Orange Level Problem Statement

Introduction

In the psychological well-being of humans, emotions are pertinent. It serves as an agent for communicating one's viewpoint or mental condition to others. Analysing audio data involves understanding the nature of sound, extracting meaningful features, and then applying machine learning techniques. Speech Emotion Recognition (SER) is one of the applications of audio analysis, which is used to detect the emotion from the speech signal. So kindly develop your own SER model on Emotion detection dataset (EDD) with the following problem statement.

Problem statement

In the vibrant town of DataScienceVille, a team of dedicated data scientists embarked on an ambitious project to build an advanced emotion classification model. They began by meticulously extracting features from their audio dataset, focusing on Mel-frequency cepstral coefficients (MFCC), Chromogram, Mel-scaled spectrogram, Spectral contrast, and Tonal Centroid. Ensuring their dataset was balanced across eight emotion categories—anger, sad, happy, neutral, calm, fearful, disgust, and surprised—they proceeded to visualize these features to understand their distributions and detect any anomalies.

Recognizing the potential impact of background noise, they applied sophisticated noise reduction techniques to enhance the quality of their audio samples. With clean and feature-rich data, they constructed a Multi-Layer Perceptron (MLP) model, meticulously fine-tuning hyperparameters such as the learning rate and activation functions, and selecting the Adam optimizer for its efficiency. The model was then trained and evaluated, with the team paying close attention to precision, recall, and F-score to gauge its performance. To ensure robustness, they employed K-fold cross-validation. To draw comprehensive insights, they compared the MLP model against Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers. Through this rigorous process, they found that while all models had their strengths, one among them excelled in capturing the complex patterns of emotional speech, ultimately achieving the best balance of accuracy, precision, recall, and overall performance.