Olympics 2024 Data Analysis Report

Prepared by: Saanvi M J

Date: 5/9/2024

The report provides a comprehensive analysis of the data related to the Olympics 2024. It includes detailed insights obtained through various data processing and machine learning techniques, such as Linear Regression, K-Means Clustering, and Decision Tree Classification. This report aims to give a clear understanding of the patterns and trends within the dataset, particularly focusing on the performance of different countries based on their medal tally.

1. Data Pre-processing:

Dataset Description: The dataset used for this analysis includes information on the performance of various countries in the Olympics 2024. Key columns include the 'Country', 'Country Code', and the number of medals (gold, silver, bronze) won by each country.

Data Pre-processing: The data was pre-processed to encode categorical variables such as 'Country' and 'Country Code' using Label Encoding. Missing values were handled appropriately to ensure the integrity of the analysis.

---

2. Analysis Engine:

Linear Regression Analysis:

- o **Objective:** To predict the medal tally based on various features.
- o **Method:** A Linear Regression model was trained on the dataset.
- o **Results:** The regression model provided a good fit, with the predicted values closely matching the actual medal counts. The model score, which indicates its accuracy, is highlighted in the analysis.
- o **Visuals:** A scatter plot showing actual vs. predicted values with a regression line is included.

K-Means Clustering:

- o **Objective:** To cluster countries based on their performance.
- o **Method:** The K-Means Clustering algorithm was used to categorize countries into three clusters.
- o **Results:** The countries were successfully grouped into clusters, and the centroids were identified.
- o **Visuals:** A scatter plot showing the clusters and centroids is included.

Decision Tree Classification:

- o **Objective:** To classify countries based on their likelihood of winning a high number of medals.
- o **Method:** A Decision Tree Classifier was trained on the dataset.
- o **Results:** The feature importance analysis showed which factors were most critical in determining a country's performance.
- o **Visuals:** A bar chart displaying the importance of different features is included.

---

3. Report Generation:

Statistics:

A detailed summary of the dataset's statistics, including measures like mean, median, and standard deviation, is provided. These statistics help in understanding the overall distribution and tendencies in the data.

Correlation Matrix:

The correlation matrix is included to show the relationships between different variables in the dataset. It highlights which factors are positively or negatively correlated, providing deeper insights into the dynamics of the dataset.

4. User Interaction:

The AI Employee for Data Analysis is designed to assist users in performing comprehensive data analysis on the Olympics 2024 dataset. The system provides a simple user interface where users can interact with various functionalities, including loading data, pre-processing it, running analyses, and generating reports. This report outlines the interaction process, detailing each step and the system's responses.

Upon starting the AI Employee, the user is greeted with a friendly welcome message:

- **Message:** "Hey, Welcome to the AI Employee for Data Analysis!"
- **Options Provided:**
    o Press 1 to load data.
    o Press 2 to pre-process the data.
    o Press 3 to run analysis.
    o Press 4 to generate the report.
    o Type 'exit' to quit.

This ensures that the user understands the available options and can easily navigate the interaction process.

Loading Data:

- **User Action:** The user is prompted to press '1' to load the data.
- **System Response:** Once the user selects this option, the system loads the dataset from the specified CSV file and confirms the successful loading of the data with the message:
    o "Data loaded successfully."
    o In case of failure, the system responds with:
    o "Failed to load data."

Pre-processing Data:

- **User Action:** After loading the data, the user can press '2' to pre-process it.

- **System Response:** The system pre-processes the data by encoding categorical variables and handling missing values. A confirmation message is provided:
    - "Data pre-processing completed."
    - If data is not loaded first, the system prompts the user:
    - "Please load data first."

Running Analysis:

- **User Action:** The user can press '3' to run the analysis on the pre-processed data.
- **System Response:** The system performs various analyses, including Linear Regression, K-Means Clustering, and Decision Tree Classification. After the analysis, the system outputs the results and, if successful, states:
    - "Linear Regression Model Score: [score]."
    - "K-Means Score: [centroids]."
    - "Decision Tree Model Score: [score]."
    - If data is not loaded or pre-processed, the system prompts:
    - "Please load and pre-process data first."

Generating Report:

- **User Action:** The user can press '4' to generate a comprehensive report based on the analysis.
- **System Response:** The system generates a report, including visualizations and summaries, and saves it as a PDF. The user is informed:
    - "Report generated successfully."
    - If data is not loaded, the system suggests:
    - "Please load and pre-process data first."

Exiting:

- **User Action:** The user can type 'exit' at any point to quit the interaction.
- **System Response:** The AI Employee gracefully ends the session with the message:
    - "Tata!"

Throughout the interaction, the AI Employee ensures that the user is guided and supported, providing feedback at every step. It effectively handles errors by notifying the user of necessary actions (e.g., loading data first) and ensures that the interaction is smooth and intuitive.

Key Features:

- **Clear Guidance:** The system provides clear instructions for every option, making it user-friendly.
- **Error Handling:** The AI Employee is equipped to handle common user errors (e.g., not loading data first) and prompts corrective actions.
- **Comprehensive Feedback:** Users are informed of the success or failure of each action, ensuring transparency in the interaction process.

Conclusion:

This report has provided a thorough analysis of the Olympics 2024 dataset, revealing significant patterns in the performance of different countries. The machine learning models used have successfully demonstrated their ability to predict outcomes and classify data points based on various features. These insights can be valuable for future predictive analysis and strategic decision-making in sports analytics.

The AI Employee for Data Analysis provides an efficient and user-friendly interface for performing complex data analysis tasks. Through clear instructions and comprehensive feedback, it ensures that users can easily navigate the data analysis process, from loading data to generating final reports. The interaction design prioritizes user experience, making the system accessible even to those with minimal technical expertise.