Speech Emotion Detection from Voice - Project Report

Abstract

Speech Emotion Recognition (SER) is a technique used to detect human emotions from voice signals. In this project, we developed a machine learning-based model that classifies emotions from audio samples using Mel-Frequency Cepstral Coefficients (MFCCs). The goal was to enable machines to understand and interpret human emotional states from speech, which has applications in virtual assistants, mental health analysis, and customer service.

Introduction

Human speech contains rich emotional information. Recognizing emotions from speech is challenging due to variations in pitch, tone, and speed. The objective of this project was to build a simple and effective speech emotion detection system using Python and machine learning. We used the RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) dataset, extracted features with Librosa, trained a Support Vector Machine (SVM) classifier, and optionally deployed the model with Streamlit.

Tools and Technologies Used

- Python: Programming language used for implementation
- Librosa: Audio processing and feature extraction (MFCC)
- Pandas: Handling datasets and dataframes
- Scikit-learn: Model training, testing, and evaluation
- Joblib: Saving/loading trained machine learning models
- Streamlit (optional): Creating a web-based UI for live emotion detection

Steps Involved in Building the Project

1. Dataset Preparation
   - Used RAVDESS dataset containing labeled `.wav` files
   - Parsed file names to extract emotion labels

2. Audio Preprocessing and Feature Extraction
   - Loaded audio using Librosa
   - Extracted MFCC features and calculated the mean for each sample

3. Feature Table Creation
   - Created a Pandas DataFrame with 13 MFCC features and emotion labels

4. Train-Test Split
   - Split data (80% train, 20% test) using train_test_split

5. Model Training
   - Trained an SVM classifier with a linear kernel on the MFCC features

6. Model Evaluation
   - Measured accuracy, precision, recall, and F1-score using classification report

7. Model Saving
   - Used joblib to save the trained model for future use

8. Web Application (Optional)
  - Built a Streamlit interface allowing users to upload `.wav` files and get emotion predictions

Conclusion

The speech emotion detection project successfully demonstrated how audio features can be used to classify human emotions. By extracting MFCCs and applying a machine learning classifier, we achieved reliable predictions on emotional speech. This system can be further improved with deep learning models, real-time audio input, and integration with AI platforms.