

Artificial Intelligence – Spring 2024

Semester Project - Phase II

Speaker Age Prediction using Regression Model

Deadline: 9th May, 2024

General Instructions:

- The project can be done in pairs or individually.
- The deadline will not be extended under any circumstances.
- All the components of the algorithm being used should be mentioned in the PDF report along with the rationale of choosing that particular technique.
- Evaluation will be done on the basis of a demo. Anyone failing to appear for the demo or unable to give up to the mark demonstration will be awarded zero.
- Evaluation will be individual.
- Data set has been provided along with the project statement.
- You can use audio processing libraries like Librosa and Python Speech Features etc.
- **You cannot use the built-in function of regression provided in scikit learn.**

Problem Statement:

The task is to predict the age of speakers from audio recordings. This can be useful in various applications such as voice-based personalization, targeted advertising, and forensic voice analysis.

Dataset:

- You have a dataset of audio recordings with associated labels for speaker age.
- The dataset includes a diverse range of speakers across different ages and genders.
- Extract acoustic features from the audio recordings, such as pitch, formant frequencies, intensity, duration, and spectral features.

What you have to do?

Step 1: Feature Extraction:

- Extract relevant acoustic features from the audio recordings using signal processing techniques.
- Commonly used features for speaker age prediction may include:
 - **Pitch:** Fundamental frequency (F0) of the voice, which is related to the perceived pitch or tone of the voice.
 - **Formant Frequencies:** Resonant frequencies of the vocal tract, which provide information about the size and shape of the vocal tract and are correlated with vowel sounds.
 - **Intensity:** Energy or loudness of the voice, which can vary with the speaker's age.
 - **Duration:** Length of speech segments, which may vary based on factors such as speaking rate and age-related changes in speech patterns.
 - **Spectral Features:** Frequency content of the voice signal, including features derived from the Fourier transform or spectrogram.

After extracting these features, your next step is to organize them into data frames for efficient analysis. Once you've created the data frames, you have to visualize all the extracted features using plots to identify any discernible trends.

Step 2: Data Preprocessing:

- Filling null values, also known as imputation, is a crucial step in data preprocessing to ensure that missing data does not hinder analysis or modeling. Evaluate the data and look for missing values. [not manually use code to do that]
- If there are any missing values, fix them using any technique (Mean/Median/Mode Imputation, Forward Fill/Backward Fill, drop NULL values)
- Normalize or scale the extracted features to ensure they are on a similar scale and have comparable magnitudes.
- You can use any scaling and normalization technique.
- Dataset is already divided into training and testing sets for model evaluation.

Step 3: Model Training:

- Choose a regression algorithm suitable for predicting continuous-valued attributes such as age.
- Train the regression model using the extracted acoustic features as input and the corresponding age labels as target variables.
- Common regression algorithms that can be used include linear regression, support vector regression, decision tree regression, or neural network regression.

Step 4: Model Evaluation:

- Evaluate the trained regression model on the testing set using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared (R^2) coefficient of determination.
- Assess the model's performance in predicting speaker age accurately across different age groups.