Mood Based Music Recommendation for a Mall using Real-time Image

[1]Afzal Mukhtar, [2]Hritika Rahul Mehta, [3]Abirami S., [4]Sukeerthi Adi, [5]Dr. Kamatchi Priya L

[1,2,3,4,5] Department of Computer Science PES University Bangalore, India
Email: [1]afzalmukhtar985@gmail.com, [2]hritika2708@gmail.com, [3]abiramisouvami@gmail.com, [4]a.sukeerthi@gmail.com, [5]priyal@pesu.edu

Abstract

Human beings have the natural ability to guess the mood of a person just through their facial expression. This ability is highly valuable to be learnt by a computing device, like computers or robots. Music affects the human emotional core and their memories very deeply, thereby affecting our mood. Similarly, the background music played in malls affects a shopper’s behaviour. Customers who listen to music they like, are more likely to have a positive experience. Time becomes more pleasant even if it is spent waiting in line or waiting to speak to a customer service associate. The proposed system is built to give customers a better and more satisfactory experience, which would make them stay longer and purchase more. It also helps the workers to be in a better mood. All leading to an increase in sales in the mall. The system uses the latest crowd image and finds the 2 most common facial moods using a CNN. In the backend, the songs present in the database are classified based on audio and lyrical features. Finally, a playlist of songs is recommended based on a percentage mapping of moods found in facial features.

Keywords— Music Mood Classification, Mood Recognition, Music Mood, Lyrics Mood, SVM, XGBoost, Convolutional Neural Network, CatBoost, GeniusAPI, SpotifyAPI

I. INTRODUCTION

The Mood of a user is a complex emotional function of the human brain. Music can enhance the productivity of the listener and can help in the release of a hormone called serotonin which can put them in a better mood. This can increase their duration of focus and understanding capabilities. Based on research, people linger longer in places where they feel happy or calm, thereby a mood-based music system can make people linger longer in places. The system will calculate the average mood of the crowd of people entering a particular floor of the mall and suggest a playlist of songs to better their current mood. This can give the marketers a better chance to attract customers to their stores, even if they have been window-shopping. The longer the customers stay, the higher the chances of them buying something. Which in turn can increase the profit of the mall. Thereby mood plays a vital role in improving the chances of gaining higher profit or doing a certain work longer. The analysis of crowd emotion, and how it can affect the outcome of a particular situation has been a topic of great interest among researchers. Psychologically, music has a great impact on the human mind and affects each one differently, but music of certain kinds tends to uplift the human mood in various situations. Humans often listen to music based on their mood, and the type of music can vary from person to person, but the general effect music can have on people, regardless of their personal preference, ensures to improve the mood of the user. Psychologically, certain musical audio and lyrical features have similar effects on a gathering, irrespective of the preference of each individual. We try to bring these together to provide a better music playlist for a crowd setting.

II. LITERATURE SURVEY

Aya Hassouneh, et al. in [1] focussed on developing a real-time emotion recognition system for physically disabled people using facial landmarks and EEG signals. A multimodal human-machine interface (HMI) was developed to classify facial expression into six emotions, where Facial feature extraction was performed by capturing grayscale images of subjects' faces and using the Lucas Kande algorithm to place the virtual marker. CNN with throughput and Adam optimizer improved the performance. The authors used an LSTM network for emotion detection using EEG signals. Notwithstanding its advantage for Autism kids to perceive the emotions of others, this system can help physically disabled people too. It can further improve business results and gauge the emotive responses of a crowd. The disadvantages are that data collected was from a single geographical location (Kuwait) and many students could not contribute to the data because of conflicts in their schedule.

Peter Dunker, et al. review the common mood models for their flexibility and generate a new reference set for multi-modal mood classification in [2]. They have used two classification models – Gaussian Mixture Models and SVM. The prime challenge of this paper was to create a reference set for multimodal classification and for which a two-step approach was used, where they collected already tagged data which included images from Flickr of different moods and music across a range of genres. A personal review of the tags assigned for photos and music was performed. The paper considered here came up with an audit of different mood models and the determination of a universal dimension and classification-based model. The mood model permits a simple mapping of measurement and class model type and a multi- modular use for all occurrences in this publication. A generic characterization structure was introduced that can deal with pictures and music in an equal way and permits a stream- lining of various system parameters. These resulting average recognition rates appear to be substandard when contrasted with instances of the previously evaluated endeavours. We must emphasize the fact that a random approach achieved 25% recognition. This points to the order that the considered reference set is very diverse.

A. V. Iyer, et al. have come up with a novel Android application 'EmoPlayer' [3] which can utilize the device's front camera and take an image of the user, and recognise the face using the Viola-Jones Algorithm. It sends this image to the server which detects the emotion and sends it back to the application. It retrieves a playlist for the detected mood. The playlist is aimed to improve mood to a happier mood. The application is compatible with Android devices. There is a limitation on the emotions recognized by the application, which are Angry, Happy and Sad.

Junting Qi, et al. in [4] talk about real-time face detection which is scale-invariant and has a large receptive field. The authors talk about the recent face recognition technologies that are based on frontal faces and similar sizes, which are good if we ignore the background of the faces. YOLOFKP is designed for real-time face detection with dense and low-scaled images. This model is scale-invariant and with a large receptive field, which is achieved using convolutional layers to increase the context area of the selected region. Using hole convolution increases the extent of the observable world, and keeps kernel size constant. Thus the model is robust for small and dense faces.

Jun Yang, et al. have focused on improving the accuracy of the classification of facial expression recognition using a deep network in [5]. The initial image is normalized, then introduced to a signal to add Gaussian noise to the data. This helps in overcoming the noisy effect of the facial expression images and improves the overall classification accuracy and stability of the model. The feature extraction implements the use of a convolution denoising autoencoder. XGBoost is a Gradient Boosting algorithm that is applied on a Decision Tree, which performs efficient utilization of multi-threads in a CPU while improving the accuracy of the classification. The experimental results showed that when the model was used with superimposed noise, it performed with higher accuracy than it did without.

III.PROPOSED METHOD

The mood-based music system will identify the mood of ev- ery customer entering a floor of the mall from images captured on the CCTV camera. It will calculate the cumulative mood for the crowd of people and suggest a playlist according to the mood the admin wants the customers to be in and the ambience of the floor. The system is built to give customers a better and more satisfactory experience, which would make them stay longer, thereby increasing their chances of purchasing more products. It also helps the workers to be in a better mood to service the customer hence making a healthy and happy work environment. However, the system will not cater to the needs of every customer. The playlist needs to be updated manually. It needs constant images from the CCTV cameras for detecting mood. Mood detection is not possible with a mask on.

IV. IMPLEMENTATION DETAILS

A. *Face Emotion Recognition*

The dataset (FER-2013, Kaggle), was an imbalanced dataset that had a large variation in the number of images for each type of emotion. In this project, we have focused only on the five major moods that are usually seen in the crowd in a mall

– Happy, Sad, Angry, Surprised and Neutral. The significance of data imbalance reduced on choosing these five emotions

– –

– –

–

TABLE I: CNN ARCHITECHTURE

Layer	Feature Map	Size	Kernel Size
Input	1	48 x 48 x 1	-
Conv2D	32	48 x 48 x 2	3 x 3
Conv2D	64	48 x 48 x 64	3 x 3
Batch Normalisation			
MaxPooling	64	24 x 24 x 64	2 x 2
Dropout(0.25)			
Conv2D	128	24 x 24 x 128	3 x 3
Conv2D	256	22 x 22 x 256	3 x 3
Batch Normalisation			
MaxPooling	256	11 x 11 x 256	2 x 2
Dropout(0.25)			
Flatten			
Dense	-	1024	-
Dropout(0.50)			
Dense	-	5	-

We designed a Neural network with the following sequence of 2 Convolution, Batch Normalization, Max Pooling, and Dropout, which repeats for another set. The output from the last Max Pooling layer is flattened and then passed through a dense layer and final output layer *(Table I)*.

Each layer employs the use of the ReLu activation function, and the final output layer has a Softmax activation function.

