Song Recommendation System based on Mood Detection using Spotify's Web API

Anusha Bhowmick
School of Computing
SRM Institute of Science & Technology
Kattankulathur, Tamil Nadu, India
ab6681@srmist.edu.in

Kankshini Shamkuwar
School of Computing
SRM Institute of Science & Technology
Kattankulathur, Tamil Nadu, India
ks9624@srmist.edu.in

Dr. J D Dorathi Jayaseeli
School of Computing
SRM Institute of Science & Technology
Kattankulathur, Tamil Nadu, India
dorathij@srmist.edu.in

Dr. D Malathi
School of Computing
SRM Institute of Science & Technology
Kattankulathur, Tamil Nadu, India
malathid@srmist.edu.in

2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC) | 978-1-6654-5687-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/IIHC55949.2022.10060806

***Abstract*— When most existing music recommendation systems are content-based, collaborative, hybrid or context-based with a limited songs database, this study examines these systems and designs a new content-based recommendation system, with a vast number of songs. It takes into account the user's current emotion and then recommends songs based on their previous listening history with the help of song features retrieved using Spotify's Web API. The emotion recognition model gives an accuracy of approximately 67% and the recommendation system successfully generates a 20 song playlist based on the user's Spotify listening history and the current emotion detected.**

Keywords—content-based, clustering, K-Means, spotipy, Web API, CNN, OpenCV, Viola-Jones, emotion detection

I. INTRODUCTION

Music finds a presence in everyone's daily life. With the coming of streaming services, a huge variety of music is now available and accessible to people. Music differs according to place and culture. And people have different choices, likes and dislikes. This people's taste in music also varies. So, to find what type of music a person might like listening to and develop a recommendation system to help reach different artists, songs and genres, old and new, to users.Finding the relevance between various songs is a tedious task. It might be possible that one song which one user likes or belongs to a genre/type that one likes, may be disliked by other users or the same user over a period of time.

The system will determine the musical preferences of users based on the analysis of their interaction with the Spotify app during use, i.e. their recent top tracks information willbe queried using Spotify's web API. We also try to find out if the emotional context of the user can be used in a music recommendation system (RS) and to try to improve accuracy and provide better recommendations than existing models. Current recommendation systems like Spotify use a hybrid model (content-based + collaborative), whereas other research work have used direct emotion-based playlists as recommendations.We intend to bring a new approach such that new, old, popular, as well as not popular songs have an equal weight in being recommended while keeping in mind the user's general affinity to a type of songs (using recent top tracks) + the mood the user is currently in (emotional context).

II. STATE OF THE ART (LITERATURE SURVEY)

In our literature survey, we reviewed the following papers and have described the work done in brief:

In Music Recommendation System [1], a recommendation system was developed using the musical preferences of users. A large dataset of songs was taken and the content was analyzed. Recommendations were made on the basis of users making choices for the songs they liked, based on which they got recommended other songs.

This paper helped develop an understanding for the methodology to be followed and the system to be designed in this study.

Soleymani, Aljanaki, Wiering, Veltkamp et al. [2] made a RS based on psychological aspects of musical choices of the users. Thus this became another way of recommending other than usual genre-based recommendations. Auditory modulation features were analyzed to detect the attributes and regression was done. This research was unique as it performed better than the RS based on genres or the user-based RS based on the root-mean-square error, thus motivating us to choose factors other than genre for our recommendation system.

Ning, Li et al. [3] proposed an algorithm uses the user's contextual information to give recommendations. Uses content-based recommendation and logistic regression which give both user-based and item-based collaborative filtering recommendation results.

This research work was significant however the focus of the study is to create better recommendations that can help the user discover new music without taking popularity or charting positions into consideration.

In [4] by Mallegowda, Rane, Krishnan, Goyal, Hector et al., a hybrid (content+collaborative) approach was used. Convolutional Neural Network (CNN) was used to train song information. High dimensional data was reduced and a 5-song recommendation playlist was given.

