



## Overview

This assignment guides you through building and improving a machine learning model to predict the type of Olympic medal (Gold, Silver, Bronze) using a real dataset. You'll progress through stages: from basic data preparation to building a neural network, then enriching the dataset to improve your model.

### Level 1: Exploratory Data Analysis (EDA) & Preprocessing

Goal: Understand and prepare the data for machine learning.

Tasks:

- Load the dataset
- Evaluate relevant features
- Visualize distribution of `Medal` column
- Encode categorical and target variable

### Level 2: Baseline ML Models

Goal: Build and evaluate a simple classification model.

Tasks:

- Split data into training and test sets (e.g., 80/20)
- Train Logistic Regression and/or Decision Tree Classifier
- Perform Model Evaluation

### Level 3: Multilayer Perceptron (MLP)

Goal: Train a simple neural network using Keras.

Tasks:

- Normalize numerical features
- Convert target to categorical using `to\_categorical`
- Build an MLP with Keras
- Train using validation split
- Plot training/validation accuracy and loss
- Evaluate model accuracy

#### Level 4: Feature Enrichment & Impact Analysis

Goal: Enhance model performance by adding domain-driven features.

Tasks:

- Create these 2 new features:

1. **`country\_medal\_count`**: Represents the **total number of medals** won by each country across all years and sports. This uses group by to count how many medals each country has won overall and then maps this count back to each row using the Country column.

Country	country_medal_count
United States	1992
Italy	486
Soviet Union	1021

2. **`sport\_popularity`**: Represents the **total number of medals awarded in each sport** — a proxy for how prominent or competitive a sport is. This calculates how many medals have been awarded per sport (globally), and maps this back to each row using the Sport column.

Sport	sport_popularity
Diving	180
Athletics	1,XXX
Rowing	3XX

3. Add them to your dataset and re-train the MLP

- Compare performance before and after.
- And reflect:
  - Did accuracy improve?
  - Did the model train faster/slower?
  - Which feature seemed more predictive?

#### Submission Checklist

- Python script or Jupyter Notebook with code and comments
- Performance comparison (before/after enrichment)
- Short reflection paragraph on results and learnings

#### Evaluation Rubric (Total: 10 points)

Level	Criteria	Points
Level 1: EDA & Preprocessing	Data is correctly cleaned, visualized, and encoded.	2
Level 2: Baseline ML Model	Baseline model is implemented and evaluated properly.	2
Level 3: MLP Model	MLP is well-constructed, trained, and analyzed with results.	3
Level 4: Feature Enrichment	New features are added, model is re-evaluated, and impact is discussed.	3