B. *Audio Mood Classification*

The audio features are obtained using the Spotify API. There are two dataframes created from the information re- turned by the API call, one which stores information about the song metadata and the other which stores the features needed to supply to the model. No transformation of features is necessary other than to take the subset of audio features we consider important.

The Audio Mood Model is an XGBoost model which has parameters as follows – learning rate=0.4, max depth=12, subsample=0.6, colsample bytree=0.4, num classes=4, objective=multi:softprob, n estimators=550, gamma=0.3.

C. *Lyrics Mood Classification*

Lyrics data is gathered using the Genius API using the artist name and song title, then the data is removed of characters which hold very little or no meaning associated with it, and it is also translated to English language if the language is not the English, which is detected using the Detectlanguage API and GoogleTranslate API. This is then converted into a vector representation using the Spacy language model.

The Lyrics Mood model is an ensemble of three dif- ferent models, which include XGBoost, and two Deep Neural Networks. We got the best configuration of XG- Boost after performing a RandomSearch over a large number of parameters, where the results had the pa- rameters as follows – colsample bytree=0.5, max depth=9, min child weight=4, n estimators=140, nthread=4, objec- tive=multi:softprob, seed=27, and subsample=0.6.

The Deep Neural networks had a wide and a narrower configuration. Both the networks had an alternative fully connected layer and a dropout layer. The wider network's fully connected layer sequence is – 128, 256, 256, 32, 32, 4. The narrow network's fully connected layer sequence is – 64, 64, 128, 128, 64, 64, 4.

V. EVALUATION METRICS

To measure the performance of the model's ability to classify the data, we use two classification metrics. Since there is an imbalance in the data, we choose F1-score which is the harmonic mean of the precision and recall, and it accounts for the trade-off between them. The second metric used is Accuracy to measure the overall effectiveness of the model to accurately predict the class of the data.

VI. RESULTS

A. Face Emotion Recognition

For measuring our classification performance, we use two metrics, namely overall accuracy and F1-score. Our Face Emotion Recognition model performed with an average accuracy of 68%, which is comparatively bad to the other pre-existing model VGG-16, which gives an accuracy of around 71%. A ConvXGBoost model was made by taking the output from the first dense layer of VGG-16 and training XGBoost on that output to get the final prediction. The parameters used were – min child weight=5, n estimators=450, subsample=0.8, grow policy='lossguide', colsample bytree=0.4, tree method='gpu hist', max depth=8, objective='multi:softprob', num classes=5.

TABLE II ACCURACY OF FER MODELS

Model	**Overall Accuracy**
CNN	68%
VGG - 16	71%
ConvXGBoost (Using VGG-16)	72%

TABLE III: F1-SCORE OF FER MODELS

Model	**Happy**	**Sad**	**Angry**	**Surprised**	**Neutral**
CNN	0.81	0.54	0.58	0.83	0.61
VGG-16	0.85	0.58	0.63	0.83	0.64
ConvXGBoost	0.85	0.57	0.64	0.84	0.64

B. Audio Mood Classification

We tested multiple models for the classification of our audio mood, like SVM, XGBoost, ANN, and Denser-ANN models. The XGBoost model on average performs much better than the other models, thereby it being the model we chose to be used in the product.

TABLE IV: ACCURACY OF AUDIO MOOD MODELS

Model	**Overall Accuracy**
SVM	82%
XGBoost	86%
CNN	83%
Dense-ANN	85%
Deep-NN XGBoost	83%

TABLE V: F1-SCORE OF AUDIO MOOD MODELS

Model	**Happy**	**Energetic**	**Calm**	**Sad**
SVM	0.69	0.84	0.89	0.81
XGBoost	0.72	0.81	0.96	0.88
CNN	0.68	0.79	0.96	0.86
Dense-ANN	0.74	0.72	0.97	0.91
Deep-NN XGBoost	0.69	0.75	0.97	0.88

C. Lyrics Mood Classification

Lyrics classification was performed with multiple models, and a majority vote ensemble technique is used to get the best of each model *(Table VI & VII)*.

Choosing an ensemble model helped us to generalize the problem better as each model was trained on a different random sample. And the few of the best ensemble models are represented in the following tables.