Kathavate et al. [5] used a hybrid approach. This system also takes into consideration the user's contextual information, i.e., here whether they are doing work or dancing. Four different recommendation strategies were used and a playlist with top-N songs were recommended. Strategies are not described in detail and supervised kNN model is used with accuracy of 96%.

Stefanos, Georgios, Xarilaos, Savvas et al. [6] discuss various content-based recommendation systems like LIBRA,

which is a book recommendation system, CBMRS (music RS), PRES (home improvement RS), Cobra (blogs and RSS feeds filter). Describes advantages and disadvantages of Recommendation Systems and different techniques for user modeling and item analysis. However this paper analyzes recommender systems in general

Hirve, Jagdale, Banthia, Kalal, Pathak et al. [7] capture facial expression using camera. They used Viola-Jones algorithm for face detection and multiclass SVM for emotion detection.

Iyer, Pasad, Sankhe, Prajapati et al. [8] is about `EmoPlayer', is an application which recommends songs based on user's current mood.

Shin, Jang, Lee, Jang, Kim et al. [9] worked on a music service based on brain signals of the user. These signals were analyzed and then songs are recommended based on the user's emotions.

Florence, Uma et al. [10] contained emotion extraction, Audio extraction and Emotion-Audio extraction module. Used CK+ and HELEN datasets and for face detection use HAAR and HOG algorithms.

George, Suneesh, Sreelakshmi, Paul et al. [11] use the Convolutional Neural Network (CNN) model to classify 7 facial emotions. Contained three modules: Emotion, Music Classification and Recommendation Modules. The Emotion modules identifies the current mood of the user by taking facial image as input on which CNN is used. Music Classification Module classifies songs into 4 emotion classes based on song features. The Recommendation Module recommends songs to the by connecting their emotions to the mood type of the song.

Tan, Fan, Su, Zhang et al. [12] used an animated figure to show the emotion of the song using animated expressions. The emotion or mood type of the songs is found based on audio features. HAAR algorithm and OpenCV used for face detection.

Mahadik, Milgir, Jagan, Kavathekar, Patel et al. [13] studied existing systems, and used Keras algorithm (DL based technique) for up to five distinct facial emotions (Happy, Sad, Anger, Surprise and Neutral). The system uses a CNN built with the help of Keras. The backend is TensorFlow in Python & OpenCV is used for image processing for face detection.

Alrihaili, Alsaedi, Albalawi, Syed et al. [14] had pre-created 4 playlists to enhance the mood of the user. Viola-Jones algorithm and Principal Component Analysis techniques were used. The highest accuracy achieved for a mood is 98%.

III. MODULES OF WORK

There are three modules in our project, namely: Face Detection Module, Emotion Detection Module and Song Recommendation Module.

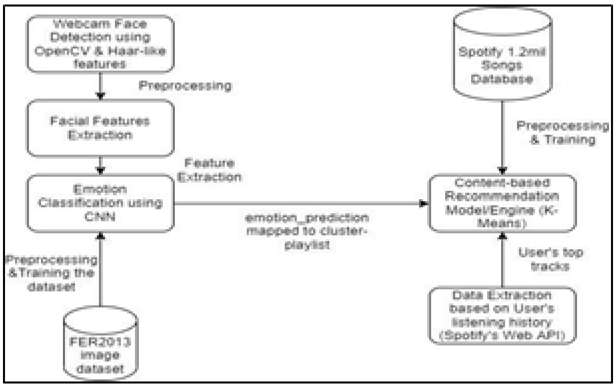


Fig. 1. Chart showing flow between modules of system

Fig. 2. Sample images from dataset

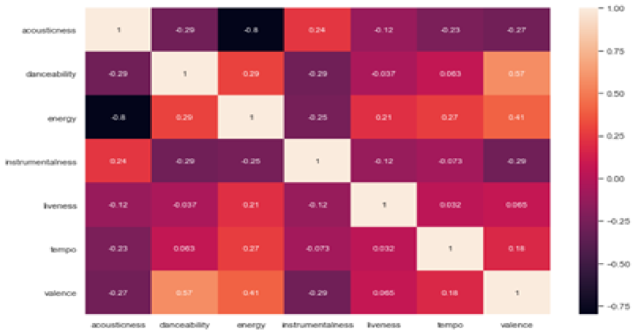


Fig. 3. Correlation map between song features

The Face Detection module looks for Haar-like features in the webcam input of a face and detects the face using Viola-Jones algorithm. OpenCV is the library used.

When the face is detected, the CNN facial expression detection model from the Emotion Detection Module is used to classify the facial input into an emotion class.

The emotion class detected is fed into the song recommendation engine. The user is asked to login to their Spotify account and the user's recent top tracks data is extracted using Spotify's web API. Top track data is fed into the K-means clustering model in the form of a mean vector generated from the song features, which then finally recommends nearest 10 songs from the most similar cluster. Additionally, 10 more songs are recommended from the cluster of songs corresponding to the emotion detected from facial expression.

We can consider this as a content+context based song recommendation system that uses K-Means clustering, CNN and Viola-Jones algorithm Machine Learning/Deep Learning Algorithms.

IV. MODULE EXPLANATION/DESCRIPTION

A. Emotion Detection Module

1) Getting Data: Our system uses the dataset FER-2013 which contains facial expression images with the dimensions 48*48 pixels, each image mapped to an emotion label (0:Angry, 1:Disgust, 2:Fear, 3:Happy, 4:Sad, 5:Surprise, 6:Neutral), i.e. 3 columns, Emotion, Pixels and Usage. Usage can befor training or testing.

2) Preparing Data: Preprocessing is done on the data to change it to the right format. Dataset is divided into X_train, X_test and y_train, y_test where the former is for pixels in a string format, and the latter is for emotion label (integer encoded labels). 7 emotion classes are used denoted by the variable num_classes. Data is converted to a 4d tensor for training, which has the following parameters: row_num, width, height, and channel.

3) Image Augmentation: This is done to improve the performance the model, and so that the model can generalize better. It is done before passing it to the model, using ImageDataGenetrator provided by Keras. Data generator is ideal when it comes to training a large amount of data. Rescaling is done which is dividing the pixel value by 255. A horizontal flip flips the image horizontally. fill_model fills the image if not available. Image is rotated using rotation_range in the range of 0– 90 degrees. On testing data, only rescaling is applied. Also, here the batch size is 64.

4) Build model and train: For the CNN model, Blocks are created using Conv2D layer, Batch- Normalization, Max- Pooling2D, Dropout, Flatten are used to create blocks and these blocks are then stacked together. Dense Layer is used for the end output. Model is compiled using Adam optimizer. Learning rate is kept as 0.001. Training takes 30 minutes for 1 epoch on an i3 Windows system. The model's weight is saved into a .h5 file and the architecture is converted and saved into JSON.

steps_per_epoch = TotalTrainingSamples / TrainingBatchSize

validation_steps = TotalvalidationSamples / ValidationBatchSize

B. Face Detection Module

1) Loading model: First the trained model architecture and weights are imported. Haar-cascade is used to find position of the face and after getting position the faces are cropped.

2) Preprocessing: OpenCV python library is used to read frames and for image processing & the label of emotion is returned in the form of emotion_prediction variable. Test images are rescaled by dividing them by 255. 3D matrices are converted into 4D tensors where (x,y,w,h) are the coordinates of the face detected in the input.

3) Adding Overlay: Overlay is added to output frame. Predicted emotion and confidence is displayed.

C. Clustering-based Song Recommendation Module

At first the user will get credentials from the Spotify's web dashboard (client id and client secret key). After getting credentials, the user's most listened to music will be displayed, from there those tracks their features will be taken into account (for ex: dance, tempo, mode, valence etc.) and then fed it to the recommender engine.

From the database, features will be selected which can be best for the recommendation engine, so this can be done by using heat map (visualization technique) (see Figure 4.2) thus which feature showed best relation with each other i.e. which showed more positive correlation in our case 4 features showed the best result. Hence once feature selection is done, we will feed it into the song features a separate data frame by dropping those values which are not required, thus song data frame is formed. The song data frame will undergo modeling (i.e. K- means clustering) and hence, accordingly, the clusters will be formed.

An "ear test" was done, where songs from each cluster were heard and judged as to which emotion class they would correspond best to, or which emotion class would they best serve as recommendations for. Accordingly 7 emotion classes were mapped to the 5 clusters created by our K-Means model. (See Figure 5.8)

As the facial expression of the user is detected, it is saved into an emotion_prediction label. Correspondingly, 10 songs, from the cluster playlist that emotion label belongs to, are recommended. Furthermore, a mean vector is generated from the song features of the user's top tracks, and the nearest 10 songs (nearest in terms of song feature similarity) are also recommended.

V. IMPLEMENTATION/RESULT

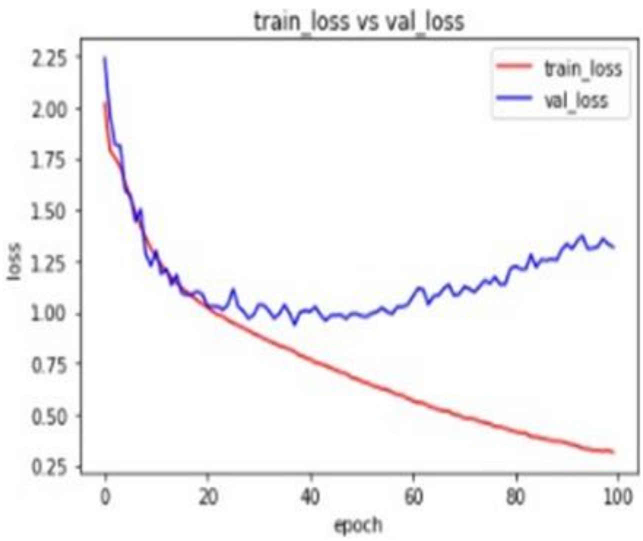


Fig. 4. training loss vs validation loss

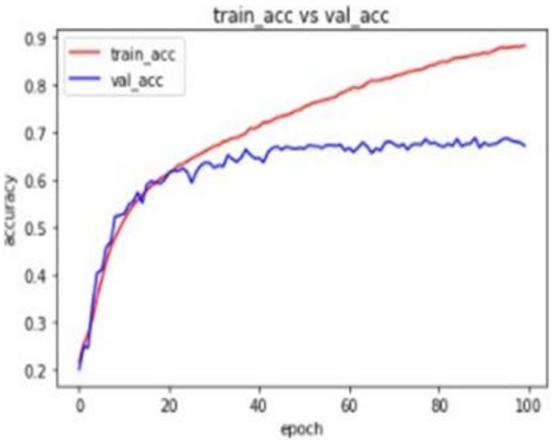
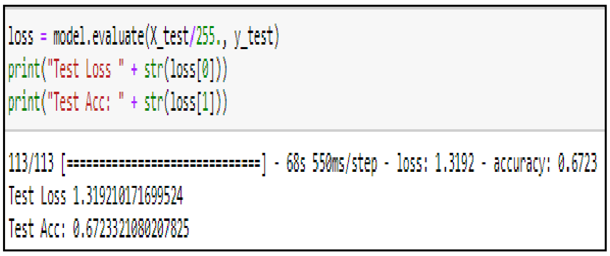


