Speech Emotion Recognition(SER)

Problem Statement:

The aim of this project is to construct and employ an LSTM classification model for Speech Emotion Recognition trained on a hybrid dataset.

GOAL: To accurately assess the emotional state of speakers in audio recordings.

HOW: Using the librosa library standardize and extract features voice signals from a dataset created for SER modeling and then build, train and test the SER model.

DELIVERABLE: MVP - A functioning LSTM model

Data:

[Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)](https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio)

[Toronto emotional speech set (TESS)](https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess)

[Surrey Audio-Visual Expressed Emotion (SAVEE)](https://www.kaggle.com/datasets/ejlok1/surrey-audiovisual-expressed-emotion-savee)

[Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D)](https://www.kaggle.com/datasets/ejlok1/cremad)

Outcomes

-

Future work:

- address class imbalances
- continue to tune and iterate or at least identify performance decline.
- Access other possible standardisations techniques, feature extraction and tuning methods.
- Utilize pretrained models such as Whisper, WavLM, and Wav2Vec 2.0, which can be fine-tuned for SER tasks.

Sources: