Moo-Sic

Mood-Based Personalized Song Recommender System

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

TEAM MEMBERS

Mudit Gupta
2020315
mudit20315@iiitd.ac.in

Srishti Jain
2020543
srishti20543@iiitd.ac.in

Siya Garg
2020577
siya20577@iiitd.ac.in

Sumit Soni
2020136
sumit20136@iiitd.ac.in

MOTIVATION

Songs are the best way to cheer someone up. However, choosing the appropriate song becomes tedious and annoying at that moment. Hence, we came up with classifying the user's mood and then suggesting a song based on their interests.

In this project, we aim to recommend songs to users by recognizing their moods using facial expressions while also considering the genres of their interest.

LITERATURE REVIEW

The paper on the *transfer learning approach for Face Recognition using Average Pooling and MobileNetV2* discusses improvising facial recognition technology. The research focuses on the implementation of 2 different facial recognition model classifiers: Average Pooling and MobileNetV2, and also draws a comparison between the results of both.
The paper concluded that MobileNetV2 has a higher accuracy rate compared to CNN with average pooling.

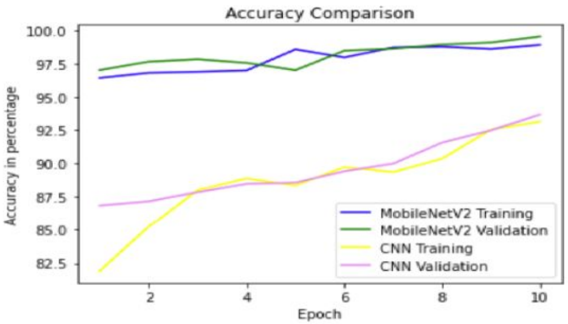
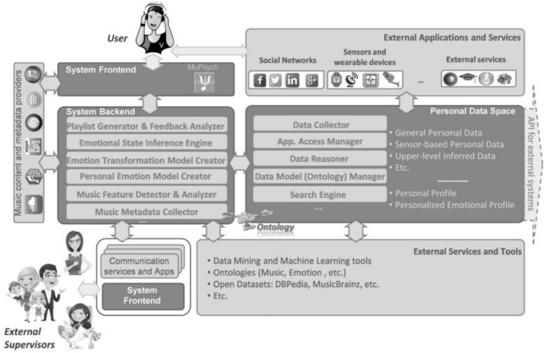


Fig. 3. Accuracy comparison of the models

The paper *Emotion-Based Music Recommender System* discusses personalized emotion-driven music recommendation systems. The approach presented in this study is targeted to provide maximum user benefits from the music-listening experience. It emphasizes the fact that to change an emotional state of a user, the main function of the system is to search for the nearest music tracks, which are defined by a certain set of music-related attributes.

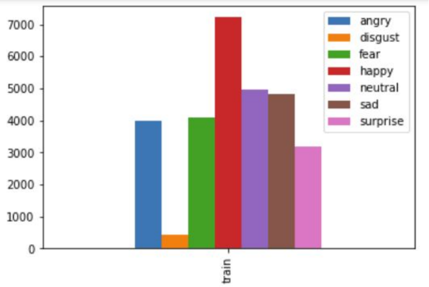


DATA DESCRIPTION: FER-2013 DATASET

FER Dataset (link)

The publicly available data consists of 48x48 pixel grayscale images of faces. The facial expression falls into one of the seven categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral) . However, we require only 2 classes, Happy and Sad, for our recommender system. An equal number of images (1000) of Sad and Happy are taken from this dataset.

For preprocessing: The input images are of size 48*48 pixels. As our models work on RGB images and require different input image sizes, we reshaped our images accordingly and extended them to 3 dimensions through duplicacy.



This plot tells about the imbalanced class distribution in the FER dataset. To have a class balance, we chose 1000 images from the 2 classes, happy and sad.

Data augmentation increases the amount of data by adding slightly modified copies of already existing data. We did this to reduce overfitting when training a machine learning model.

