

EMOTION-BASED MUSIC RECOMMENDATION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

A PROJECT PROPOSAL

Submitted by

Ayush Shakya

Bibek Thapa

Saurav Thakur

Yashuv Baskota

FUSE | MACHINES AI FELLOWSHIP 2023



fuse | machines

Problem Statement

Music is an art form encompassing the arrangement and combination of sounds, tones and rhythms in a structured and meaningful manner. It is an integral part of social, cultural and personal aspects of human life, and has the power to evoke a wide range of emotions and connect people on a profound level. Studies conducted in recent years have proven that music has a significant impact on the activity of the human brain and the general mood of the person. The timbre, beat, rhythm and pitch of music elicit different responses in the amygdala, thus leading to creation of preferences.

Human emotion also plays a large part in music choice through various factors such as emotional resonance, mood regulation and catharsis. Upbeat melodies and energetic rhythms appeal to happy minds, sad people seek out contemporary music and anxious people prefer calm and soothing music.

This project aims to provide personalized music recommendations by curating songs that empathize the person's emotions, detected using a facial recognition system. The project can be broadly divided into two parts: mood detection and music recommendation. The project aims to accurately detect the mood of the person and recommend music according to the identified emotion.

Dataset

- Source of the dataset

The CNN model for the Emotion-Based Music Recommendation System will be built using the *Kaggle* dataset called [FER-2013](#), one of the most popular datasets for facial expression recognition research. The dataset was created by collecting the results of a Google image search for various emotions. The dataset provides a valuable resource for training emotion recognition models.

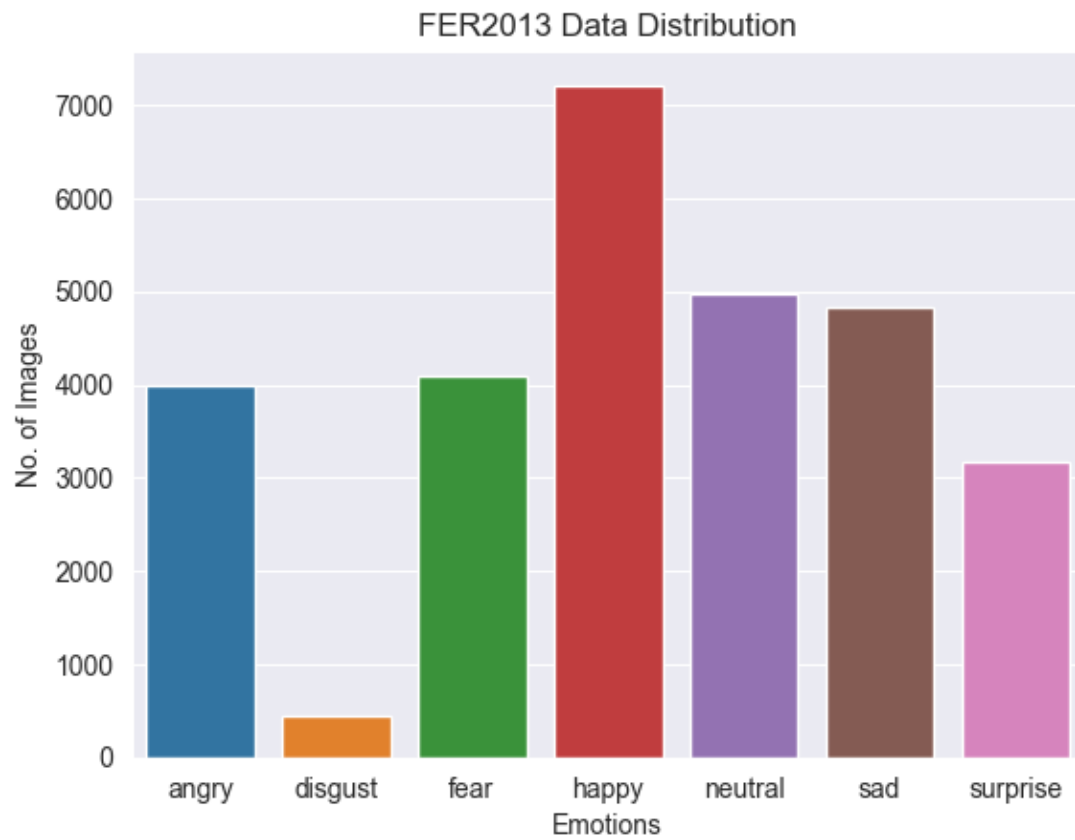
Similarly, for the music recommendation component of the project, the music dataset used to identify moods is sourced from *Kaggle* and is titled "[Spotify Music Data to Identify Moods](#)."

