Audio Feature Extraction and Analysis Report

Introduction:

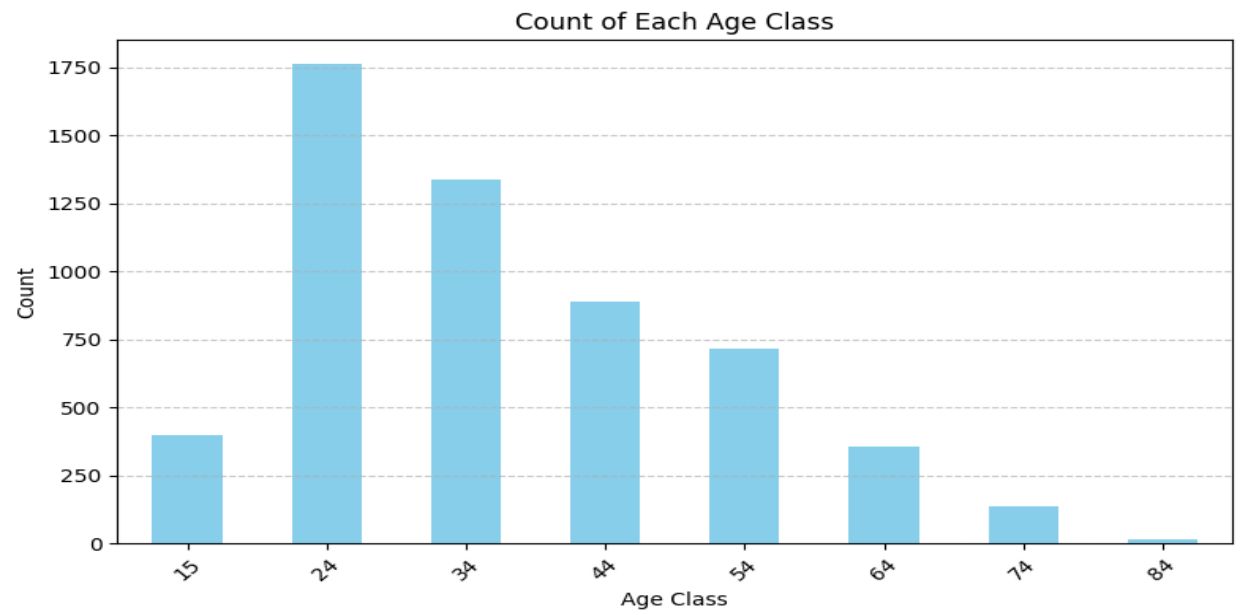
This report details the process of extracting audio features from speech samples, analyzing their relationship with age, and building a predictive model to estimate the age of speakers based on these features.

Feature Extraction:

- Utilized the **librosa** library to extract various audio features from speech samples.
- Extracted features include duration, pitch statistics (mean and standard deviation), formant frequencies, intensity, spectral centroid, MFCCs, zero-crossing rate, tempo, and spectral rolloff.
- These features were computed for both training and testing datasets.

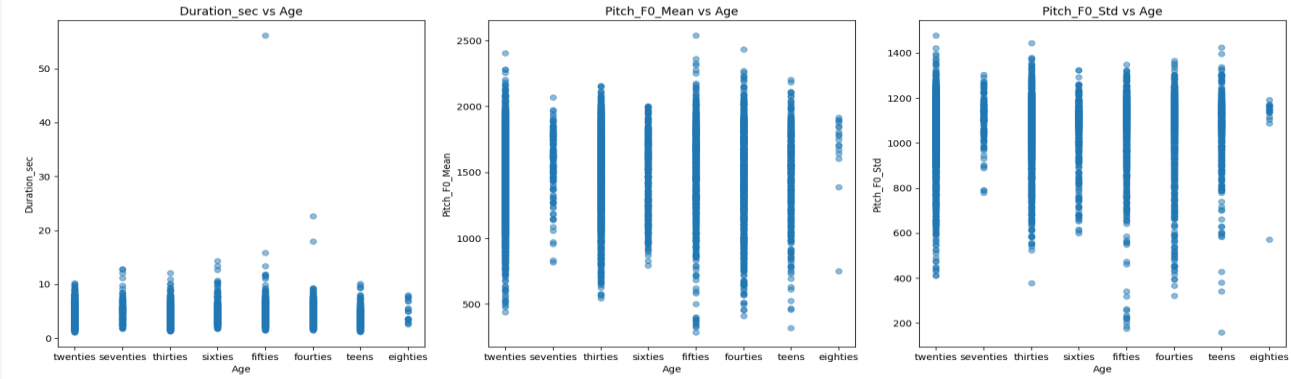
Data Preparation:

- Loaded training and testing datasets containing speech samples and corresponding age labels.
- Preprocessed the datasets to align them for feature extraction and model training.
- Cleaned the datasets by removing NaN values and irrelevant columns.