Fig. 5. training accuracy vs validation accuracy



```
loss = model.evaluate(X_test/255., y_test)
print("Test Loss " + str(loss[0]))
print("Test Acc: " + str(loss[1]))

113/113 [==============================] - 68s 550ms/step - loss: 1.3192 - accuracy: 0.6723
Test Loss 1.319210171699524
Test Acc: 0.6723321080207825
```

Fig. 6. Testing loss & Accuracy

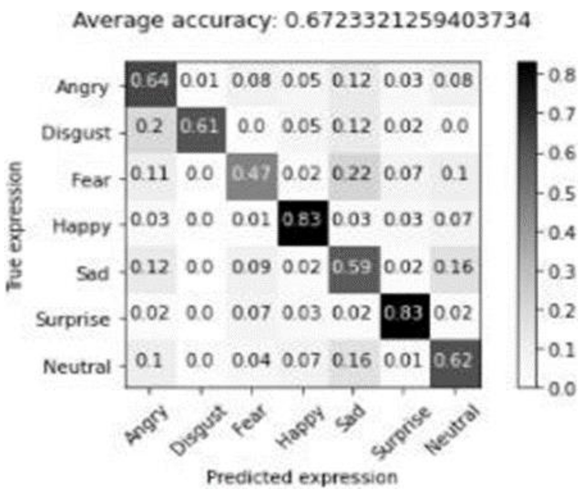


Fig. 7. Confusion matrix for CNN model for Emotion Detection

So, for this CNN model, the average accuracy is 0.6723.

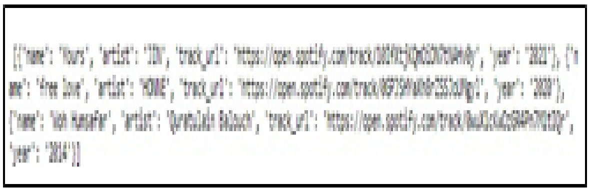


Fig. 8. A sample user's recent top tracks

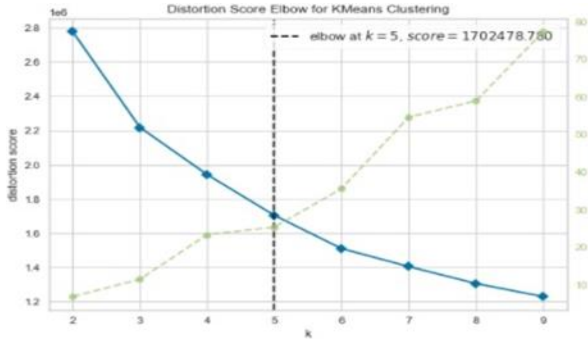


Fig. 9. Elbow method to determine ideal number of clusters for K-Means Clustering

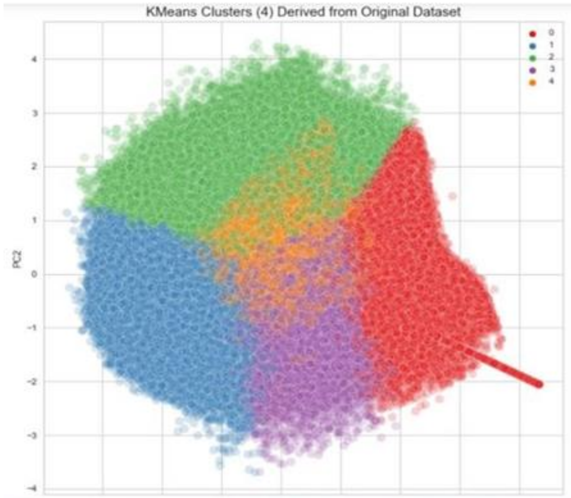
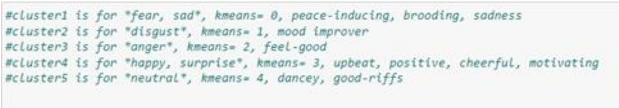


