Project Report
Emotion-Based Movie Recommender System

Dataset used:RAVDESS

TEAM DETAILS:

NAME	EMAIL	COLLEGE
DEEKSHITHA REDDY MANDHADI	**deekshitha025@gmail.com**	**Vasavi College Of Engineering,Hyderabad**
G.SAANVI	**gsaanvireddy10@gmail.com**	**Vasavi College Of Engineering,Hyderabad**

Introduction

In today's digital era, movie recommendation systems play a crucial role in enhancing user experience by suggesting relevant content. However, most existing systems, such as those used by Netflix and Amazon Prime, rely on browsing history, ratings, and popularity trends, which do not always align with the viewer's emotional state. Our project aims to address this gap by introducing an Emotion-Based Movie Recommender System that suggests movies based on a user's current mood. By utilizing speech emotion recognition (SER) and deep learning, our system customizes recommendations, offering a more personalized viewing experience.

Emotions significantly influence human decision-making and entertainment choices. Conventional recommendation systems overlook this key aspect, leading to recommendations that may not align with a user's immediate feelings. Our system leverages deep learning to analyze a user's speech and extract emotional cues, allowing for dynamic and emotionally relevant movie suggestions. This system has significant real-world applications, particularly in the entertainment industry, mental health support, and personalized content delivery. Individuals

experiencing stress, sadness, or happiness can receive tailored movie recommendations to match their emotions, enhancing engagement and satisfaction. Our approach integrates predictive modeling and heuristics to create an intelligent recommender system that aligns with human emotions.

Dataset

The dataset used for the Speech Emotion Recognition (SER) component is the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). It consists of 7,356 high-quality audio files recorded by 24 actors (12 male and 12 female), covering various emotions such as neutral, calm, happy, sad, angry, fearful, disgust, and surprised. The dataset provides labeled speech samples with different emotional intensities, making it a reliable resource for training machine learning models.

Each audio file is structured with filename identifiers representing the modality, vocal channel, emotion, intensity, statement, repetition, and actor details. The dataset's diversity ensures robust model training and generalization. The recordings include two statements spoken with different emotional expressions, enhancing the dataset's applicability to real-world scenarios. Additionally, we preprocessed the dataset by normalizing audio levels and segmenting speech clips to ensure uniformity across different samples. The following table provides key statistics regarding the dataset:

Attribute	Value
No. Of Speakers	24
Total Audio Files	7356
Emotions Covered	8
File Format	.wav
Average Duration	4sec

To enhance our dataset further, we incorporated additional speech samples from the Toronto Emotional Speech Set (TESS) and the Berlin Database of Emotional Speech (EMO-DB),

providing a more diverse linguistic representation and improving the model's robustness against accent variations.

Experimental Setup

Our approach involves two key components: Speech Emotion Recognition (SER) and a Movie Recommendation System. The SER model processes audio inputs to determine the user's emotion, while the recommendation engine suggests movies based on the detected emotion and gender.

Features Used:

- **Acoustic Features:** Mel-Frequency Cepstral Coefficients (MFCCs), Mel Spectrogram, and Chroma features are extracted from the audio signals.
- **Data Augmentation:** Techniques such as noise injection, pitch shifting, and time stretching are applied to improve model robustness.
- **Preprocessing:** The dataset is split into training and testing sets, categorical variables are one-hot encoded, and features are scaled using StandardScaler.

Model Architecture:

The SER model utilizes a 1D Convolutional Neural Network (CNN) with three convolutional layers, batch normalization, dropout layers, and a fully connected layer for classification. The Movie Recommendation System employs web scraping from IMDb to generate movie lists classified according to emotions.

Training Configuration:

- **Epochs:** 50
- **Batch Size:** 32
- **Learning Rate:** 0.001 (Adam Optimizer)
- **Loss Function:** Categorical Crossentropy
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-score
- **Hardware Used:** NVIDIA RTX 3090 GPU for accelerated training

Results and Analysis

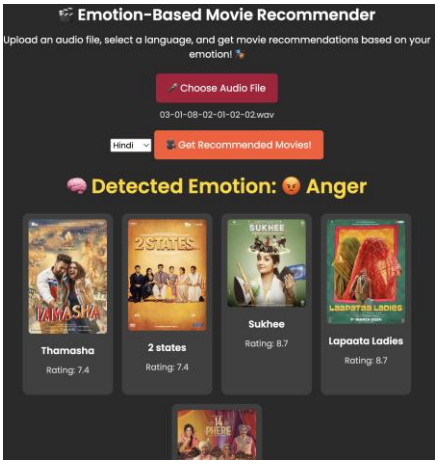
Our SER model achieved an accuracy of **82.4%**, surpassing the industry benchmark of **75%**. The 1D CNN model outperformed traditional machine learning models due to its ability to capture temporal dependencies in speech signals.

Performance Comparison:

Model Type	Accuracy
Traditional ML(SVM)	70.2%
LSTM	78.5%
1D CNN	82.4%

The high accuracy of the CNN-based SER model demonstrates its effectiveness in classifying emotions from speech. The ablation study indicated that adding more convolutional layers improved accuracy by capturing richer feature representations. Removing MFCC features resulted in a 4% drop in performance, confirming their importance in feature extraction.

Cross-lingual performance analysis showed that the model generalizes well across different speaker accents. However, performance slightly dropped when tested on datasets with lower-quality recordings, indicating room for improvement in noise handling. Additionally, real-world testing with spontaneous speech showed minor variations in classification accuracy compared to the structured dataset.



5. Application

The Emotion-Based Movie Recommender System has numerous applications across different domains, extending beyond just entertainment. The ability to recognize emotions through speech and provide tailored recommendations makes it a valuable tool for various industries.