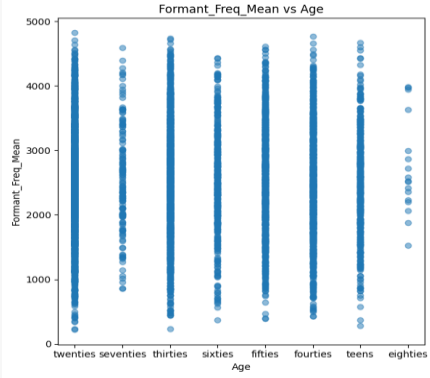
Feature Analysis:

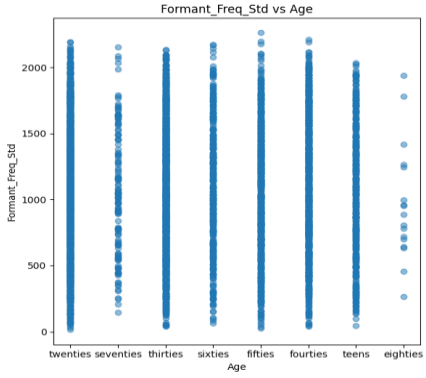
- Investigated the relationship between each audio feature and age using scatter plots.
- Analyzed how different features correlate with age and identified potential predictors.

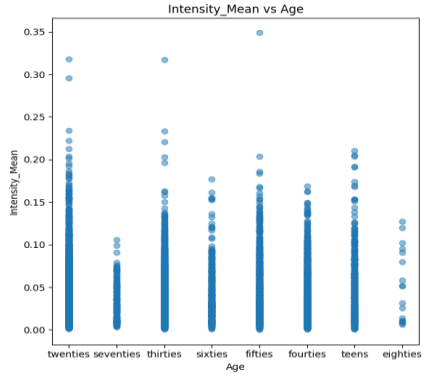

Duration_sec vs Age
Duration_sec
Age
twenties seventies thirties sixties fifties fourties teens eighties

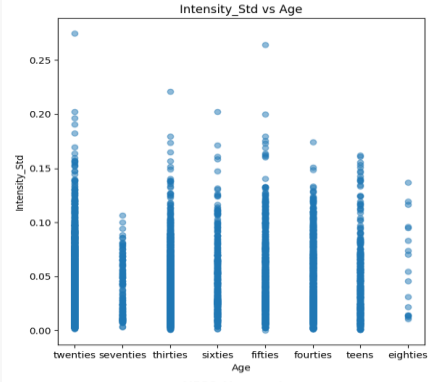
Pitch_F0_Mean vs Age
Pitch_F0_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties

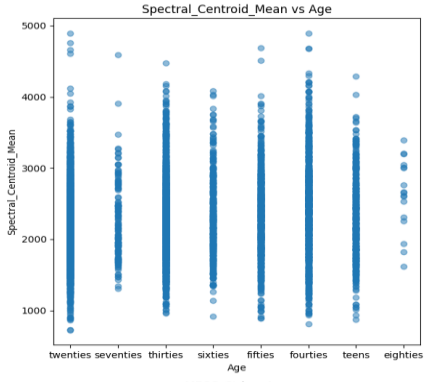
Pitch_F0_Std vs Age
Pitch_F0_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties

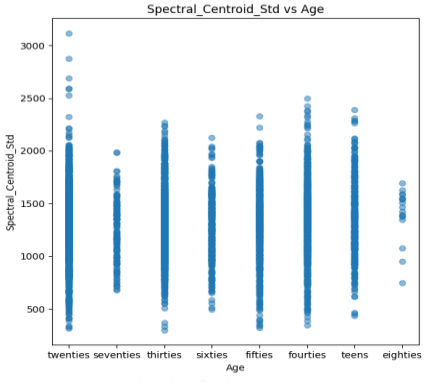

Formant_Freq_Mean vs Age
Formant_Freq_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties

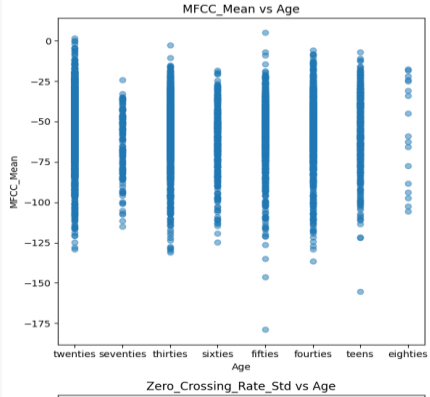

Formant_Freq_Std vs Age
Formant_Freq_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties

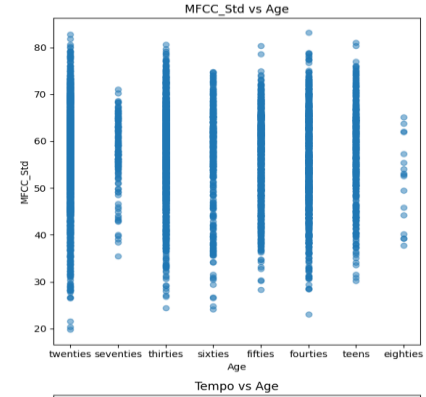

Intensity_Mean vs Age
Intensity_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties

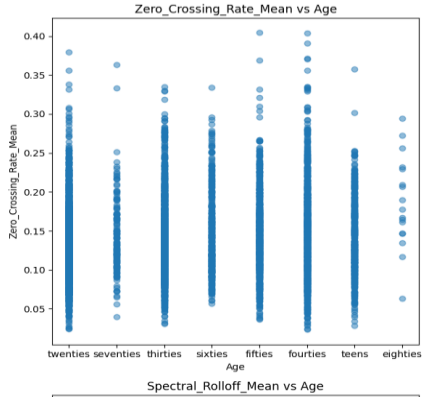

Intensity_Std vs Age
Intensity_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties

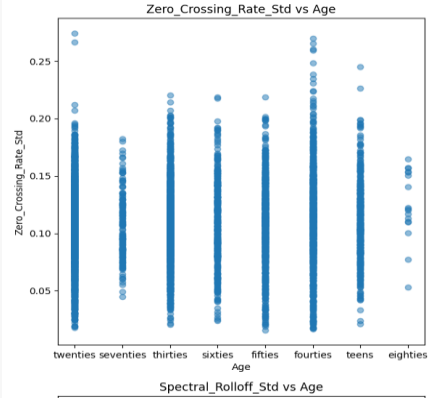

Spectral_Centroid_Mean vs Age
Spectral_Centroid_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties

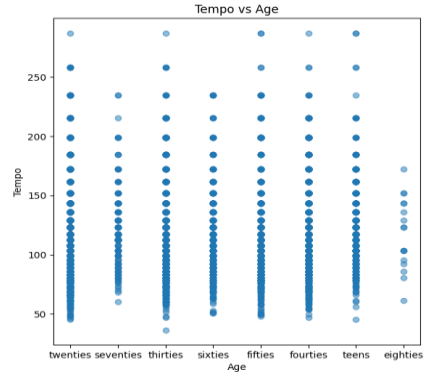

Spectral_Centroid_Std vs Age
Spectral_Centroid_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties

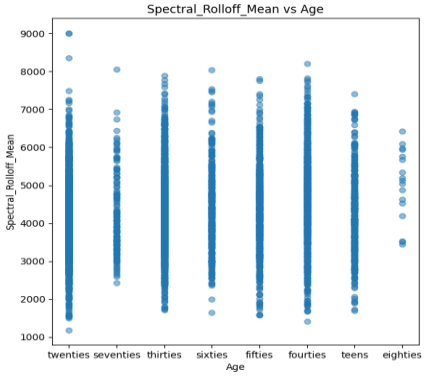

MFCC_Mean vs Age
MFCC_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties


MFCC_Std vs Age
MFCC_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties


Zero_Crossing_Rate_Mean vs Age
Zero_Crossing_Rate_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties


Zero_Crossing_Rate_Std vs Age
Zero_Crossing_Rate_Std
Age
twenties seventies thirties sixties fifties fourties teens eighties


Tempo vs Age
Tempo
Age
twenties seventies thirties sixties fifties fourties teens eighties


Spectral_Rolloff_Mean vs Age
Spectral_Rolloff_Mean
Age
twenties seventies thirties sixties fifties fourties teens eighties

Spectral_Rolloff_Std vs Age

Model Building:

- Utilized a gradient descent approach to train a linear regression model to predict age based on audio features.
- Defined a loss function to minimize the mean squared error between predicted and actual age values.
- Implemented gradient descent to iteratively update model coefficients and intercept to minimize the loss.