Fig. 10. K-Means clusters for k=5



```
#cluster1 is for "fear, sad", kmeans= 0, peace-inducing, brooding, sadness
#cluster2 is for "disgust", kmeans= 1, mood improver
#cluster3 is for "anger", kmeans= 2, feel-good
#cluster4 is for "happy, surprise", kmeans= 3, upbeat, positive, cheerful, motivating
#cluster5 is for "neutral", kmeans= 4, dancey, good-riffs
```

Fig. 11. Categorizing clusters into mood-based playlists based on 'ear-test'

	name	id	artists
370719	Memphis Pig	0stByt8dOEBWke4xd8AQ0w	['Martin Rivas']
1021911	JIGGA MAN	2EZWHyRiCDIhKUFGpfeQTy	['Sterling Hayes', 'KAMI']
293466	You and I	1KVOXBCNWzE7HbtbVaAeUQ	['Isaac Indiana']
306838	Earth Has No Sorrow	6lSfkgKcAtOYX5V71zlPTM	['Byron Cage', 'Purpose']
659008	ANGEL	3C7pJ7d2EA51JLExwpGPel	['GOT7']
953721	Stolen	0a7Amgw20ZTrqYZmfAZpci	['Nativ']
1034618	Black Milk	4SHKbhBzJ9Vt51NVckuMSG	['Animal Orchestra']
767730	Радуйся	6zsb7Y2xwPHHz2CzyEh4cN	['Alla Pugacheva']
122021	Casual Smile	7Ev7MjSuOD5NxLK51agqqK	['Steve Carlson']
299421	I H Y F I	4DydGptwyKVYCL6JQi24u8	['Mishal Moore & DJ Fonti Project']

Fig. 12. Recommendations generated based on user's top tracks

	id	name	album	album_id	artists
179549	29potP6CIMChqL09Dydkqc	Das Rheingold: Scene 2: Sanft schloss Schlaf d...	Wagner, R.: Das Rheingold	2SdxAhQrcynkYdJbT72onG	['Richard Wagner', 'Albert Dohmen', 'Ralf Luka...
195973	6MVn6Rc1tzCWoBsrx2uPSw	Der Jüngling und der Quelle, D.300 - Voice	Essential Classics: Lieder	3bZB5hJKFKw5bqy3z3nmwY	['Franz Schubert', 'Judith Raskin', 'George Sc...
396437	1GhoFBjrppVz3gFxhsEgSE	Quia ergo femina mortem instruxit	900 Years Hildegard von Bingen	6cYBI6Gjc1XmEzwUBgK9xo	['Hildegard von Bingen', 'Sequentia', 'Barbara...
402411	797q6jQVuujH4f6AMHnAsF	The Fisher Who Died in His Bed	A Retrospective 1974-1993	0o5xwOdM0mB0OiYgxoaiKO	['Figgy Duff']
912291	2NINEMXTShU4A3Mw4gvBvB	De sperme et d'eau fraiche	Knife + Heart (Original Motion Picture Soundtr...	2z7AL98OyCVQzochbAjxFu	['M83']
131162	5UCtPebshPlrqC83PF83hO	Duets for Storab: Urlar	Harrison Birtwistle: Refrains and Choruses	3LtBJqaryvBB2dsLjGhltb	['The Galliard Ensemble']
131572	4C3ScVvmJquRukTbl3TJlf	De profundis	Gregorian Chants	77Ey0aHKrYnruDFqXgVttV	['Traditional', 'Liturgical Text', 'Choir of M...
375579	0f7DqecbuMYpfm3n1g4Ymq	Pelimannit (The Fiddlers), Op. 1 (version for ...	Rautavaara, E.: Music for String Orchestra (Co...	2IC4khWDHaJqb3fnydkWAS	['Einojuhani Rautavaara', 'Ostrobothnian Chamb...
188504	5tvKDTy3TsmOqmxSMPRNY6	Solo	Too Marvelous for Words	6R5kMg58ZNGyG6zDs3klbg	['Domenico Sanna Trio']
657651	341IHxSsYveCysURwEycFe	500 Guns	Block Royalty	3J8roJbOyLfQx8m6dmMsIP	['Hollow Tip']

Fig. 13. Recommendations from cluster4 (kmeans=3) for emotion_prediction="happy"

By using VGG-16 or Resnet or by fine-tuning, accuracy can be further improved.

