IDENTITY IMPERSONATION HACKATHON 2024

1. USER Stories
2. Prepare Data:
   1. Download files Public sources which are already labeled with following categories and cleaned already
      1. Human and AI (Deepfake)
      2. Data set used fake or real

Link : <https://bil.eecs.yorku.ca/datasets/>

\*Due to available time at hand - Accent labeled and Emotion labeled datasets were explored in limited way but couldn’t be included

* 1. Extract features from Audio data files which would help train the ML models . Focus on the most widely used ones – such as Mel Spectrogram, MFCC
  2. Split the data with labels into Train and Test split of 80:20

1. Create Preliminary Model sets for comparison
   1. Build Random Forest model.
      1. Run the model with training data and tune hyperparameters with Grid search
      2. Run the test data and check results
   2. Build SVM model
      1. Run the model with training data and tune hyperparameters
      2. Run the test data and check results
      3. Run the model with training data and tune hyperparameters
      4. Run the test data and check results
2. Model selection
   1. Run the trained models with same input datasets
   2. Choose the model with best performance
3. Hosting Model
   1. Prepare the selected model to be hosted on Cloud
   2. Add Pre transformation of the Raw MP3 Raw payload to destination feature extraction and feed to the model
   3. Test the hosted model with API requests
4. FEATURE Design
   1. We opted for the use of large public datasets as for training the model,

ii) MFCC has been used to extract features as it uses log function and Mel-filter to

mimic the human hearing system. Hence, its widely used for audio analysis

1. MODEL Design

A diagram of a diagram

Description automatically generated

1. Comprehensive Design

The problem statement also included Emotion, Accent which couldn’t be covered in terms of implementation. However, our team thought upon the approach below, which can be considered to analyze emotion/Accent also along with the audio classification. As part of this process, publicly available datasets on emotion and accents will be required which can be used to train different ML or DL models.

Final implementation will include 3 separate classifications on the same audio sample.

