# Emotion Identification and Tagging Music with Appropriate Emotion

Sai Suman Chitturi (1602-18-733-097)

Praneeth Kapila (1602-18-733-116)

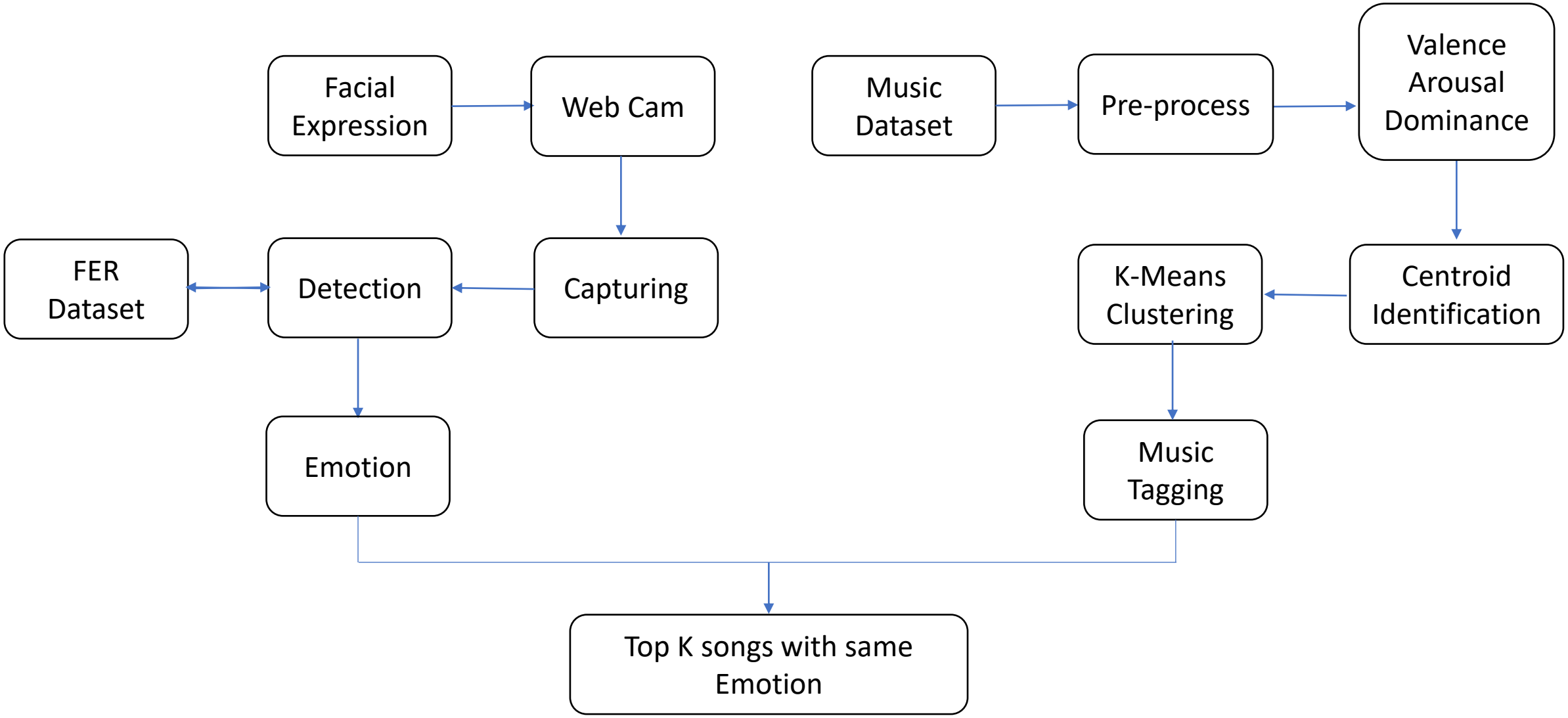
# Introduction

- Suggest Music based on User's Current Emotion
- Emotion Identification
  - Implicit: Facial Emotion, Keystrokes, Mouse-click patterns
  - Explicit: Input from User
- Tag Music & Suggest
  - Music Tagging: K-Means Clustering
  - Suggestions using Random Sampling