CONCLUSION

With millions of people listening to music all around the world on online streaming services, recommendation systems play an important role in improving a user's music taste, along with connecting one to the music from different places. At the same time, music recommendation systems help us find songs from different time periods. In this paper, our main focus was to build a recommendation system that makes use of Spotify's Web API and then recommends similar songs from the 1.2 million songs database, taking into account the emotional context of the user. Although there is scope in improving accuracy of emotion recognition model, facial emotion of the user is correctly detected and most of the recommended songs were suited as per the user's taste or their current emotion.

REFERENCES

[1] S. Shin, D. Jang, J. J. Lee, Sei-Jin Jang, and Ji-Hwan Kim, "MyMusic- Shuffler: Mood-based music recommendation with the practical usage of brainwave signals," IEEE International Conference on Consumer Electronics (ICCE), 2014.

[2] X. S. F. C. Tan, H. Fan, J. Su, and Zhang, "HeartPlayer: A Smart Music Player Involving Emotion Recognition, Expression and Recommendation," in *International Conference on Multimedia Modeling, Advances in Multimedia Modeling*, vol. 13. Springer, 2013.

[3] M. Soleymani, A. Aljanaki, F. Wiering, C. Remco, and Veltkamp, "Content-based music recommendation using underlying music preference structure," *IEEE International Conference on Multimedia and Expo (ICME)*, 2015.

[4] A. Mahadik, P. S. Milgir, V. V. B. Jagan, J. Kavathekar, and Patel, "Mood based Music Recommendation System," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 06, 2021.

[5] Y. Kodama, . S. Gayama, Y. Suzuki, S. Odagawa, T. Shioda, F. Matsushita, and T. Tabata, "A music recommendation system," *International Conference on Consumer Electronics*, 2005.

[6] A. Alrihaili, A. Alsaedi, K. Albalawi, and L. Syed, "Music Recommender System for Users Based on Emotion Detection through Facial Features," *12th International Conference on Developments in eSystems Engineering (DeSE)*, 2019.

[7] M. Mallegowda, T. S. S. Rane, N. K. Krishnan, E. Goyal, and H.G. Hector, "Music Recommendation Systems - Content Based Approach," *International Journal of Research*, pp. 2024–2029, 2013.

[8] M. S. M. Florence and Uma, "Emotional Detection and Music Recommendation System based on User Facial Expression," *3rd International Conference on Advances in Mechanical Engineering*, pp. 2020–2020.

[9] P. Iyer and P. Sankhe, "Emotion based mood enhancing music recommendation," *2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2017.

[10] S. Kathavate, "Music Recommendation System using Content and Collaborative Filtering Methods," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, pp. 2021–2021.

[11] H. Ning and Q. Li, "Personalized Music Recommendation Simulation Based on Improved Collaborative Filtering Algorithm," *Hindawi Com- plexity Volume 2020*, vol. 2020, pp. 2020–2020.

[12] A. Stefanos, D. Georgios, Z. Xarilaos, and K. Savvas, "Content-Based Recommendation Systems," 2008.

[13] R. Hirve, S. Jagdale, R. Banthia, H. Kalal, and & K R Pathak, "EmoPlEmoPlayer: An Emotion Based Music Player," *Imperial Journal of Interdisciplinary Research (IJIR*, vol. 2, no. 5, 2016.

[14] A. George, S. Suneesh, S. Sreelakshmi, and T. E. Paul, "Music Recom- mendation System Using CNN," *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, vol. 6, 2020.