Data augmentation:

happy

happy

sad

sad

DATASET DESCRIPTION: SPOTIFY DATASET

The dataset has been extracted from the Spotify's publicly available API. However, due to rate limiting restrictions we were getting multiple error responses for our HTTP GET requests.

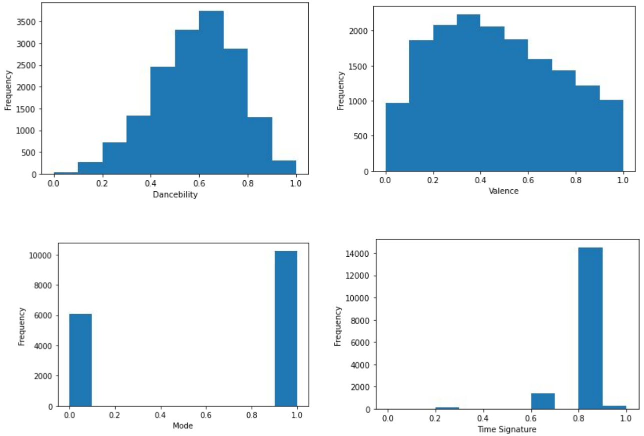
This information included meta-data and audio-features for all the songs. For every song we have extracted the following features along with their data types (as shown in the table).

We used the Spotify library to access the Spotify API for song information retrieval. First, we chose 18 genres from a total of 126 unique genres available on Spotify. We obtained information on 18,000 songs using 1000 random songs of each genre. Since We later reduced this dataset to 5000 thousands belonging to 5 unique genres: happy, sad, rock, edm and death-metal.

We also extracted the playlist data of user using custom client API.

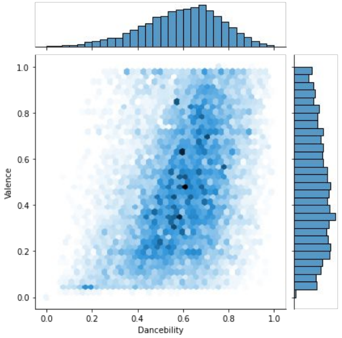
Feature Name	DataType
Song ID	String
Song Name	String
Artist Name	String
Album Name	String
Release Date	Integer
Acousticness	Float
Danceability	Float
Song Duration	Integer
Energy	Float
Instrumentalness	Float
Key	Integer
Liveliness	Float
Loudness	Float
Mode	Integer
Speechiness	Float
Tempo	Float
Time Signature	Integer
Valence	Float
Popularity	Integer

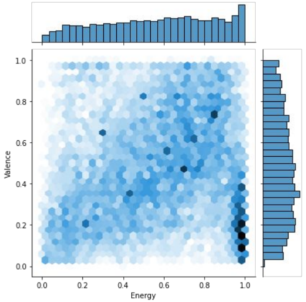
DATASET DESCRIPTION: SPOTIFY DATASET



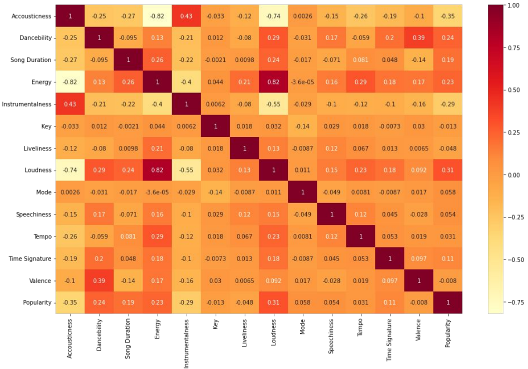
Frequency plots for a) Danceability and b) valence is distributed over a range. On the other hand, c) Time Signature and d) Mode contain only 4 unique values for the entire dataset.

Here, we can see that some features are distributed well over a range, while others mainly had the same value throughout the dataset. Thus, they perform a minimal role with our genre prediction and emotion score task and hence are mostly dropped during the feature selection task.