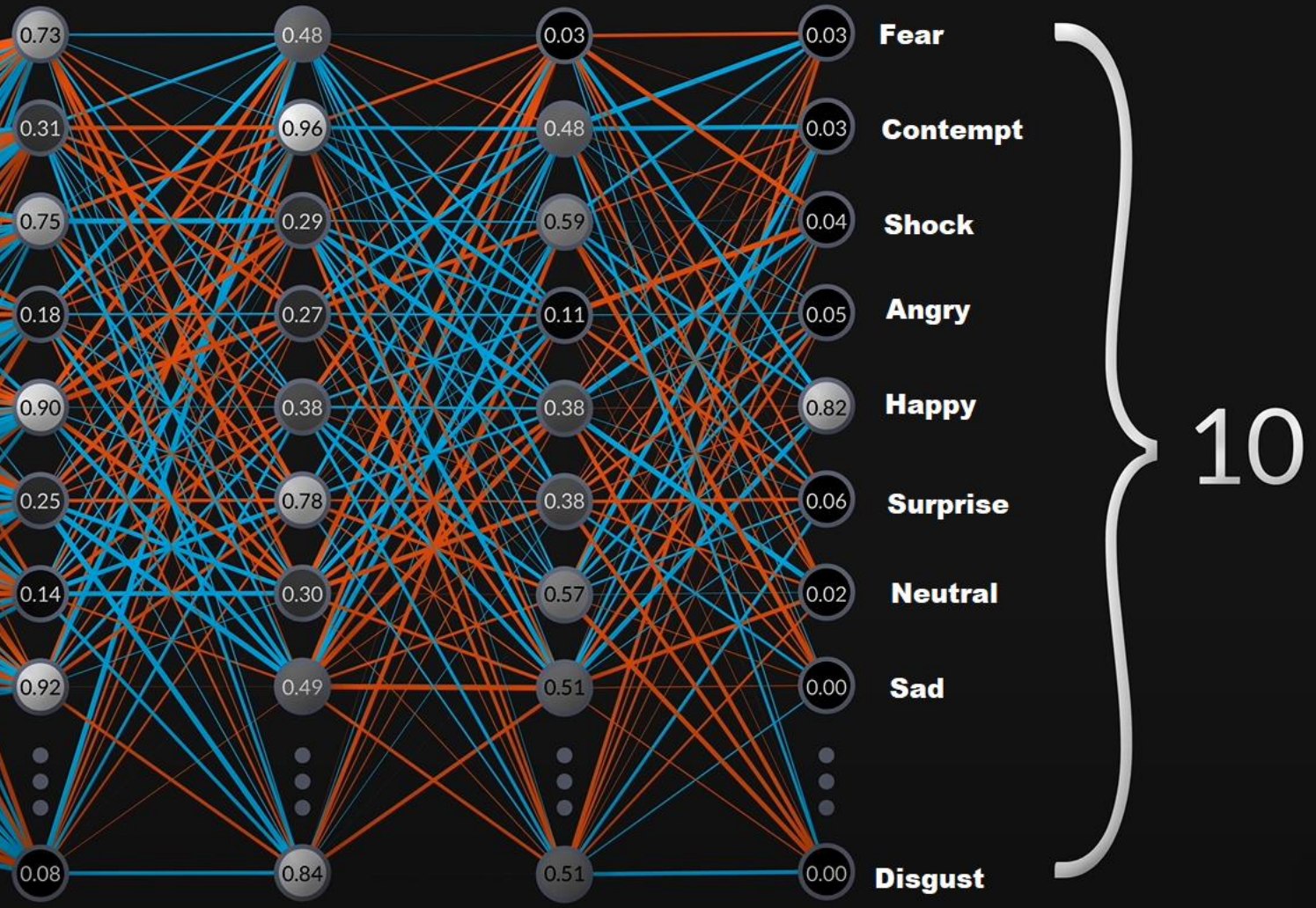
# Motivation

- Lack of Context-aware Music system
- Constantly Expanding Digital Music Libraries
  - ❖ Difficult to recall a particular song matching the current mood
- Confusion while choosing songs
- Useful when users can't reveal or express their emotion

# Flow


Facial Expression
Web Cam
Music Dataset
Pre-process
Valence Arousal Dominance
FER Dataset
Detection
Capturing
K-Means Clustering
Centroid Identification
Emotion
Music Tagging
Top K songs with same Emotion