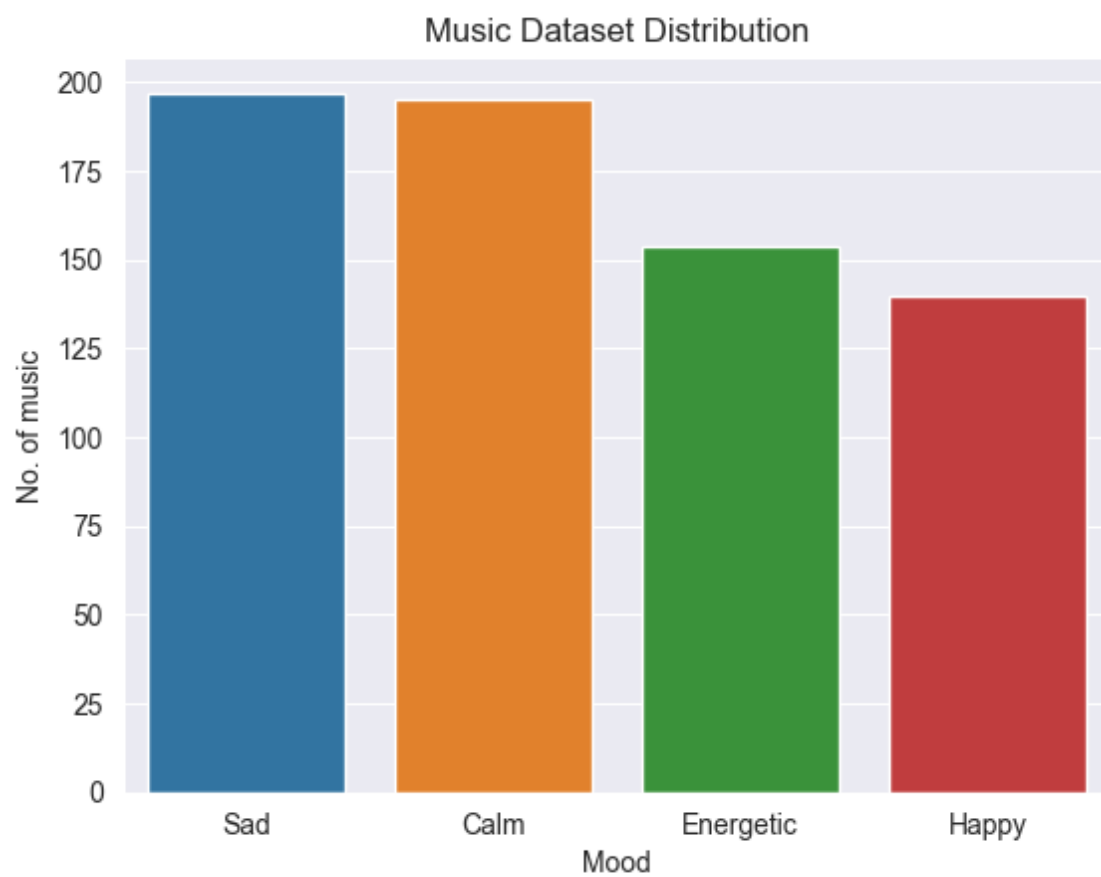
- Statistics about the dataset

- The **FER2013** dataset is divided into two parts: a training dataset and a testing dataset. The training dataset contains 24,176 images while the testing dataset consists of 7,178 images. Each image in the dataset is *grayscale* and has a resolution of 48x48 pixels. It aims to capture a wide range of facial expressions representing seven different classes of emotions: *Angry*, *Disgust*, *Fear*, *Happy*, *Sad*, *Surprise*, and *Neutral*.
- The **music dataset** is provided in CSV format and contains various attributes related to music tracks. The dataset consists of 19 *columns*, each providing valuable information for identifying the moods associated with music tracks. These columns include:

- *Album*: The name of the album to which the music track belongs.
- *Artist*: The name of the artist or group responsible for the music track.
- *ID*: A unique identifier for each music track.
- *Release Date*: The date when the music track was released.
- *Popularity*: A measure of the popularity of the music track.
- *Length*: The duration of the music track in seconds.
- *Danceability*: A metric indicating the suitability of the track for dancing.
- *Acousticness*: A measure of the level of acoustic elements in the track.
- *Energy*: The energy level of the track.
- *Instrumentalness*: A measure of the presence of instrumental elements in the track.
- *Liveness*: The likelihood of the track being performed live.
- *Valence*: A metric indicating the positivity or happiness of the track.
- *Loudness*: The loudness level of the track.
- *Speechiness*: A measure of the presence of spoken words in the track.
- *Tempo*: The tempo or speed of the track.
- *Key*: The key in which the track is composed.
- *Time Signature*: The time signature of the track.
- *Mood*: The mood or emotional category associated with the track.

