MINI-PROJECT SYNOPSIS
ON
SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

1. INTRODUCTION

Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch. This is also the phenomenon that animals like dogs and horses employ to be able to understand human emotion. SER is tough because emotions are subjective and annotating audio is challenging. In this Python mini project, we will use the libraries librosa, soundfile, and sklearn (among others) to build a model using an MLP Classifier. This will be able to recognize emotion from sound files. We will load the data, extract features from it, then split the dataset into training and testing sets. Then, we'll initialize an MLP Classifier and train the model. Finally, we'll calculate the accuracy of our model. This will be able to recognize emotion from sound files. We will load the data, extract features from it, then split the dataset into training and testing sets. Then, we'll initialize an MLP Classifier and train the model. Finally, we'll calculate the accuracy of our model.

2. OBJECTIVE OF THE PROJECT

As human beings speech is amongst the most natural way to express ourselves. We depend so much on it that we recognize its importance when resorting to other communication forms like emails and text messages where we often use emoji's to express the emotions associated with the messages. As emotions play a vital role in communication, the detection and analysis of the same is of vital importance in today's digital world of remote communication. Emotion detection is a challenging task, because emotions are subjective. There is no common consensus on how to measure or categorize them. We define a SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them. Such a system can find use in a wide variety of application areas like interactive voice based-assistant or caller-agent conversation analysis. In this study we attempt to detect underlying emotions in recorded speech by analysing the acoustic features of the audio data of recordings.

3. REQUIREMENTS ANALYSIS AND SPECIFICATION

librosa

librosa is a Python library for analysing audio and music. It has a flatter package layout, standardizes interfaces and names, backwards compatibility, modular functions, and readable code.

JupyterLab

JupyterLab is an open-source, web-based UI for Project Jupyter and it has all basic functionalities of the Jupyter Notebook, like notebooks, terminals, text editors, file browsers, rich outputs, and more. However, it also provides improved support for third party extensions.