Project Proposal: Voice-based Emotion Detection

a. Project Title

Voice-based Emotion Detection

b. Problem Statement

Understanding human emotions through voice has significant applications in customer support, mental health, and human-computer interaction. Current systems often fail to capture emotional nuance across diverse voices, accents, and tones. This project aims to develop a scalable solution for real-time voice-based emotion recognition.

c. Novelty/Complexity

The project stands out by addressing emotional subtleties in diverse speech data, requiring large-scale datasets for generalization. It integrates robust preprocessing, model training, and real-time deployment. Complexity arises in handling accents, tones, background noise, and monitoring drift when new voice patterns appear.

d. Objectives

- Build an audio preprocessing pipeline (noise reduction, feature extraction).
- Train models on large labeled emotion datasets (RAVDESS, CREMA-D).
- Deploy a real-time inference system usable in call centers or mental health apps.
- Integrate monitoring to detect performance drift as new voices/accents are encountered.
- Provide a scalable MLOps pipeline for continuous retraining and improvement.

e. Dataset Requirement

- RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)
- CREMA-D (Crowd-sourced Emotional Multimodal Actors Dataset)

Why dataset-heavy? Emotions are subtle; a large volume of varied voices is needed to generalize across accents, tones, and recording conditions.

f. MLOps Elements

- Audio preprocessing pipeline.
- Real-time inference deployment.
- Continuous monitoring for model drift.
- Feedback-driven retraining and updating.

g. Extension: Emotion-Aware Music Recommendation

As a natural extension, the detected emotions can be mapped to music preferences. By integrating with datasets like the Million Song Dataset, the system can recommend mood-matching songs in real-time, acting like a 'Spotify that senses your mood.'

h. Team Members

Aman
Chakrahar Reddy
Hemanth V