Project Report

Voice Emotion Recognition Using Machine Learning

Intern: Rania Khan

Internship Program: AI/ML

Internship Institute: Technik Nest Pvt Ltd

1. Problem Statement

Humans express emotions through their voice, and understanding these emotions can be helpful in fields like virtual assistants, therapy, and customer support. The goal of this project is to automatically detect a speaker's emotion from audio speech using machine learning techniques. Recognizing emotions from speech can enable machines to respond more empathetically and naturally.

2. Approach

We used the RAVDESS dataset, which contains emotional speech recordings from 24 professional actors(12 male,12 female). Each audio file represents one of eight emotions. The key steps included:

1. **Audio Preprocessing:** Extracting meaningful features from audio signals using MFCCs.
2. **Label Mapping:** Decoding emotion codes from filenames into readable labels like "happy" or "sad".
3. **Model Training:** Building a Random Forest Classifier to predict emotions based on the extracted features.
4. **UI Integration:** Adding a simple interface using Gradio to make predictions on user-uploaded audio files.

3. Tools and Libraries Used

- **Programming Language:** Python
- **Libraries:**
    - librosa (for MFCC feature extraction)
    - scikit-learn (for building the machine learning model)
    - Gradio (for user interface)
    - NumPy, Pandas, OS (for data handling)
- **Platform:** Google Colab
- **Dataset:** RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)

4. Key Results

- The model successfully classifies audio input into one of the eight emotions: *neutral, calm, happy, sad, angry, fearful, disgust, surprised*.
- A simple Gradio interface allows users to upload .wav files and receive emotion predictions.
- Accuracy was evaluated, and the classifier showed promising performance on unseen audio samples.

5. What I Learned

- Gained hands-on experience with audio processing and feature extraction using MFCC.
- Learned how to implement and evaluate a machine learning classifier (Random Forest) for audio-based emotion recognition.
- Explored how to build a user interface using Gradio, making the model interactive and user-friendly.
- Understood real-world applications of ML in areas like emotional AI, mental health tech, and voice assistants.

6. Applications and Future Scope

- **Applications:**
  - Emotion-aware virtual assistants (Alexa, Siri)
  - Voice-based mental health monitoring
  - Smart customer service agents
- **Future Enhancements:**
  - Real-time emotion detection using microphone input
  - Using deep learning models (CNN, LSTM) for improved accuracy
  - Multi-language support and robustness across accents