The Ensemble – 2 performs better on average compared to the rest of the models, giving around 72% of overall accuracy *(Table VIII & IX)*.

D. User Interface of Web Application

The user needs to first sign-up or login to use the applica- tion. The login/sign-up UI is represented in the Fig 1.

TABLE VI: ACCURACY OF LYRICS MOOD MODELS

Model	Overall Accuracy
DeepNN-1	67%
DeepNN-1	73%
CatBoost	62%
SVM	60%
XGBoost-1	60%
XGBoost-2	60%

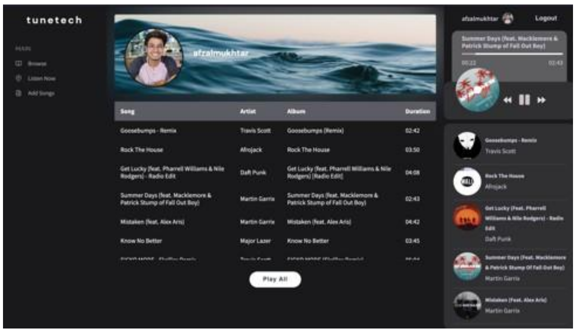
TABLE VII: F1-SCORE OF LYRICS MOOD MODELS

Model	***Happy***	***Angry***	***Relaxed***	***Sad***
DeepNN-1	*0.68*	*0.76*	*0.62*	*0.61*
DeepNN-2	*0.77*	*0.82*	*0.67*	*0.68*
CatBoost	*0.65*	*0.71*	*0.54*	*0.57*
SVM	*0.65*	*0.73*	*0.56*	*0.44*
XGBoost-1	*0.63*	*0.71*	*0.55*	*0.53*
XGBoost-2	*0.62*	*0.72*	*0.53*	*0.51*

TABLE VIII: ACCURACY OF ENSEMBLE MODELS

Model	**Overall Accuracy**
Ensemble-1 (DeepNN-1, DeepNN-2)	71%
Ensemble-2 (DeepNN-1, DeepNN-2, XGBoost-2)	73%
Ensemble-3 (DeepNN-1, DeepNN-2, CatBoost)	69%
Ensemble-4 (DeepNN-1, DeepNN-2, CatBoost, XGBoost-2)	68%

Once the user logs into the application, they can access the application. The music player dashboard provides the users with numerous features to browse, generate recommendations and play songs, to name a few.

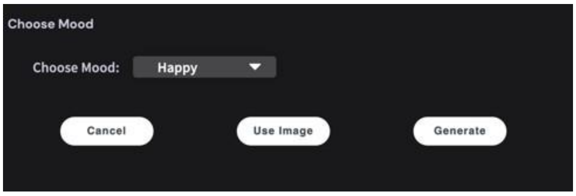


Song	Artist	Album	Duration
Goosebumps - Remix	Travis Scott	Goosebumps (Remix)	02:42
Rock The House	Afrojack	Rock The House	03:50
Get Lucky (feat. Pharrell Williams & Nile Rodgers) - Radio Edit	Daft Punk	Get Lucky (feat. Pharrell Williams & Nile Rodgers) [Radio Edit]	04:08
Summer Days (feat. Macklemore & Patrick Stump of Fall Out Boy)	Martin Garrix	Summer Days (feat. Macklemore & Patrick Stump of Fall Out Boy)	02:43
Mistaken (feat. Alex Aris)	Martin Garrix	Mistaken (feat. Alex Aris)	04:42
Know No Better	Major Lazer	Know No Better	03:45

tunetech
MAIN
Browse
Listen Now
Add Songs
afzalmukhtar
afzalmukhtar
Logout
Summer Days (feat. Macklemore & Patrick Stump of Fall Out Boy)
00:22
02:43
Play All
Goosebumps - Remix
Travis Scott
Rock The House
Afrojack
Get Lucky (Feat. Pharrell Williams & Nile Rodgers) - Radio Edit
Daft Punk
Summer Days (Feat. Macklemore & Patrick Stump Of Fall Out Boy)
Martin Garrix
Mistaken (Feat. Alex Aris)
Martin Garrix

Fig. 2. Music Player Dashboard

Users can generate playlist using the crowd image that is stored in their database, or by manually selecting the mood of songs they want. They can regenerate playlist if they wish to get different songs.