- Sample visualization of data points





Possible Approaches

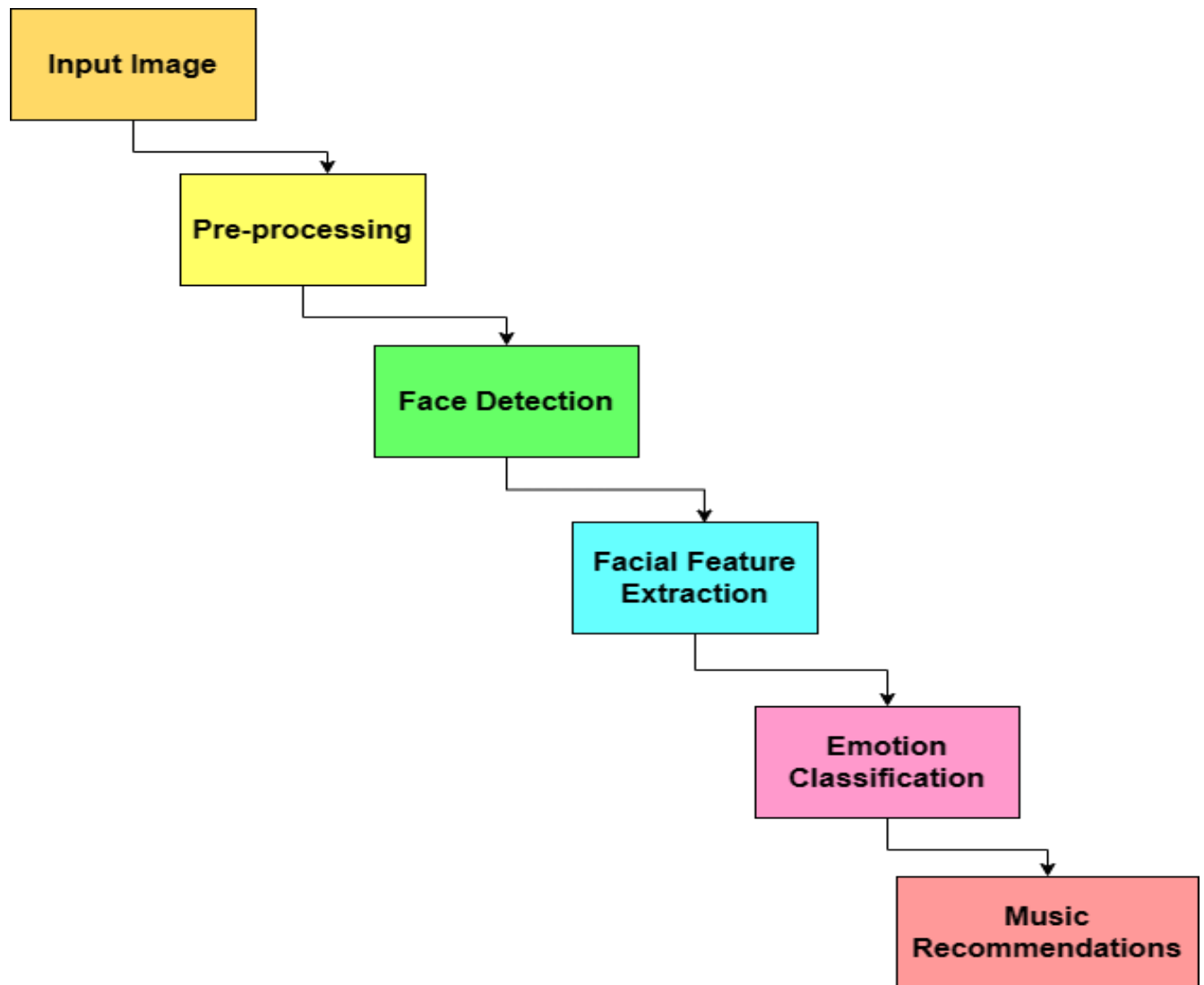


Fig: Expected Project Roadmap

Experimentation Tracking Process

We plan to use ML Flow, a comprehensive experiment tracking tool, to track the experimentation process of our project. It will allow us to easily track the different parameters and configurations that we use for our experiments, as well as the results of those experiments. This will help us identify the best-performing models and configurations and reproduce our results. We expect it to improve our efficiency and the effectiveness of the project.

Evaluation Metric

For the classification of emotions, we plan to use accuracy as a primary evaluation metric, which is a reasonable starting point to assess the classification performance of the system and compare different models or iterations during the development process. Accuracy is defined as:

$$\textit{Accuracy} = \frac{\textit{Number of Correct Predictions}}{\textit{Total Number of Predictions}}$$

Accuracy can be calculated for the emotion recognition and music recommendation.

Expected Outcome

The goal of the project is to accurately recognize the emotional states of the listeners through facial expression recognition, and provide personalized music recommendations based on those emotions. In the proposed application, the user provides their face to the model through a webcam, which detects the emotion and provides personalized music recommendation as output.