EEG-BASED GAME RATING CLASSIFICATION (PEGI & ESRB)

Dataset Overview

This dataset contains EEG-derived features extracted from gameplay sessions. The objective is to **predict age-based content ratings** for video games using neural signals.

Data Column	**Description**
PEGI	Target rating label (e.g., 3, 7, 12, 16, 18) based on the **Pan-European Game Information** system
ESRB	Target rating label (e.g., E, T, M) based on the **Entertainment Software Rating Board**
EEG Features	2268 numerical features derived from EEG signals: statistical measures (mean, RMS, std, kurtosis, etc.) over time-frequency windows and channels
Total Samples	900
Data Issues	Some feature columns contain **missing values (NaN)**

Objective

You are required to build **two classification models**:

1. Predict the **PEGI** label (numerical multi-class classification).
2. Predict the **ESRB** label (categorical multi-class classification).

Each model must operate on **imputed EEG features**, meaning you must handle missing values **explicitly and appropriately** before training.

Restrictions

- Do **not drop rows or columns** with missing data unless explicitly justified and explained.
- Do **not use dimensionality reduction techniques** (e.g., PCA, UMAP, Autoencoders).
- You **must use imputation techniques** (e.g., SimpleImputer, KNNImputer, IterativeImputer) to fill in missing values.
- You should apply **feature selection or feature importance ranking** to reduce the number of input variables.

Tips & Hints

- Apply imputation **before** splitting data into training/test sets for consistency (or separately on training/test with same method).
- Try different values for k in KNN imputer (e.g., 3, 5, 10).
- Select the **smallest number of features** that still provides **high accuracy**.
- Make sure to handle the PEGI (numerical) and ESRB (categorical) labels **separately** and evaluate both.