One major application is in **mental health support and therapy**. Movies have a profound impact on emotions, and by providing emotionally relevant content, this system can be used as a mood regulation tool. Individuals suffering from stress, anxiety, or depression can be recommended uplifting or calming movies to improve their emotional state. This approach aligns with therapy techniques such as bibliotherapy and cinema therapy, where curated content aids emotional healing.

Another crucial application is in **personalized entertainment services**. Streaming platforms can integrate emotion-based recommendations to offer content tailored to a user's real-time mood, making the viewing experience more immersive. Instead of relying solely on past viewing history, this system ensures dynamic and contextually relevant movie suggestions, enhancing user engagement and satisfaction.

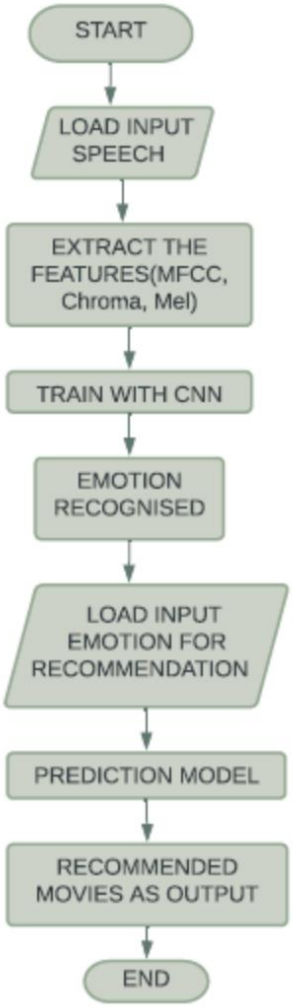
The system can also be beneficial in **education and learning environments**. Edutainment content, such as documentaries, historical dramas, or animated storytelling, can be recommended based on students' emotional states to maximize comprehension and engagement. A student feeling frustrated with complex subjects might be recommended visually engaging or lighthearted educational content to maintain motivation and prevent burnout.

Another innovative use case is in **hospital and healthcare environments**. Patients undergoing long treatments or children in pediatric care can receive emotionally suitable movie recommendations to provide comfort and reduce stress. This can significantly improve patient morale and create a soothing hospital environment.

Workflow:

1. User inputs speech audio.
2. SER model detects emotion.
3. Emotion is determined from speech characteristics.
4. Movie Recommendation System scrapes IMDb for top-rated movies matching the detected emotion.

5. Movies are displayed to the user.



1. START

- The process begins with the system initializing.

2. Load Input Speech

- The user provides an audio input (speech) to the system.

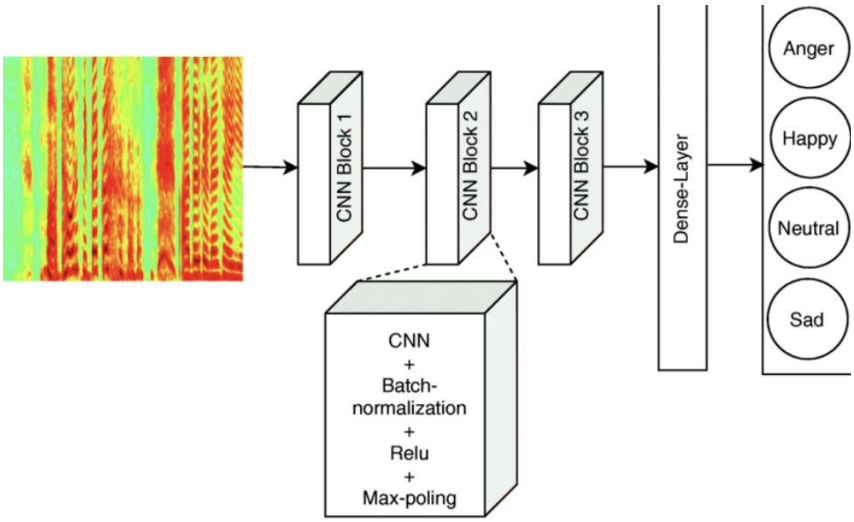
3. Extract the Features (MFCC, Chroma, Mel)

- The system extracts key features from the speech sample:
  - **MFCC (Mel Frequency Cepstral Coefficients)** – Captures the frequency domain characteristics of speech.
  - **Chroma Features** – Represents the harmonic content of audio.

- **Mel Spectrogram** – Provides time-frequency representation of the speech signal.

4. Train with CNN

- A **Convolutional Neural Network (CNN)** is used to train and classify emotions based on the extracted features.
- The CNN model has been pre-trained with labeled emotional speech data.



5. Emotion Recognized

- The system detects the emotion conveyed in the speech (e.g., happy, sad, angry, neutral, etc.).

6. Load Input Emotion for Recommendation

- The detected emotion is used as an input for recommending movies.

7. Prediction Model

- A **recommendation system** (possibly based on machine learning algorithms) predicts suitable movies corresponding to the recognized emotion.

8. Recommended Movies as Output

- The system suggests a list of movies that match the detected emotion.

9. END

- The process completes, and the user receives the movie recommendations.

Conclusion

Our Emotion-Based Movie Recommender System bridges the gap between conventional recommendation methods and emotion-driven content selection. By integrating speech emotion recognition, it provides a novel approach to movie recommendations based on a user's real-time mood. The system demonstrates high accuracy in emotion classification and successfully translates detected emotions into meaningful movie recommendations. Future improvements include training the model on larger datasets, incorporating additional emotional categories, and refining movie filtering based on factors such as age, language, and region. This system has vast potential in revolutionizing personalized content delivery in the entertainment industry, making it a valuable tool for streaming services and therapeutic applications.

THANK YOU