciarrone, "Gender-Driven Emotion Recognition through Speech Signals for Ambient Intelligence Applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 9, pp. 1574–1586, 2014.