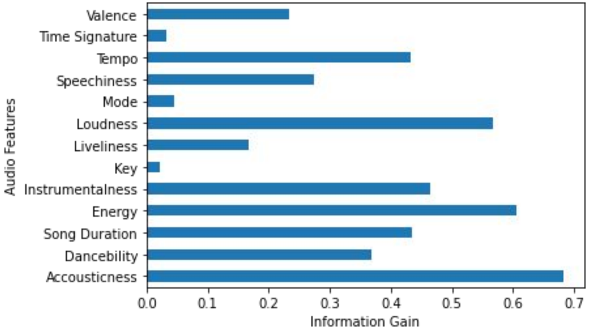




Features like valence, danceability, and energy contribute highly to emotion score prediction tasks, which is verified by the correlation graph as shown above.



Correlation Matrix



Information Gain Graph

METHODOLOGY

Facial Mood Recognition Task

Applied Transfer Learning and modified the following pre-trained CNN based models to perform Facial mood recognition task.
We have used Categorical cross-entropy loss along with adam optimizer with a learning rate of 0.001.
VGG16 model: very small (3 × 3) convolution filters are used. It pushed the depth to 16–19 weight layers,
making it approximately 138 trainable parameters.
Inception_V3 model: It has a total of 42 layers and a lower error rate than its predecessors.
MobileNetV2 model: There are two types of blocks. One is a residual block with a stride of 1. Another one is a block with a stride of 2 for downsizing.

We have also implemented this model on live photos.

METHODOLOGY

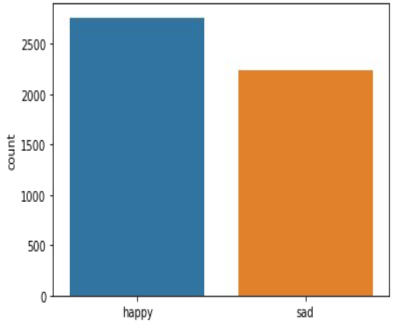
Music Emotion Recognition Task

Since no datasets available provide labels for emotions based on the above-discussed feature set, we have assumed the songs belonging to the ‘happy’ and ‘sad’ genres to also belong to ‘happy’ and ‘sad’ emotions, respectively. We trained our models to classify songs of other different genres as happy or sad emotionally.

The Emotion score of a song has been defined as the softmax probability of the song belonging to the ‘happy’ genre.

Multiple classification models like Logistic Regression, Multinomial Naive Bayes (MnB), SGD Classifier, K-Nearest Neighbor (KNN), Decision Tree Classifier, Random Forest Classifier (RFC), Support Vector Classifier (SVC), and Multi-Layer Perceptron (MLP) were used

Of these seven, the best results were observed in SVC and MLP. Further, a grid search was performed on both models to get better hyperparameter values.



