**Emotion Extraction from Audio Data:**

Models Utilized:

1. superb/wav2vec2-base-superb-er (Emotion Classification Based on Voice)
2. distil-whisper/distil-small.en (Speech-Recognition)
3. j-hartmann/emotion-english-distilroberta-base (Emotion Classification Based on Language)

Roberta-Base-Model

Extract Features from Language (Text Data)

Adding two probabilities Array and Applying SoftMax layer on top of it & Selecting Higher probability Emotion

Return Probability array Target Class:

[‘neu’,’hap’,’ang’,’sad’]

Distil Whisper small Model

Speech-Recognition

Return Probability array Target Class:

[‘neu’,’hap’,’ang’,’sad’]

Extracting latent features from voice

Wave2vec2 Model

**Details For Whisper Model for Transcribe:**

For 16 Second Audio File Time taken by model for processing given below:

distil-whisper/distil-medium.en – [ Size = 800 mb, Processing Time = 27 sec]

distil-whisper/distil-small.en – [ Size = 400 mb, Processing Time = 12 sec]

Predicted Text:

Small Model: “Colonel, Lieutenant Kinderk ordered the code red didn't he because that's what you told Lieutenant Kendrick to do. Object! When it went bad, you cut these guys loose! Your Honor, you had more than inside of the money transferred. You've done your dialogue, but damn it! You coerced the doctor! Consider yourself! Lieutenant! Colonel, Justin!”

Medium Model: “Colonel, Lieutenant Kendrick ordered the code, Dinni, because that's what you told Lieutenant Kendrick to do. Object! And when it went bad, you cut these guys loose! Your Honor, you had margins. Your Honor, you had margins. Your Honor, you doctored the logbook. You coerced the doctor. Consider yourself in contempt. Colonel Joseph!”