**Title: Report on Audio Feature-Based Age Prediction**

**1. Introduction:**

This section provides an overview of the aim of the report, which is to develop a machine learning model capable of predicting the age of individuals based on features extracted from audio signals, specifically speech. It sets the context for the subsequent discussion by highlighting the importance and potential applications of such a model, such as in voice-controlled systems or demographic analysis.

**2. Data Collection and Preprocessing:**

* **Data collection:** Mention the sources of the datasets used for training and testing the model, which are "cv-valid-test.csv" and "truncated\_train.csv". Briefly describe the contents of these datasets.
* **Feature extraction:** Explain the process of extracting various features from audio signals, including pitch, intensity, duration, formants, and spectral features like centroid, bandwidth, etc. Describe each feature extraction technique briefly.
* **Data cleaning:** Outline the steps taken to clean the data, including handling missing values (dropping NaNs) and normalization techniques used (z-score and min-max scaling).

**3. Model Training:**

* **Linear Regression Model:** Describe the implementation of the linear regression model using the normal equation approach. Explain how the model is trained using the extracted audio features and the target variable, which is age.
* **Evaluation Metrics:** Discuss the evaluation metrics used to assess the model's performance, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score.

**4. Model Performance Assessmen**t:

Calculate and interpret the evaluation metrics obtained from the trained model. Discuss what each metric signifies in terms of the model's accuracy and predictive capability.

Assess the model's performance across different age groups by calculating MAE for each group. Interpret the results to understand how well the model performs for different age ranges.

**5. Comparison with Library Implementation:**

Describe the process of repeating the model training using scikit-learn library's linear regression implementation. Highlight any differences in implementation or results compared to the custom implementation. Visualize the actual vs. predicted ages using graphical representations and discuss any observations or insights gained from the comparison.

**6. Conclusion:**

Summarize the key findings and implications of the study. Emphasize the effectiveness of the developed model in predicting age based on audio features. Discuss potential applications and future directions for further improvement or refinement of the model.

**7. Recommendations:**

Provide suggestions for future research or development based on the insights gained from the study. This could include exploring additional features, optimizing model parameters, or deploying the model in practical applications.Highlight any areas where further investigation or experimentation could yield valuable insights or improvements in the model's performance.

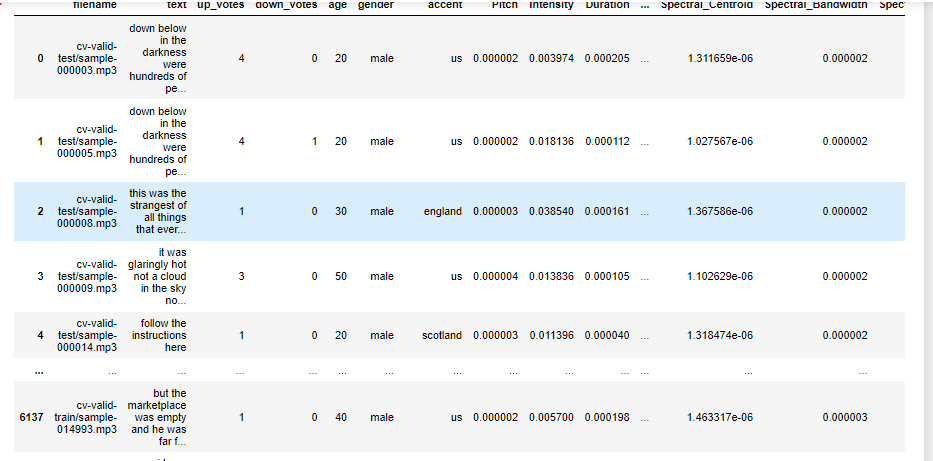
**8. Comparison with Library Implementation:**

To ensure the robustness and correctness of the developed model, a comparison was made with the implementation provided by the scikit-learn library.

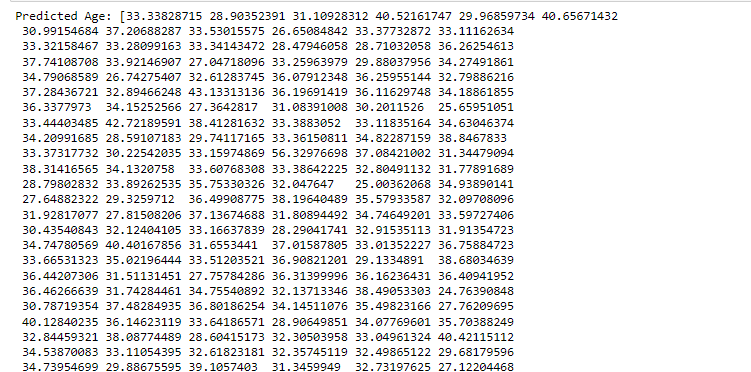
* **Custom Implementation:** The model was trained using a linear regression approach implemented without using external libraries. This involved deriving coefficients using the normal equation and making predictions based on the extracted audio features.
* **Library Implementation:** The same dataset was used to train a linear regression model using the scikit-learn library. This involved initializing and training the model using the Linear Regression class provided by scikit-learn and making predictions on the test data.
* **Visualization:** Actual vs. predicted ages were visualized using scatter plots for both custom and library implementations. The plots provided a visual comparison of the predictions made by each implementation method.
* **Performance Evaluation:** Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score were calculated for both implementations. This facilitated a quantitative comparison of the performance of the custom and library-based models.
* **Observations:** Any notable similarities or differences in performance between the two implementations were observed and discussed. Insights gained from the comparison were used to validate the correctness of the custom implementation and assess its effectiveness relative to a standard library-based approach.

**9. Screenshot:**

* **Merging dataset of both files with data cleaning and scaling and normalization**

****

* **Prediction age:**

****

* **Graph:**

**A graph with blue dots and a dotted line

Description automatically generated**

* **Accuracy:**

**A screenshot of a computer error

Description automatically generated**

* **Assessment:**

**A screenshot of a computer

Description automatically generated**