Choose Mood

Choose Mood: Happy

Fig. 3. Generate Playlist

TABLE IX: F1-SCORE OF ENSEMBLE MOOD MODELS

Model	***Happy***	***Angry***	***Relaxed***	***Sad***
Ensemble-1	*0.74*	*0.79*	*0.66*	*0.64*
Ensemble-2	*0.73*	*0.81*	*0.66*	*0.68*
Ensemble-3	*0.73*	*0.79*	*0.62*	*0.61*
Ensemble-4	*0.72*	*0.77*	*0.61*	*0.62*

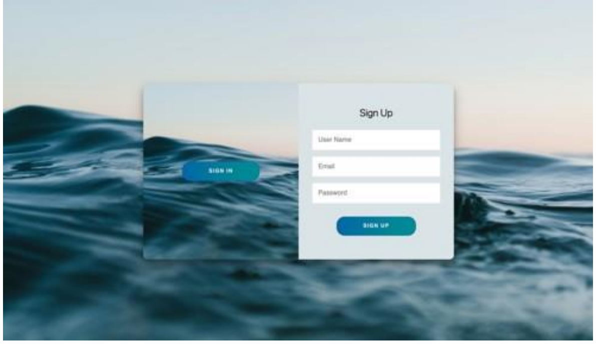
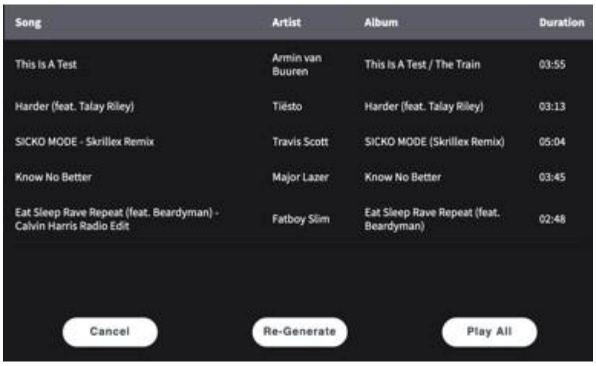


Fig. 1. User Login/Signup Page



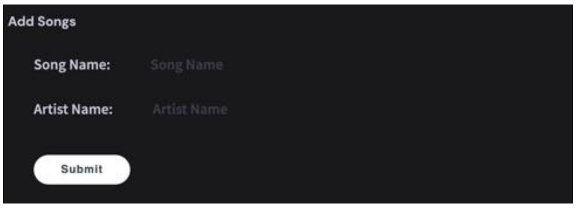
Song	Artist	Album	Duration
This Is A Test	Armin van Buuren	This Is A Test / The Train	03:55
Harder (feat. Talay Riley)	Tiësto	Harder (feat. Talay Riley)	03:13
SICKO MODE - Skrillex Remix	Travis Scott	SICKO MODE (Skrillex Remix)	05:04
Know No Better	Major Lazer	Know No Better	03:45
Eat Sleep Rave Repeat (feat. Beardyman) - Calvin Harris Radio Edit	Fatboy Slim	Eat Sleep Rave Repeat (feat. Beardyman)	02:48

Fig. 4. Example Recommended Playlist

Every user has to upload their own songs, for which the meta data and lyrics is extracted using Spotify and Genius APIs. These data after preprocessing, predicts the mood of the song and stores it in the database

VII. CONCLUSIONS

Listening to music can entertain us, and there is proof backed by research that it can make us healthier. The powerful effects of music range from improving cognitive performance, reducing stress, eating less and improving memory.



Add Songs

Song Name: Song Name

Artist Name: Artist Name

Fig. 5. Add Songs

Our project led us to focus on the effect music has on the mood of a person. It has been proven that music can make us happier. We try to recognise the current facial mood of users and improve it to make them happier Our product is a website that mall owners can sign up for and use to enhance the mood of their customers and thereby increase sales. We designed the website keeping in mind the functional and security aspects of our project.

