National College of Ireland

National College of Ireland
Programming for Artificial Intelligence (H9PAI)

Programme: MSc in Artificial Intelligence - MSCAI1A
Continuous Assessment (30%)

Release Date: 6th November 2024
Submission Date: 10th December 2024

Project Report

Title: NASA APOD Data and Iris Dataset

Abstract: This report describes the project in detail, which includes extracting data from NASA's APOD (Daily Astronomy Photo) API, storing the data in a JSON file, analyzing the data, and exporting it to a CSV file. Additionally, there are sections focused on processing and visualizing the Iris dataset. The study provides a comprehensive account of the process, challenges encountered, and related solutions.

Problem 1: Retrieving and Storing NASA APOD Data in a JSON File

Approach:

- The dotenv library was used to securely load the API key from environment variables using os.getenv(), therefore safeguarding sensitive information.
- It was made possible to send GET calls to the NASA APOD API and get data for certain time periods using the get_apod_data() method.
- All fetched data was appended to apod_data.json, ensuring existing data wasn't overwritten.
- One-second break was set between API calls so that they would not go over the rate limits..

- Exception handling (try-except blocks) was included to manage potential issues like network errors or invalid dates.

Problems and Ways to Fix Them:

- **Rate Limiting**: Implementing time.sleep(1) between requests solved the risk of exceeding the API's rate limit.
- **JSON file management:**To ensure accurate data and speedy processing, methods for reading and writing large JSON files were provided.

Key Outputs:

- Successfully appended data entries in the apod_data.json file, showing that data retrieval was smooth.

Problem 2: Reading and Analyzing JSON Data

Approach:

- A function, read_apod_data(), was developed to read and load the content of the JSON file into a Python dictionary.
- The analyze_apod_media() method iterated over the data to quantify the total number of photos and videos while identifying the date with the most extensive explanation.

Challenges and Solutions:

- **Working with Large Files**: Enhanced memory utilisation and reading methodologies were used to process massive data files effectively.
- **Error Handling**: Extensive try-except blocks were included to deal with issues like missing or corrupted files.

Key Outputs:

- The total number of media types and the date with the most comprehensive description were recognized and displayed accurately.

Problem 3: Working with NumPy Arrays

Approach:

- Created a 2D NumPy array with 20 rows and 5 columns filled with random numbers from 10 to 100.
- Conditional verifications confirmed that the sum of each row was even and that the overall array sum was a multiple of 5.
- Elements divisible by both 3 and 5 were selected, and those exceeding the mean were substituted with the array mean.
- Statistical studies, including mean, median, standard deviation, and column variances, were conducted.

Key Outputs:

- Results included calculated statistics such as the mean, median, standard deviation, and variance for each column.

Problem 4: Analyzing the Iris Dataset

Approach:

- The Iris dataset was imported into a pandas DataFrame, then fundamental data investigation was performed.
- Errors in lines 35 and 38 have been corrected and verified.
- Two new features, Petal Ratio (petal length divided by width) and Sepal Ratio (sepal length divided by width), were added to the DataFrame.
- Pairwise correlations between numeric columns were calculated, and the highest positive and negative correlations were identified.
- Leaf Ratio was plotted against Leaf Ratio in a scatter plot with points separated by species and trend lines included for each species.
- A pair plot of the original and new features was generated to visualize feature relationships.

Key Outputs:

- Identification of the highest positive and negative correlations.
- Scatter plot and pair plot visualizations that provide insights into the data.

References

- Matplotlib Developers (2023). *Matplotlib.pyplot.savefig documentation*. Available at: https://matplotlib.org/2.0.2/api/pyplot_api.html#matplotlib.pyplot.savefig (Accessed: 27 November 2024).
- Dev.to Contributors (2024). *Community articles and tutorials*. Available at: https://dev.to/ (Accessed: 26 November 2024).
- NASA (2024). *NASA APOD API Documentation*. Available at: https://github.com/nasa/apod-api (Accessed: 27 November 2024).
- Python Software Foundation (2023). *Python official documentation*. Available at: https://docs.python.org/3/ (Accessed: 27 November 2024).
- Stack Overflow (2024). *Programming Q&A platform*. Available at: https://stackoverflow.com/ (Accessed: 27 November 2024).
- Google LLC (2024). *Google Translate*. Available at: https://translate.google.com/ (Accessed: 30 November 2024).