

Grammar Scoring Engine Evaluation Report

1. Objective

The main objective involved developing a trained model to forecast grammar quality scores from (0 to 5) which would evaluate spoken English audio recordings. The proposed scores will match MOS Likert ratings by evaluating grammatical accuracy together with complexity.

2. Approach Summary

- **Model Type:** A regression model was trained to output continuous grammar scores.
- **Feature Extraction:** Audio features such as MFCCs were extracted from each .wav file using librosa.
- **Regression Head:** A fully connected regression layer mapped the audio embeddings to grammar scores.

3. Preprocessing Steps

- **Audio Normalization:** All audio files were standardized to a fixed sample rate.
- **Padding/Truncation:** Applied to ensure consistent input shape.
- **Feature Extraction:** MFCCs (Mel Frequency Cepstral Coefficients) were computed to represent audio characteristics.

4. Model Architecture

- **Input:** .wav audio files (45–60 seconds).
- **Feature Extraction:** MFCCs computed using librosa.
- **Model:** A simple neural network with:
 - Input layer matching MFCC shape.
 - Dense layers with ReLU activation.
 - Output layer producing a single float (grammar score).
- **Loss Function:** Mean Squared Error (MSE).
- **Evaluation Metric:** Root Mean Squared Error (RMSE).

5. Evaluation Results on Submission Data

- **Number of Samples:** 204
- **Mean Predicted Score:** 3.21
- **Min Score:** 2.07
- **Max Score:** 4.82
- **Standard Deviation:** 0.41