# Facial Emotion Recognition: How it Works



# Literature Review: Convolutional Neural Networks

# [1]. Facial Emotion Recognition using an Ensemble of Multi-Level Convolutional Neural Networks

- Proposed a CNN based on multi-level features for Facial emotion identification.
- Hierarchy of Characteristics are considered to improve the classification job.
- Tested on the FER2013 dataset:
  - Found to be similar to existing state-of-the-art approaches in terms of performance

# Literature Review: Multi-task Cascaded Neural Network

# [2]. Research on Face Detection Technology Based on MTCNN

- Multitask Neural Network model for face detection.
- Image pyramid is used to transform the scale of the initial image.
- Require GPUs to train faster
- Very accurate and Robust

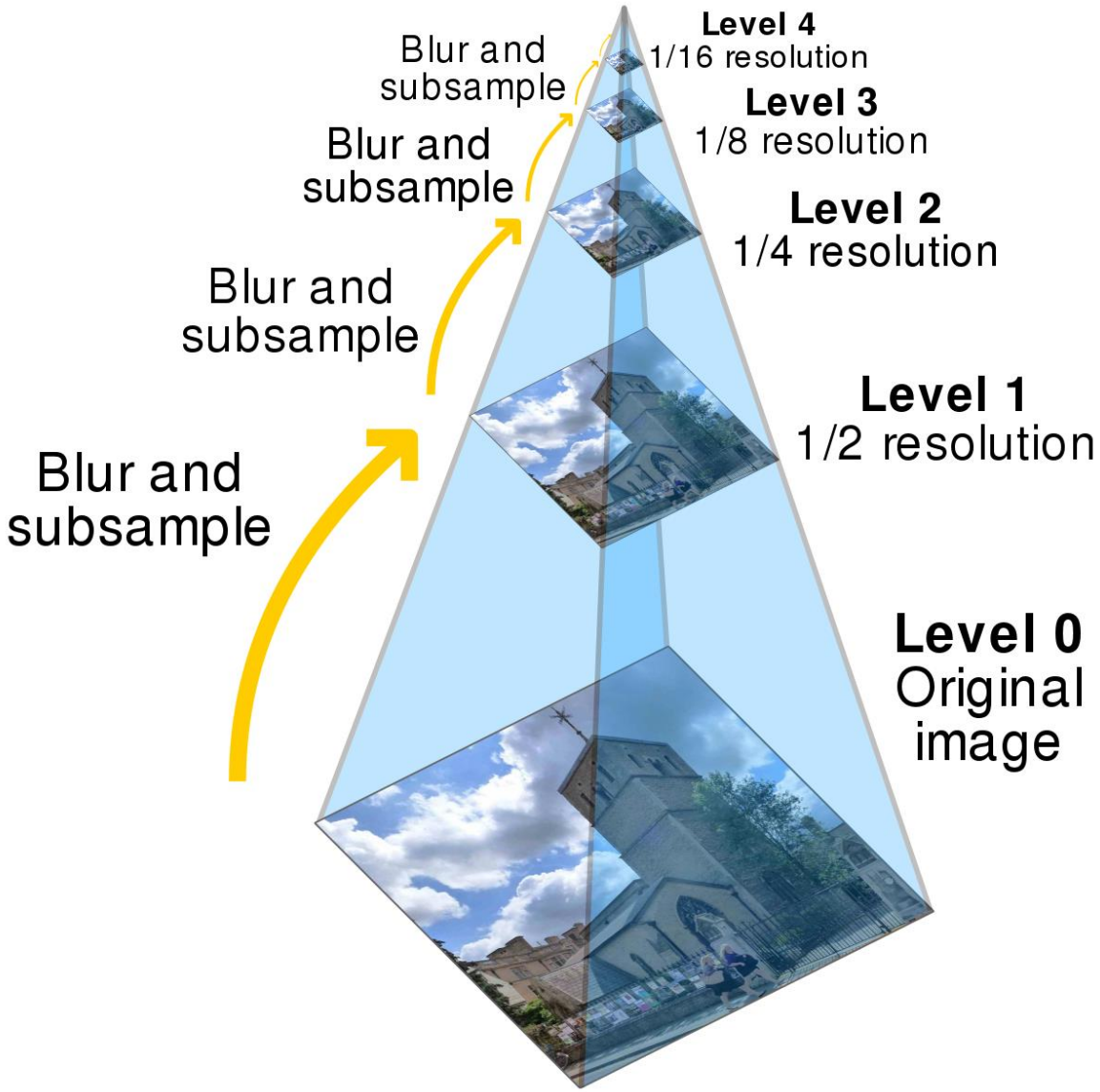



Image Pyramid

# Literature Review: Compare & Contrast

| Parameter | CNN | Deep Face (DNN) | FER (MTCNN) |
|---|---|---|---|
| Accuracy | High | Low | High |
| Train time | Low | High | High |
| Validation time | Low | High | High |
| Advantages | Speed | Light Weight | Self-alignment of Face |
| Disadvantages | Large Training Data | Low Accuracy | High Train Time |
| Common | Unable to detect rare facial expressions like Disgust | | |

# Proposed Methodology: Emotion Detection

- Based on the accuracies, both, CNN and MTCNN, seem a good fit for Facial Emotion Detection
- Choose CNN if time is an Important Factor
- Choose MTCNN if Accuracy is more important
- Time factor can be reduced by using GPUs.
