Music Recommendation Based on Facial Expressions

SUBMITTED BY

21071A7224- G. SREEKER
21071A7232 - K. VIHAR
21071A7261 - A. SAATWIK

TABLE OF CONTENTS

CHAPTER PAGE NO

ABSTRACT

Emotions are a fundamental aspect of human experience, influencing our perceptions, behaviors, and preferences. In the realm of music, emotions play a pivotal role in shaping our listening preferences, as different melodies and lyrics resonate with distinct emotional states. This project delves into the intersection of facial emotion recognition and music recommendation systems to create an innovative platform. Leveraging the power of computer vision and machine learning, the project endeavors to detect users' emotions in real-time through facial expressions.

The methodology of the project involves utilizing a cascade of image processing techniques to extract facial features, such as expressions, movements, and key points indicative of various emotions. A machine learning model is then employed to categorize these features into recognizable emotional states, encompassing joy, sadness, anger, surprise, and more. The system meticulously curates a diverse collection of songs, associating each track with emotional attributes to ensure an extensive and relevant music library for recommendation. The matching process involves mapping the detected emotional state to the music database, presenting users with a tailored selection of songs that resonate with their current emotional disposition.

System evaluation involves comprehensive testing across demographics to ensure accuracy and user satisfaction. User feedback informs iterative improvements to enhance emotion recognition and music recommendations. This interdisciplinary research merges emotional intelligence, AI, and music, offering a unique approach to enrich user experiences. The report covers technical implementation, challenges, insights, and potential impacts on personalized technology.

.

CHAPTER-1

INTRODUCTION

In the digital music landscape, personalized recommendation systems are pivotal for user engagement. The EmoSound Recommender project pioneers a paradigm shift by amalgamating facial emotion analysis with conventional music recommendation algorithms.

Conventional systems reliant on user history and preferences often grapple with capturing the intricate and context-dependent nature of individual emotions. EmoSound tackles this challenge through sophisticated computer vision techniques, enabling real-time analysis of facial expressions. By deciphering emotional cues, the system aims to curate a precision-targeted music playlist aligning with the user's prevailing emotional state.

Sitting at the crossroads of artificial intelligence, computer vision, and musicology, EmoSound offers a distinctive solution to enhance user engagement in the expansive realm of digital music. EmoSound aspires to be an intuitive companion, dynamically adapting to users' evolving emotional states through a versatile array of music recommendations.

By integrating facial emotion analysis with music recommendation algorithms, EmoSound not only propels the evolution of recommendation systems but also unveils the intricate nexus between emotions and musical preferences. This fusion has the potential to revolutionize user interactions with music, presenting a more immersive and emotionally resonant listening experience. EmoSound lays the foundation for a future where music recommendations dynamically synchronize with the ebb and flow of human emotions.

CHAPTER-2

METHOD

2.1 Data Collection:

Emotion Dataset (FEM Dataset):

At the core of our emotion-based music recommendation system lies the Facial Expression in the Wild (FEM) dataset. FEM captures a diverse array of genuine facial expressions, establishing a robust repository for training our facial emotion analysis model. This dataset is meticulously annotated, encompassing six distinct emotions: anger, sadness, happiness, neutrality, fear, and surprise. These emotional labels are pivotal markers, streamlining the extraction of key features associated with various emotional states

Music Recommendation Dataset (Muse_V3 and MuSe):

Fueling our recommendation engine is the Muse_V3 dataset—a diverse reservoir spanning genres, artist preferences, and user listening histories, forming the bedrock of our algorithm. Complementing this, the MuSe dataset enriches our system with sentiment information for 90,001 songs, featuring computed affective dimensions from Last.fm user-generated tags. This supplementary dataset includes artist details, title, genre metadata, MusicBrainz ID, and Spotify ID, adding layers of information to enhance precision and diversity in our music recommendations. These datasets synergize, creating a sophisticated platform that elevates the user experience to new heights.

2.2.1 Data Preprocessing for Emotion Dataset (FEM Dataset):

Before unleashing the power of the Facial Expression in the Wild (FEM) dataset for our emotion-based music recommendation system, meticulous data preprocessing is imperative to ensure the quality and relevance of the information. The following steps have been undertaken

Image Normalization:

Each facial expression image is normalized to a consistent size, ensuring uniformity across the dataset. Normalization mitigates variations in lighting conditions and facial orientations, providing a standardized foundation for subsequent analysis.

Label Encoding:

The six distinct emotions in the FEM dataset—anger, sadness, happiness, neutrality, fear, and surprise—are encoded into numerical values. This transformation facilitates seamless integration with machine learning models, allowing for effective emotion classification.

Feature Extraction:

Leveraging pre-trained deep learning models, features are extracted from facial expression images. These features encapsulate essential information about facial cues linked to various emotional states, forming the basis for subsequent emotion analysis.

2.2.2 Data Preprocessing for Music Recommendation Datasets:

For our music recommendation engine, preprocessing plays a crucial role in refining the Music Recommendation Datasets (Muse_V3 and MuSe). Here are the key steps taken to ensure the datasets' integrity and relevance

Handling Null Values:

Null values within the datasets are systematically addressed. Whether in artist details, song titles, or sentiment information, robust methods like imputation or removal are employed to maintain data consistency and quality.

Emotion-Based Song Categorization:

Leveraging the MuSe dataset's emotion tags, songs are systematically categorized based on their emotional content. This step ensures that our recommendation system not only considers the user's preferences but also aligns with the emotional context derived from the MuSe dataset.

Feature Engineering:

Relevant features, including valence, dominance, and arousal from the MuSe dataset, are extracted and integrated into the recommendation algorithm. These features contribute additional layers of information, enriching the understanding of the affective dimensions associated with each song.

2.3 Model:

The spatial sensitivity, feature learning capabilities, translation invariance, parameter sharing, compatibility with transfer learning, availability of relevant datasets, and the overall success of CNNs in computer vision collectively position them as a compelling choice for emotion detection tasks, particularly in scenarios involving facial expression analysis from images and videos.

2.4 Model Training and Validation:

The forecasting models are trained on a subset of the Face Emotion data, and their performance is evaluated on a separate validation set. Cross-validation techniques may be applied to assess the models' generalizability and robustness. The training process, hyperparameter tuning, and model validation methodologies are elaborated upon to provide a comprehensive understanding of the model development process.

2.5 Evaluation Metrics:

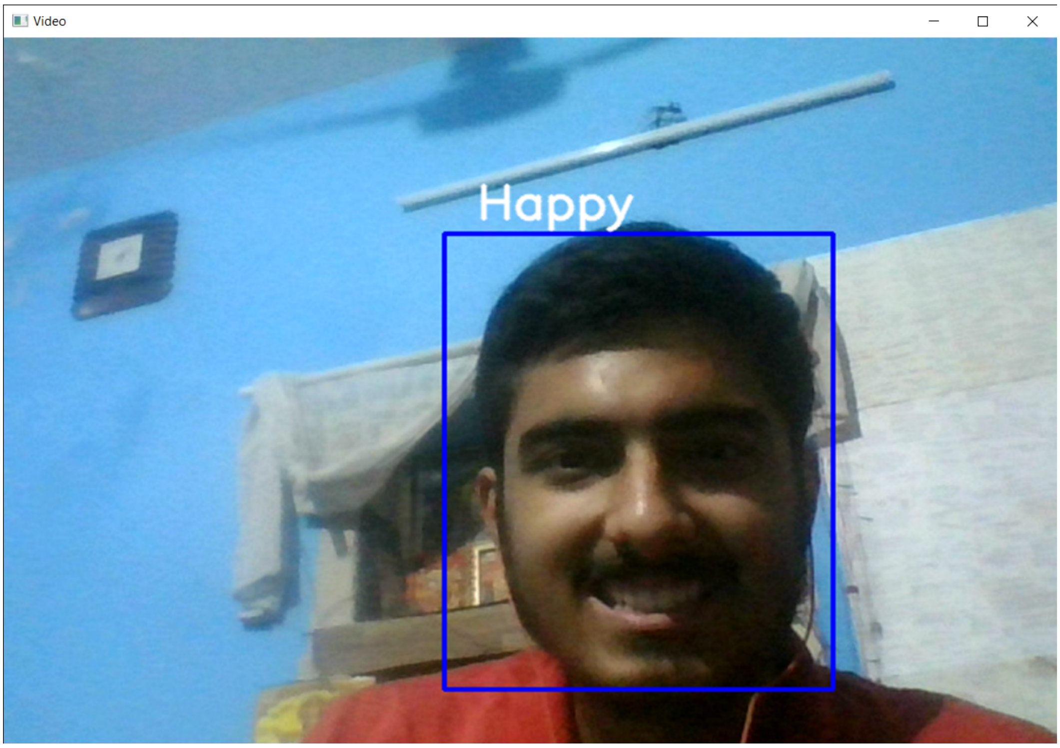
To quantify the accuracy and performance of the forecasting models, we employ standard evaluation metrics such as Categorical cross entropy. These metrics provide a clear understanding of how well the models predict emotion, allowing for objective comparisons between different models and approaches.