Obtained balanced class distribution

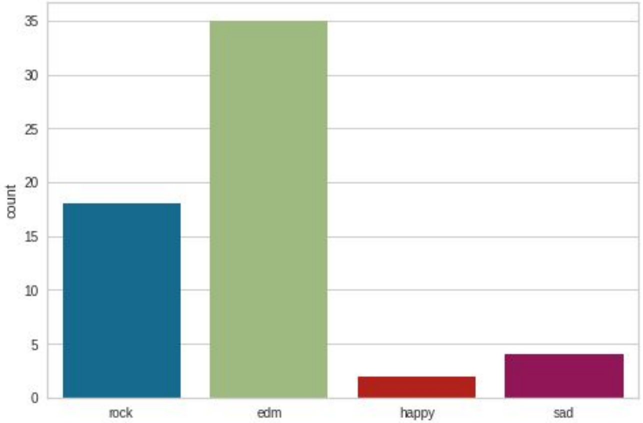
METHODOLOGY

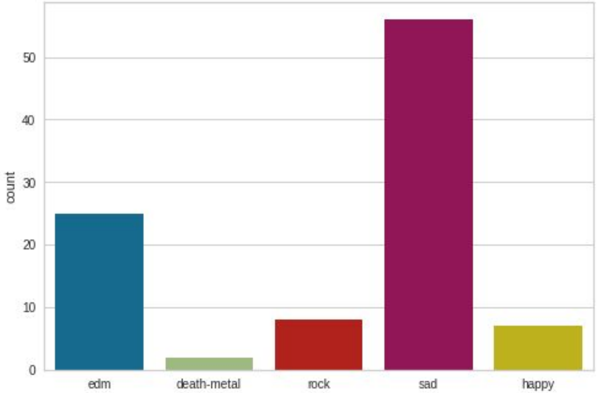
Playlist Genre Recognition Task

We used our custom Spotify Client API to access a user's playlist and got a list of the audio features of all the songs in this playlist. The aim was to classify each of these newly retrieved songs into a genre in order to retrieve which genre of songs the user generally listens to.

We had the genres of the initial custom dataset and used them for training purposes. We used the same seven classification models as the previous task. Of these seven, the best results were observed in MLP and RFC. Further, a grid search was performed on both models to get better hyperparameter values. The best-estimating hyperparameters for each came to be

Finally, RFC was used since it gave better results than fine-tuned MLP.





METHODOLOGY

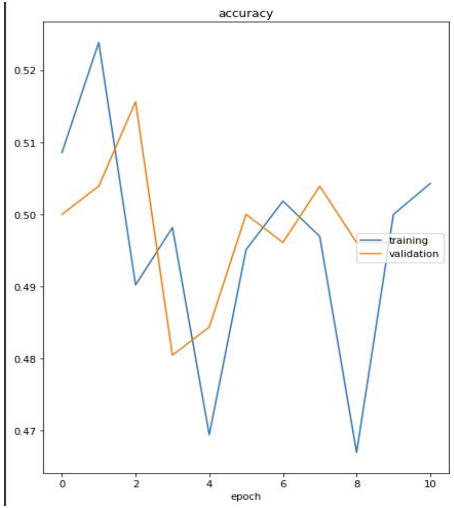
```
User ID user_5
User image
```

```
(224, 224, 3)
User favourite genre: rock
User human mood score: 0.015704870223999
listen to:
Song: Hold Me Closer By: Elton John
Song: The Beach By: The Neighbourhood
Song: Rosemary By: Deftones
Song: What Once Was By: Her's
Song: Delusions of Saviour By: Slayer
```

[Full-pipeline Recommendation Model]()

RESULTS

Facial Mood Recognition Task



Model Name	Training Accuracy	Testing Accuracy
InceptionV3	**0.96**	**0.91**
Vgg16	0.90	0.86
MobileNetV2	0.85	0.82

Accuracy of Inception_V3

Validation Set: **84.5%**

Testing Set: **82%**

RESULTS

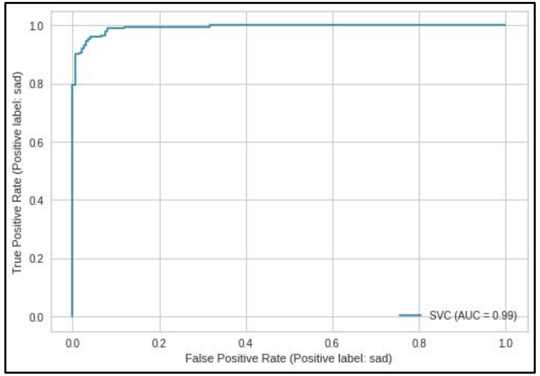
Classification Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.9525	0.9535	0.9524	0.9525
Multinomial Naive Bayes	0.9275	0.9298	0.9277	0.9274
SGD Classifier	0.9500	0.9524	0.9498	0.9499
K-Nearest Neighbor	0.9475	0.9478	0.9474	0.9475
Decision Tree	0.9400	0.9402	0.9399	0.9400
Random Forest Classifier	0.9525	0.9540	0.9524	0.9524
SupportVector Classifier	0.9525	0.9540	0.9524	0.9524
Multi-Layer Perceptron	**0.9525**	**0.9535**	**0.9524**	**0.9525**

