Predicting Olympic Medal Counts

Project Overview

In this project, you will explore basic machine learning (ML) and deep learning (DL) techniques to predict the number of Olympic medals a country will win. The dataset provided includes features such as GDP, population, and sports index, along with the actual number of medals won. You will build and evaluate different models to understand which factors are most influential in predicting Olympic success.

Dataset Description

The dataset (check folder for Data) includes:

- iso: Country ISO code
- ioc: International Olympic Committee code
- name: Country name
- continent: Continent of the country
- population: Population of the country
- gdp: Gross Domestic Product (GDP) of the country
- olympics_index: An index indicating the country's overall performance in the Olympics
- sports_index: An index indicating the country's sports infrastructure and support
- olympicsIndex: A calculated index related to Olympic performance
- sportsIndex: A calculated index related to sports
- total: Total number of medals won
- gold: Number of gold medals won
- silver: Number of silver medals won
- bronze: Number of bronze medals won

Project Steps

1. **Data Preprocessing**
    - **Data Cleaning**: Handle missing values, check for duplicates, and address any inconsistencies in the data.
    - **Feature Engineering**: Create or modify features as necessary (e.g., interactions between GDP and population). Normalize or standardize features if needed.
    - **Data Splitting**: Divide the dataset into training and testing sets (e.g., 80% training, 20% testing).

2. **Exploratory Data Analysis (EDA)**
    - **Descriptive Statistics**: Compute basic statistics for numerical features (mean, median, standard deviation).
    - **Visualizations**: Create plots to visualize the relationships between features and medal counts. Examples include scatter plots, histograms, and correlation matrices.
    - **Feature Analysis**: Determine which features have the strongest correlation with the number of medals won.
3. **Machine Learning Models**
    - **Linear Regression**: Implement a simple linear regression model to predict the total number of medals. Evaluate its performance using metrics such as Mean Absolute Error (MAE) and R-squared.
    - **Decision Trees**: Use decision tree regression to model non-linear relationships. Assess the model's performance and interpret its predictions.
    - **Random Forest**: Apply random forest regression to improve prediction accuracy and handle complex feature interactions.
4. **Deep Learning Models**
    - **Neural Network**: Build a basic feedforward neural network using a library like TensorFlow or Keras. Experiment with different architectures, such as varying the number of layers and neurons.
    - **Hyperparameter Tuning**: Adjust hyperparameters (e.g., learning rate, number of epochs) to optimize model performance.
5. **Model Evaluation**
    - **Performance Metrics**: Evaluate the performance of all models using MAE, Mean Squared Error (MSE), and R-squared.
    - **Model Comparison**: Compare the performance of machine learning models (Linear Regression, Decision Trees, Random Forest) with deep learning models (Neural Networks).
6. **Interpretation and Insights**
    - **Feature Importance**: Analyze which features are most influential in predicting the number of medals.
    - **Model Interpretation**: Discuss the results of different models and what they reveal about the factors affecting Olympic success.
7. **Documentation and Presentation**
    - **Report**: Write a detailed report covering data preprocessing, model building, evaluation, and insights. Include visualizations and explanations.

- **Presentation**: Prepare a presentation summarizing the key findings and model performance. Include visuals and key takeaways.

Deliverables

- Python scripts or Jupyter notebooks with data preprocessing, model training, and evaluation.
- A detailed project report with methodology, findings, and insights.
- A presentation slide deck summarizing the project (upload pdf).

Files to be uploaded: 1 Notebook + 1 PDF (your presentation)

Note: This is just a hint structure for your project, you can modify your steps and you all can even create a classification model.