Speech Emotion Detection Using Classical ML

Team: Mahashweta Panigrahi (PES2UG23CS315)
        Lakshita Negi (PES2UG23CS301)

1. Problem Statement

The project aims to classify emotions (neutral, happy, sad, angry, calm, fearful, disgust, surprised) from short audio clips using classical ML methods. Speech emotion recognition has applications in customer support, assistive tech, and entertainment.

2. Dataset

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

- 8 emotions, 24 speakers (12 male, 12 female), WAV 48 kHz

- Filenames encode speaker, emotion, intensity, take

Preprocessing: Extracted MFCC, chroma, spectral contrast, and zero-crossing rate features using librosa. Aggregated mean features per clip.

3. Methodology

- **Feature Extraction:** MFCC & spectral features aggregated into fixed vectors.

- **Models:** Random Forest and SVM (RBF), standardized with StandardScaler.

- **Tuning:** SVM hyperparameters (C=10, gamma='scale') tuned via GridSearchCV.

- **Evaluation:** 80/20 train-test split; metrics — Accuracy, Precision, Recall, F1.

Tools: Python, librosa, scikit-learn, numpy, joblib

5. Results

Model	Accuracy	F1-Score
Random Forest	Random Forest	0.55
SVM (tuned)	66 %	0.61

Inference: SVM outperformed RF, especially on angry, fearful, and surprised emotions. RF underperformed without tuning.

6. Challenges & Future Work

During the project, we faced issues like dataset imbalance, similar acoustic patterns between certain emotions, and minor background noise.
In the future, we plan to add delta MFCC features, perform deeper hyperparameter tuning, and explore CNN-based models on spectrograms to boost accuracy.Minor noise and speaker variation

7. Conclusion

The SVM model achieved 66% accuracy on RAVDESS, which is within the expected range for MFCC-based approaches(60-75%) , providing a strong baseline for future improvements.