In the future, we can focus on developing an application that will be a more portable form to use the product in other areas of application like individual usage. We can add features to keep in mind the previous listening preferences of users while giving suggestions of songs. A few more moods can be taken into account while classifying the facial emotions if we can get access to the required data. The data pre-processing step for lyrics classification is slow, we can try to improve the speed, and adding a feature to upload song data from the user, if the song is not available on Spotify or Genius.

VIII. ACKNOWLEDGMENT

We would like to express our gratitude to Professor Ka- matchi Priya, Department of Computer Science and Engi- neering, PES University, for her continuous guidance, assis-tance, and encouragement throughout the development of this Project. We are grateful to the Capstone Project Coordinator, Dr. Sarasvathi V, Associate Professor, for organizing, manag- ing, and helping with the entire process. We take this opportu- nity to thank Dr. Sandesh B J, Chairperson, Professor, Depart- ment of Computer Science and Engineering, PES University, for all the knowledge and support we have received from the department. We would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help. We are deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro-Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to us various opportunities and enlightenment every step of the way. Finally, this Capstone Project could not have been completed without the continual support and encouragement We have received from our family and friends.

REFERENCES

1. A. Hassouneh, A. M. Mutawa and M. Murugappan, ”Development of a Real-

Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods," *Informatics in Medicine Unlocked,* vol. 20, 2020.

2. P. Dunker, S. Nowak, A. Begau and C. Lanz, "Content-based Mood Classification for Photos and Music," in *1st ACM international confer- ence on Multimedia information retrieval (MIR '08),* New York, 2008.
3. A. V. Iyer, V. Pasad, K. Prajapati and S. R. Sankhe, "Emotion Based Mood Enhancing Music Recommendation," in *2nd IEEE International Conference On Recent Trends in Electronics Information & Communi- cation Technology (RTEICT),* Bangalore, 2017.
4. J. Qi, C. Wang, L. Cheng, S. Jiang, X. Zhang and H. Jing, "YOLOFKP: Dense Face Detection Based on YOLOv3 Key Point Network," in *Inter- national Conference on Computing and Pattern Recognition*, Xiamen, 2020.
5. J. Yang, D. Zhang, Z. Pan, D. Liu and J. Chen, "Facial Expression Recognition Based on Convolutional Denoising Autoencoder and XG- Boost," in *IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC 2019),* 2019.
6. H. Drieberg, "The effect of background music on emotional processing : evaluation using a dot probe paradigm," 2013. [Online]. Available: https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1097&context=theses hons. [Accessed 2020].
7. M. Khan and A. Ajmal, "Effect of Classical and Pop Music on Mood and Performance," *International Journal of Scientific and Research Publications*, vol. 7, no. 12, pp. 905-911, 2017.
8. S. Swaminathan and E. G. Schellenberg, "Current Emotion Research in Music Psychology," *Emotion Review*, vol. 7, no. 2, pp. 189-197, 2015.
9. Z. Ruifang, J. Tianyi and D. Feng, "Lightweight face detection network improved based on YOLO target detection algorithm," ACM, 2020.
10. S. Thongsuwan, S. Jaiyen, A. Padcharoen and P. Agarwal, "ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost," *Nuclear Engineering and Technology*, vol. 53, no. 2, pp. 522 - 531, 2020.
11. K. R. Tan, M. L. Villarino and C. Maderazo, "Automatic music mood recognition using Russell's twodimensional valence-arousal space from audio and lyrical data as classified using SVM and Na¨ıve Bayes," in *The International Conference on Information Technology and Digital Applications*, Manila, 2019.
12. L. Ren, W. Xian, H. Tang, Y. Jiang, H. Jia and J. Li, "Pedestrian and Face Detection with Low Resolution Based on Improved MTCNN," in *9th International Conference on Computing and Pattern Recognition (ICCPR 2020),* Xiamen, 2020.
13. Z. Ruifang, J. Tianyi and D. Feng, "Lightweight face detection network improved based on YOLO target detection algorithm," in *ACM ISBDAI conference (ISBDAI'20),* Johannesburg, 2020.