Prototype Documentation: Aggression Detection Prototype

Approach

This prototype focuses on detecting aggression in audio data. It utilizes both audio-based features and machine learning techniques to classify emotions. The primary steps include data extraction, feature engineering, model training, and prediction.

Features Chosen

Audio Features: MFCC (Mel-Frequency Cepstral Coefficients): These capture the spectral characteristics of audio.

Chroma Features: Represent the 12 different pitch classes.

Mel Features: Reflect the spectral characteristics of sound.

Preprocessing Steps

Data Extraction: Sound files are read and transformed into feature vectors using the extract_feature function.

Data Split: The dataset is divided into training and testing sets.

Libraries and Tools

Librosa: Used for audio feature extraction.

Soundfile: Used for reading sound files.

Numpy: Used for numerical operations.

Scikit-learn: Used for machine learning.

Pandas: Used for data organization and visualization.

Pickle: Used for model serialization.

Challenges Faced

Variability in Audio Data: Differences in audio quality and recording conditions can impact model performance.

Data Labeling: Collecting labeled data for training the model can be resource-intensive.

Model Optimization: Tuning hyperparameters to achieve higher accuracy and robustness is a challenge.

User Manual

Input:

Audio Files: Users can provide their own audio files for aggression detection. The accepted audio file formats include:

MP3,WAV,Other common audio formats containing spoken content.

Analysis: Feature Extraction: The prototype extracts MFCC, chroma, and mel features from the audio data.

Model Training: A Multi-Layer Perceptron (MLP) classifier is trained on the extracted features to classify emotions, with a focus on detecting aggression.

Output:

The prototype reports an accuracy score, indicating how well the model performs in aggression detection.

F1 Scores: F1 scores for individual emotions (e.g., 'calm,' 'disgust,' 'happy,' 'fearful') are provided.

Emotion Predictions: Users can predict emotions for their audio data using the trained model. A sample prediction ('disgust') is demonstrated in the code.

Model Serialization: The trained model is saved to a file for future use.

Conclusion:

This prototype offers a solution for aggression detection in audio data. Users can input audio files and receive emotion predictions based on the model's training. The use of audio features and machine learning enables the model to classify emotions, with a focus on detecting aggression.