  - ✓ MTCNN is better than CNN

# Implementation of Facial Emotion Recognition

- Datasets used:
  - FER2013: 28k train images + 7k test images; 48x48; B&W
  - Affect Net:  49k train images + 4k test images; 224x224; Colored

| Parameter | CNN | Deep Face (DNN) | FER (MTCNN) |
|---|---|---|---|
| Train Accuracy | 74.83 % | 87.35 % | 92.19 % |
| Train Dataset | FER 2013 Train | FER 2013 Train | FER 2013 Train |
| Test Dataset | Affect Net | Affect Net | Affect Net |

# Tagging Music with Appropriate Emotion

# Subset of LastFM Million Song Dataset

MuSe: The Musical Sentiment Dataset

| Column Label | Description |
|---|---|
| lastfm_url | Last.fm page of the song |
| track | Song title |
| artist | Artist name |
| seeds | The initial keyword(s) that seeded the scraping of this song |
| number_of_emotion_tags | Number of words that contributed to the emotion score of the song |
| valence_tags | Pleasantness dimension of the song |
| arousal_tags | Intensity dimension of the song |
| dominance_tags | Control dimension of the song |
| mbid | MusicBrainz Identifier of the song |
| spotify_id | Spotify Identifier of the song |

# Tagging Music

- Tagging Music involves clustering based on Valence, Arousal and Dominance values.
- VAD values identify the emotion associated with the song.
- These are floating-point values.
- Therefore, the dataset is plotted in 3-D space.
- K-Means clustering is used to group similar content.
- 7 Clusters are obtained: Each uniquely identifies an emotion.

# K-Means clustering

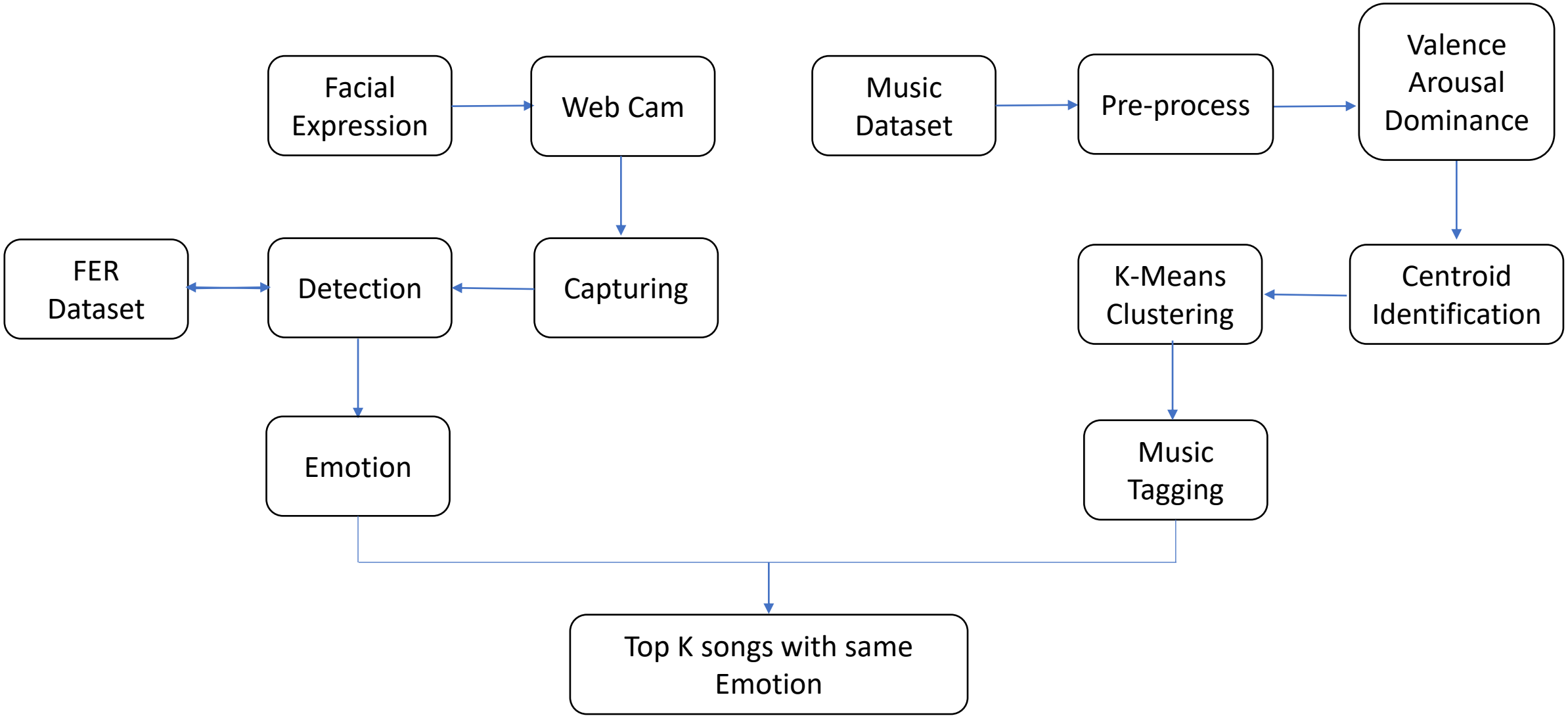
- VAD values are relative:
  - ➢ They change as the range of the dataset varies.

- Based on the range of VAD in MuSe dataset, initial centroids were identified.
- The 7 identified initial centroids uniquely determine the emotion associated with the song.
- K-Means Clustering is then performed with the initial centroids.
- 7 Clusters are obtained at the end, each identifying an emotion.

# Results

| Method | Accuracy/Silhouette score |
|---|---|
| Without Initial Centroids | 0.31 |
| With Initial Centroids | 0.35 |
| Split into Train/Test | 87% |

# Flow: Recap


Facial Expression
Web Cam
Music Dataset
Pre-process
Valence Arousal Dominance
FER Dataset
Detection
Capturing
K-Means Clustering
Centroid Identification
Emotion
Music Tagging
Top K songs with same Emotion

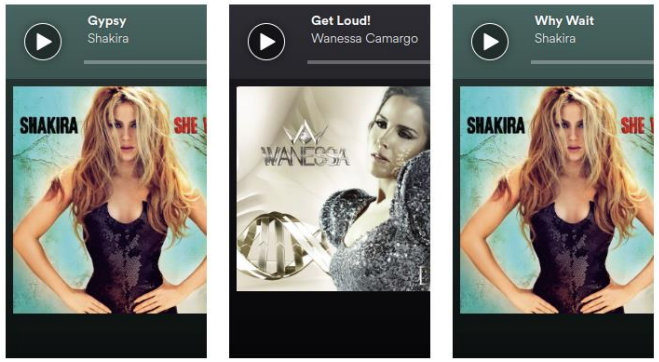
# Top picks for Identified Emotion

- Based on the identified emotion, top 15 songs that are tagged with same emotion are picked and displayed.
- Clickable Spotify Embed widgets are displayed that can be used to play the song.

Clear photo

**Identified Emotion: Surprise**

Top Tracks



Thank you