Assignment 4

Food Item Classification

Prepare a clean, well-documented Kaggle notebook that enhances the performance of a previously implemented classification task on a food nutrition dataset. The original task used Multinomial Logistic Regression (MLR) and its regularized forms (Ridge, Lasso, ElasticNet) to classify food items into three dietary categories: **In Moderation**, **Less Often**, and **More Often**.

Task Description

The aim of this assignment is to **refine and improve classification accuracy** and **evaluation metrics** by addressing data imbalance, performing advanced preprocessing, and comparing modeling strategies **(Multinomial vs. One-vs-Rest).** Your final notebook should reflect thoughtful experimentation and interpretation of your results

You can access the Notebook from here: Notebook

Requirements:

1. Dataset Description and Problem Definition

- Provide a brief description of the dataset.
- Define the problem.

2. Exploratory Data Analysis (EDA)

- Descriptive statistics for numerical features.
- Visualizations to explore distributions and relationships.
- Identify issues like skewness or outliers.

3. Preprocessing

- Address class imbalance using techniques such as:
  - SMOTE
  - Random oversampling or undersampling.
  - Class weighting in logistic regression.

4. Modeling and Evaluation

- Train and evaluate the following models using **the balanced dataset:**
  - Multinomial Logistic Regression and its regularized models (if applicable).
  - One-vs-Rest Logistic Regression

- Use these evaluation metrics:
  - Accuracy
  - Precision, Recall, F1-Score (macro and weighted)
  - Confusion Matrix
  - ROC-AUC **(per class, if possible)**
- Plot:
  - Confusion matrices
  - Precision-Recall curves
  - ROC curves
  - Feature importance or coefficient comparison (for interpretability)

5. Model Comparison and Interpretation

- Compare Multinomial vs. One-vs-Rest strategies.
- Discuss:
  - How regularization affected model performance and generalization.
  - How data balancing methods improved classification results.
- Justify the best-performing model using **quantitative metrics**.

Instructions:

- All codes should be written **independently** without using AI tools.
- Submit your notebook by **uploading it to Kaggle** and sharing the public notebook link.
- **Inside your GitHub repository named IEEE-ML-2025, add a folder named Food_Item_Classification,** which must contain:
  - Your **.ipynb** notebook.
  - A **README.md** file specific to the project.
  - A **plots/ folder** to store relevant visualizations and metrics.

Expected Deliverables:

- **Your GitHub repository link**
- **Your public Kaggle notebook link** demonstrates your full implementation.

Submission Deadline: Wednesday 30/7/2025 before 11:59 PM