Speech Sentiment Analysis using Deep Learning

Objective:

To classify human emotions from speech signals using the RAVDESS dataset and deep learning (CNN + LSTM) with MFCC features.

Dataset Used:

RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)

Contains 8 emotions: neutral, calm, happy, sad, angry, fearful, disgust, surprised

Preprocessing & Feature Extraction:

Librosa used to load audio and extract MFCC (Mel Frequency Cepstral Coefficients)

Each MFCC feature is reshaped to: (13, 174, 1)

Model Architecture (Total: 9 Layers):

Input Layer Accepts MFCCs shaped (13, 174, 1)

Conv2D Layer 1 32 filters, 3x3 kernel, ReLU, padding='same'

MaxPooling2D Layer 1 2x2 pool size

Conv2D Layer 2 64 filters, 3x3 kernel, ReLU, padding='same'

MaxPooling2D Layer 2 2x2 pool size

Conv2D Layer 3 128 filters, 3x3 kernel, ReLU

Reshape Layer Flattens to shape suitable for LSTM

LSTM Layer 1 128 units, returns sequences

LSTM Layer 2 64 units

Speech Sentiment Analysis using Deep Learning

Dense Layer 128 units, ReLU + Dropout (0.5)

Output Layer Softmax layer with num_classes=8

Total Layers: 11 (including Reshape, Dropout, and Output)

Training Details:

Loss: sparse_categorical_crossentropy

Optimizer: Adam

Epochs: 30

Batch Size: 32

20% of training data used for validation

Evaluation:

Metrics used:

- Confusion Matrix

- Classification Report (Precision, Recall, F1-score)

- Accuracy & Loss Curves

Final training and testing accuracy reported

Overfitting monitored via plots

Real-Time Emotion Prediction:

Records live speech (35 seconds)

Extracts MFCC, reshapes, and passes to trained model

Speech Sentiment Analysis using Deep Learning

Predicts and displays emotion in real time using sounddevice and librosa

Results:

Accurately detects emotions from speech

Real-time prediction system works efficiently

Can be extended to datasets like TESS or CREMA-D

Conclusion:

This project shows that MFCC features combined with a CNN-LSTM hybrid model (11 layers) can effectively classify emotions from speech. The real-time system is accurate and can be applied in emotion-aware applications like virtual assistants or therapy bots.