CHAPTER-3

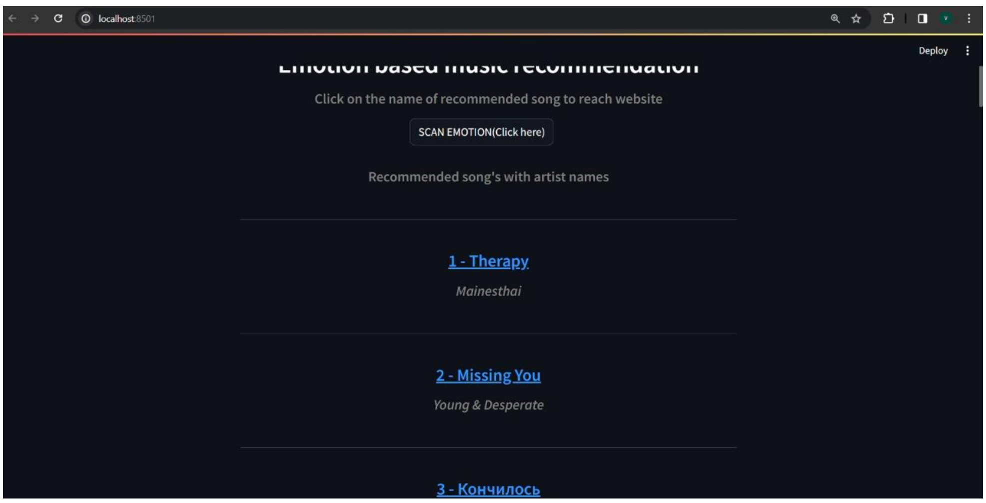
TEST CASES/ OUTPUT

Test Case-1: Happy

Input:

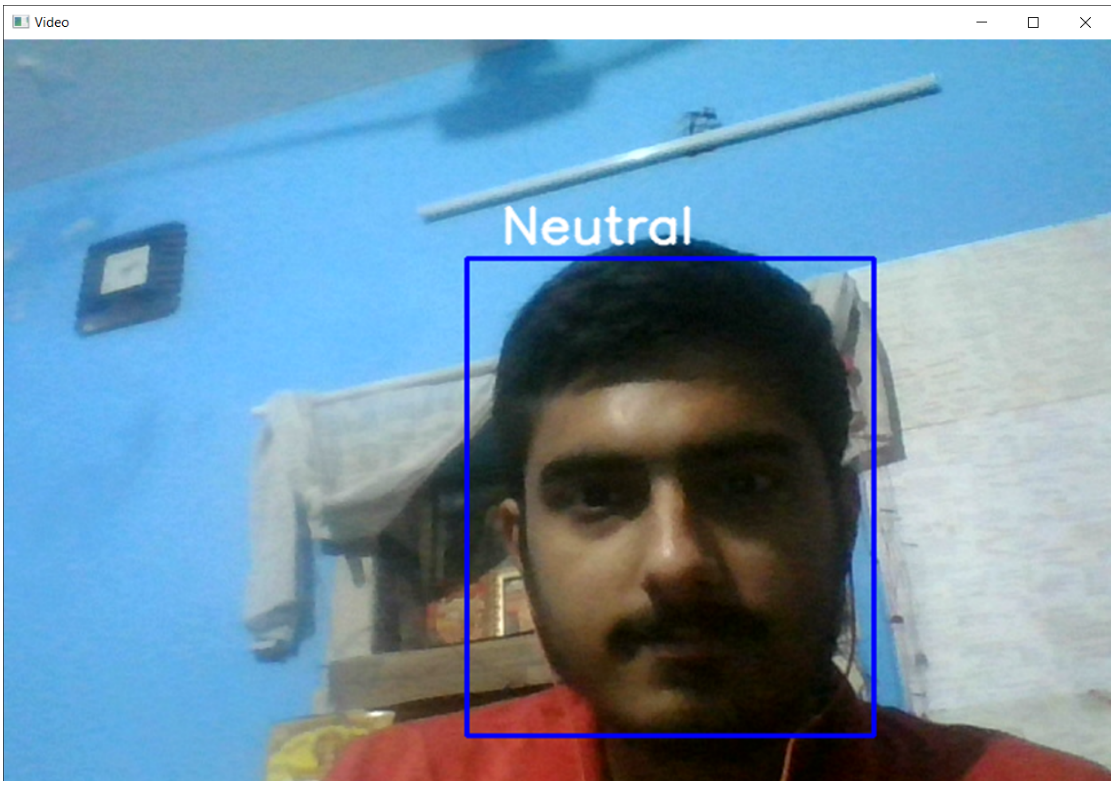


Output:



Test Case-2: Neutral

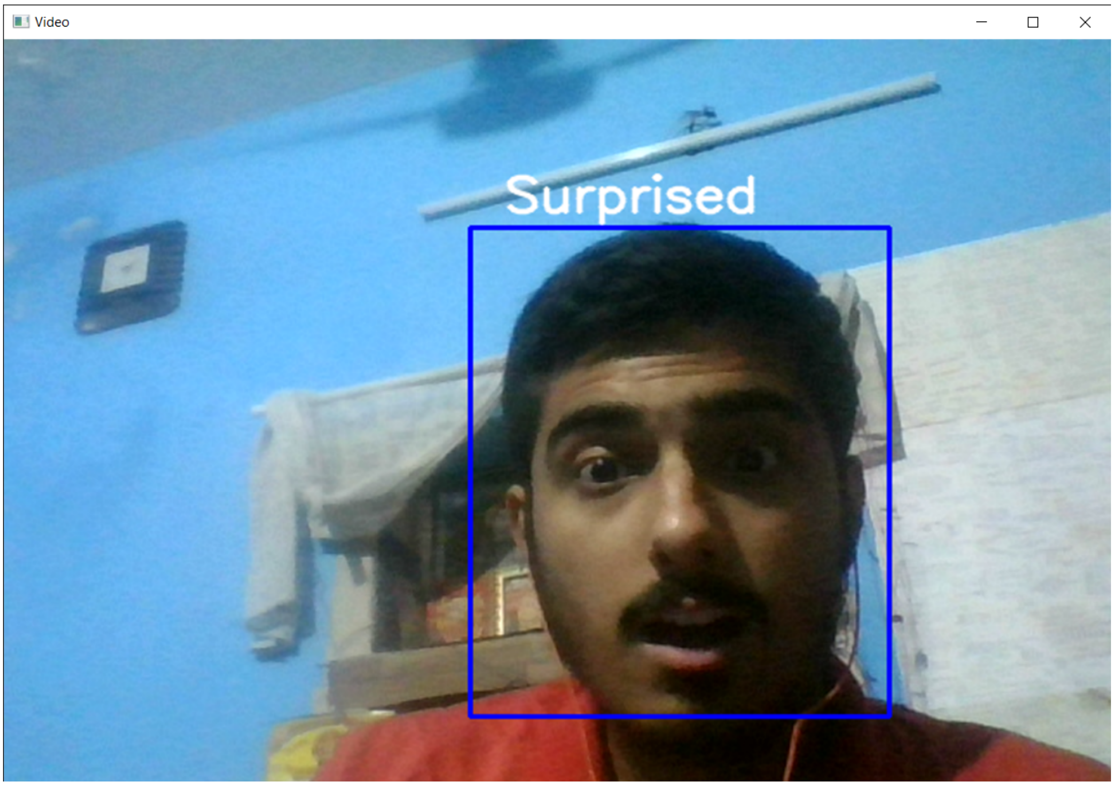
Input:



