# Grammar Scoring Engine

(Report)

**Objective:**

The goal of this project is to build a **Grammar Scoring Engine** that evaluates spoken English audio samples and assigns a **continuous grammar score between 0 and 5**, based on a provided rubric. The model is trained on .wav files along with MOS Likert grammar scores.

**Preprocessing Steps:**

**Audio Conversion**

* The original .wav files failed to load due to encoding issues.
* We used pydub to convert all audio files to PCM 16-bit WAV format to ensure compatibility with librosa.

**Feature Extraction**

* We extracted **MFCC (Mel Frequency Cepstral Coefficients)** using librosa, a standard feature set for audio modelling.
* Each audio file was transformed into a fixed-length 13-dimensional feature vector using the **mean of MFCCs** over time.

**Pipeline Architecture**

* **Model Input:** 13-D MFCC feature vector for each audio file.

 **Model Used:** **RandomForestRegressor** from **sklearn** for its robustness and ability to model non-linear relationships.

* Training **Size:** 444 audio samples.
* Train**/Test Split:** 80/20 for validation.
* **Evaluation Metrics:**
* **RMSE** (Root Mean Squared Error)
* **R² Score** (Coefficient of Determination)

**Evaluation Results:**

| **Metric** | **Value** |
| --- | --- |
|  |  |
| RMSE | 1.0857 |
| R² Score | 0.1353 |

**Interpretation:**

* **RMSE ≈ 1.08** indicates that predictions are, on average, off by ~1 grammar point (on a 0–5 scale).
* **R² Score ≈ 0.135** shows the model explains ~13.5% of the variance in grammar scores — better than random, but with significant room for improvement.

**Improvement Areas:**

* **Feature Engineering:** Introduce additional features such as pitch, zero-crossing rate, and spectral contrast.
* **Hyperparameter Tuning:** Use cross-validation to fine-tune the Random Forest parameters.
* **Advanced Models:** Explore deep learning models or pretrained audio embeddings (e.g., Wav2Vec2, OpenL3, etc.).

**Conclusion:**

This baseline system demonstrates a working pipeline from **raw audio to grammar scoring**, using classic audio processing and machine learning techniques. While the current performance is modest, the foundation is solid and ready for further enhancement.