Performance:

- Evaluated the model's performance using mean squared error (MSE), R-squared, and accuracy metrics.

Learning Rate Vs Accuracy:

lr = 0.00000001

Mean Squared Error (MSE) on test data: 421.52445589292154

R-squared (coefficient of determination) on test data: -0.9470258169770482

Accuracy on test data (within +/- 3 years): 19.13099870298314

lr = 0.000000001

Mean Squared Error (MSE) on test data: 268.2649599860618

R-squared (coefficient of determination) on test data: -0.239118621899983

Accuracy on test data (within +/- 3 years): 15.304798962386512

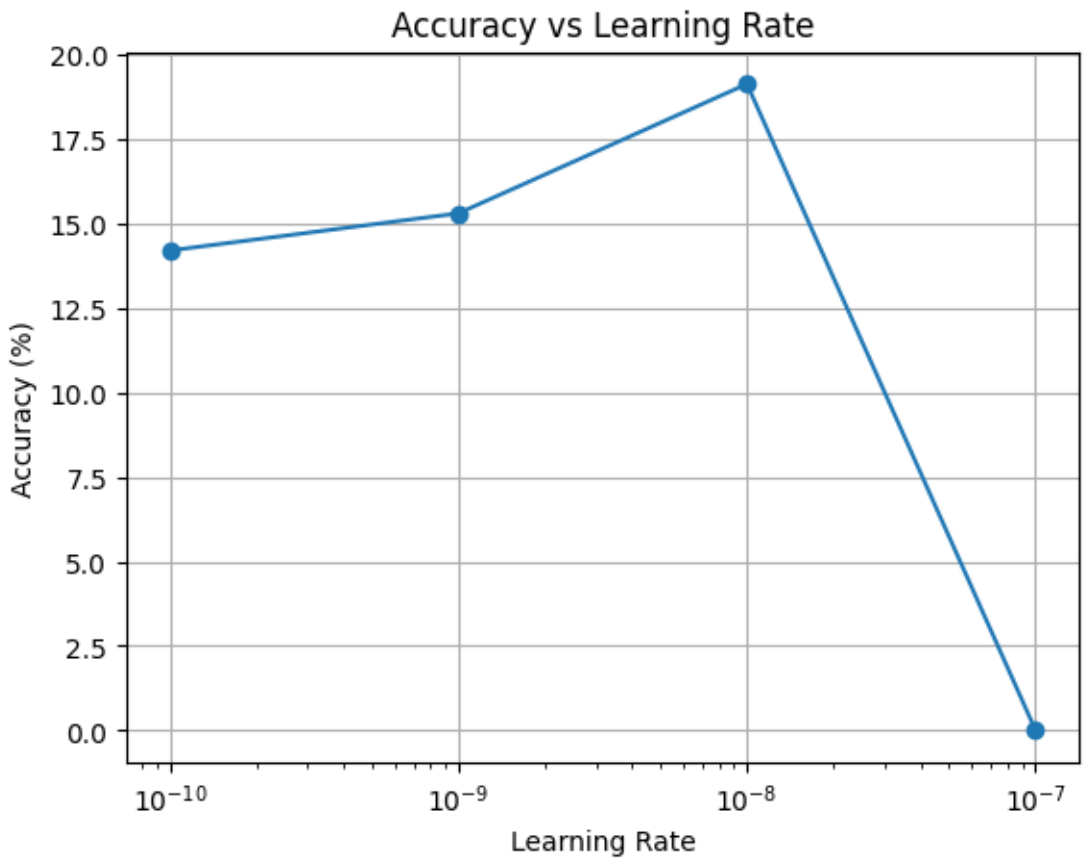
–

lr = 0.0000000001

Mean Squared Error (MSE) on test data: 255.88316330034516

R-squared (coefficient of determination) on test data: -0.18192697507943634

Accuracy on test data (within +/- 3 years): 14.202334630350194


Accuracy vs Learning Rate
Accuracy (%)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
10^{-10}
10^{-9}
10^{-8}
10^{-7}
Learning Rate