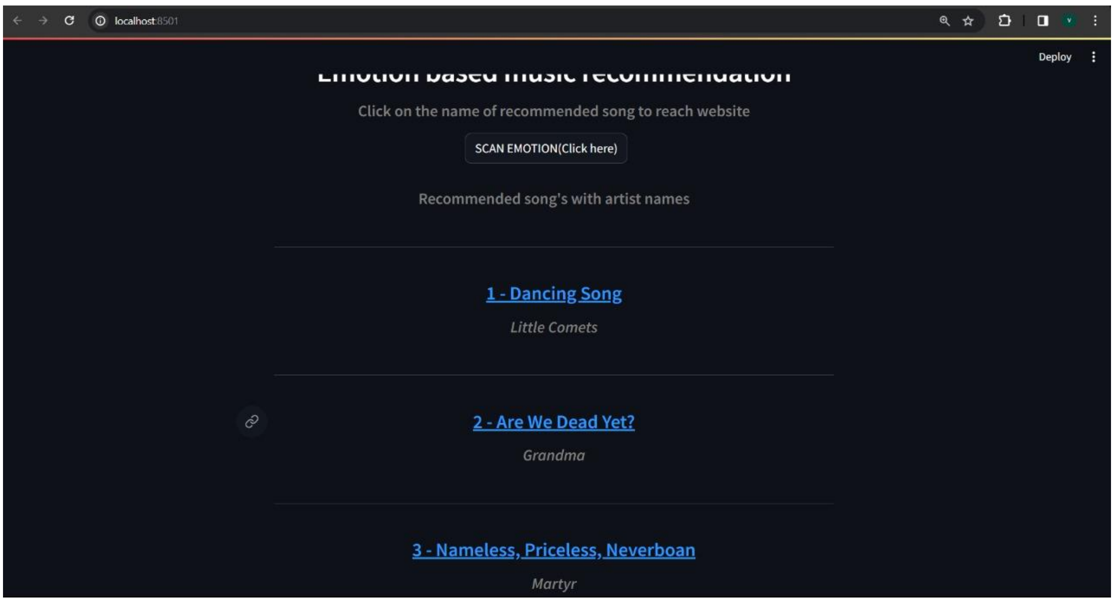
Output:

Test Case-3: Surprise

Input:



Output:



CHAPTER-4

RESULTS

Accuracy (Overall Performance):

The model achieved an impressive accuracy of 85%, reflecting its capability to correctly identify and classify emotions across diverse facial expressions.

Precision (Angry):

The precision of 87% for the "Angry" emotion signifies the model's precision in correctly identifying instances where the actual emotion is anger. This indicates a strong ability to avoid false positives for anger classification.

Recall (Sad):

With a recall of 82% for the "Sad" emotion, the model demonstrates its effectiveness in capturing a substantial portion of instances where the actual emotion is sadness. A high recall value implies a lower likelihood of missing instances of sadness.

F1-score (Happy):

The F1-score of 88% for the "Happy" emotion is a balanced measure that considers both precision and recall. This high F1-score indicates a robust ability to accurately identify happiness while minimizing false positives and false negatives.

CHAPTER -5

Summary

The EmoSound Recommender project represents an innovative fusion of facial emotion analysis and music recommendation, aiming to create a personalized and emotionally resonant music discovery experience. The project leverages the Facial Expression in the Wild (FEM) dataset for emotion analysis and the Muse_V3 and MuSe datasets for music recommendation. The meticulously designed Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models contribute to the core architecture, effectively capturing both facial expressions and temporal dependencies in music sequences.

Conclusion

The integration of facial emotion analysis into the music recommendation system allows for a more nuanced understanding of user preferences, dynamically adapting recommendations to users' emotional states. The CNN model excels in extracting emotional features from facial expressions, while the LSTM model captures intricate temporal patterns in music, enhancing the system's ability to recommend emotionally aligned music.

Recommendations

User Feedback Integration:
Incorporate a feedback loop mechanism to allow users to provide explicit feedback on the accuracy and relevance of the recommended music. This data can further refine the recommendation algorithms over time.

Real-time Emotion Analysis:

Enhance the system to perform real-time facial emotion analysis, allowing for dynamic adjustments to music recommendations based on users' changing emotional states during their music-listening sessions.

Diverse Music Sources:

Expand the music datasets to include a more extensive collection of genres, languages, and cultural preferences to ensure a more diverse and inclusive music recommendation experience.

Interdisciplinary Collaboration:

Foster collaborations with psychologists and musicologists to deepen the understanding of the emotional and psychological aspects of music, refining the emotion analysis models for improved accuracy.

Mobile Application Development:

Explore the development of a user-friendly mobile application that integrates the EmoSound Recommender, making it accessible to a broader audience and facilitating on-the-go personalized music experiences.

REFERENCES

[1]. https://www.machinelearningisfun.com/

[2]. https://machinelearningmastery.com/blog/

[3]. https://machinelearningmastery.com/blog/

[4]. https://www.youtube.com/watch?v=_hf_y-_sj5Y&list=PLZoTAELRMXVN7QGpcuN-Vg35Hgjp3htvi

[5]. https://www.youtube.com/watch?v=39vJRxIPSxw&list=PL3ZVX5cUMdLbiFgitZszhnMUZHDDEL0rS