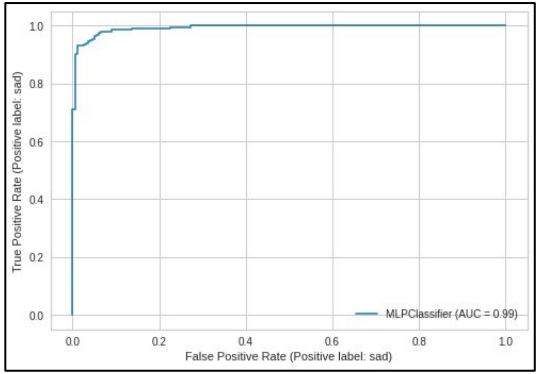
Music Emotion Score Task

Classification Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.7080	0.7112	0.7084	0.7094
Multinomial Naive Bayes	0.5480	0.5712	0.5600	0.5486
SGD Classifier	0.6573	0.6768	0.6658	0.6330
K-Nearest Neighbor	0.7120	0.7162	0.7111	0.7094
Decision Tree	0.6987	0.6983	0.6960	0.6969
Random Forest Classifier	**0.8093**	**0.8068**	**0.8091**	**0.8072**
Support Vector Classifier	0.7187	0.7199	0.7184	0.7186
Multi-Layer Perceptron	0.7547	0.7529	0.7549	0.7536

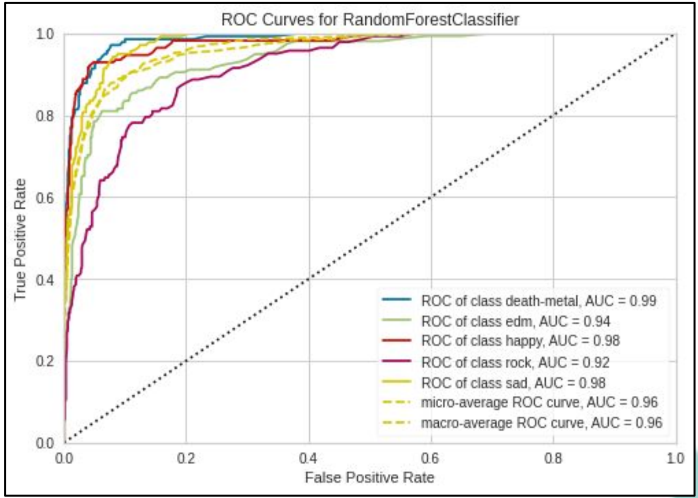
Playlist Genre Recognition Task

ANALYSIS





Music Emotion Score Task



Playlist Genre Recognition Task

TIMELINE

- We have managed to complete the project as proposed, following the tentative timeline we had given. We curated the datasets, trained and tested a model to predict human mood, stored song emotion scores for 5k songs, and classified the song's genre of different users' playlists.

- In the future, we plan to work on making this into a full-fledged deployable app or web extension. Also, we aim to test this pipeline with multiple users and make our model better.

INDIVIDUAL EFFORTS

- **Mudit Gupta**: Literature review, Data Extraction and Collection, Extraction and EDA of Spotify dataset, Analysis and inference of the data and results, Extraction of user's Spotify playlist, and making Genre prediction classification model.
- **Siya Garg**: Literature review, Data Extraction and Collection, EDA, feature selection, and data augmentation of FER 2013 dataset. Analysis and inference of the data and results, song emotion score model, and final pipeline for recommender system.
- **Srishti Jain**: Literature review, Data Extraction, and Collection, Extraction and EDA of Spotify dataset, Analysis and inference of the data and results, song emotion score model, the final pipeline for recommender system.
- **Sumit Soni**: Literature review, Data Extraction, and Collection of working Facial recognition model, Analysis, and inference of the data and results, creating a pipeline for capturing images and predicting facial mood score.

Drive folder containing all related material : link
GitHub : link

Any Questions?

Thank You

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**