PROJECT TITLE: Human Emotion Detection from Voice

1. Abstract:

This project presents a voice-based emotion detection system that classifies human emotions from speech using machine learning. The solution leverages Mel Frequency Cepstral Coefficients (MFCCs) as audio features and employs a Random Forest classifier for emotion recognition. The system is trained on a labeled dataset of audio samples, with each sample annotated for its emotional content. A user-friendly web interface built with Streamlit allows users to upload voice recordings and receive real-time emotion predictions. This application demonstrates the potential of combining machine learning with audio signal processing for affective computing.

2. Introduction:

Emotions play a crucial role in human communication, and recognizing them from speech can greatly enhance human-computer interaction. Traditional emotion recognition methods often rely on visual cues, but speech offers a non-invasive and widely available alternative. This project aims to detect emotions from voice recordings using machine learning techniques. It extracts key audio features from speech signals and classifies them into predefined emotion categories such as happy, sad, angry, and surprised. The system not only trains a model using a supervised learning approach but also provides an interactive interface for real-world usage.

3. Tools Used:

- **Programming Language**: Python – Primary programming language for data processing, model training, and web deployment.
- **Libraries**: librosa, numpy, scikit-learn, joblib, streamlit, pydub
- **Dataset**: RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)
- **Model**: Random Forest Classifier
- **Deployment**: Streamlit

4. Steps Involved in Building the Project:

- **Dataset Collection**

  A pre-labeled speech dataset (e.g., RAVDESS or similar) was used, where each audio file corresponds to a specific emotional expression (e.g., happy, sad, angry).The dataset was organized into folders per speaker/actor.

- **Preprocessing and Feature Extraction**

  - Each audio file was loaded using the librosa library.
  - MFCC (Mel Frequency Cepstral Coefficients) features were extracted, which capture the short-term power spectrum of the audio signal.
  - Extracted features were stored as numerical vectors for training.

- **Model Training**

  - Split the data into training and test sets (80:20).
  - Trained a Random Forest classifier on the extracted features.
  - Achieved high accuracy on validation using classification_report() and accuracy_score().

- **Model Saving**

  - Serialized the trained model using joblib for later inference.

- **Web Application Development**

  - Built an interactive UI using Streamlit.
  - Implemented audio uploading and format conversion using pydub.
  - Extracted features from user-uploaded audio and predicted the emotion in real-time.

5. Conclusion:

This project successfully demonstrates a machine learning-based voice emotion detection system. By leveraging MFCC audio features and a Random Forest classifier, the system can accurately classify emotions from speech. The integration of a Streamlit web interface enhances usability, making it accessible to end-users for real-time predictions. The combination of data preprocessing, model training, and user-friendly deployment showcases the effectiveness of ML in affective computing. Future improvements could include real-time microphone input, multi-language support, and the use of deep learning for improved accuracy and robustness.