CHALLENGE DESCRIPTION

Challenge 1: (Folder 1)

1. **Natural Language Processing (NLP) for Cyber-Attack Investigation:**
    - **Objective:** The goal is to automate the process of identifying files relevant to a cyber-attack investigation.
    - **Approach:** Develop an NLP script to parse file contents and extract files containing specific keywords related to the investigation, such as "MPESA," "HAWALA," and "KAFIR."
    - **Output:** The script should move identified files to a directory named 'evidence' and generate a CSV report listing the files containing the specified keywords.

Challenge 2: (Folder 1)

2. **File Format Classification:**
    - **Objective:** Automate file categorization based on their formats to streamline analysis.
    - **Approach:** Design a classification model using provided file signatures dataset to categorize files in the folder into different formats like PDF, DOCX, JPEG, HTML, etc.
    - **Output:** The model should classify each file accurately into its respective format.

Challenge 3: (Folder 2)

3. **Malware Detection with Virus Total AI:**
    - **Objective:** Detect malware among the files extracted from RAM Dump using VirusTotal's machine learning API.
    - **Approach:** Utilize VirusTotal API to scan the extracted files and generate a CSV report of identified malware files.
    - **Output:** A detailed CSV report listing files identified as malware, which can also be used to train a malware detection ML model for future use.

Challenge 4: (Folder 3)

4. **Text Classification for File Reconstruction:**

- **Objective:** Develop an ML model to detect potential file manipulation by analyzing binary file headers and footers: Reconstruct files with corrupted extensions found in Folder 3 using ML algorithms.
- **Approach:** Train a model to identify anomalies in file structures indicating unauthorized modifications: Design an ML model capable of reconstructing files to their original format using provided headers and footer dataset.
- **Output:** The model should successfully reconstruct corrupted files based on the provided dataset or applicable ML algorithms: The model should flag files with potential manipulation, aiding investigators in identifying unauthorized changes.

Challenge 5: (Folder 4)

5. **Facial Recognition in Video Clips:**
    - **Objective:** Develop an ML model for facial recognition in video clips to identify specific individuals.
    - **Approach:** Utilize state-of-the-art facial recognition algorithms to parse video clips and accurately extract known faces.
    - **Output:** The model should accurately identify and extract specified individuals' faces from the video footage and indicate the match score.

Challenge 6: (Folder 5)

6. **Predictive Insights through Machine Learning:**
    - **Objective:** Predict the most frequently visited category in the suspect's browsing history.
    - **Approach:** Construct a classification model to predict the browsing habits category based on the suspect's browsing history dataset.
    - **Output:** The model should predict the most visited category, providing insights into the suspect's browsing habits.

Challenge 7: (Folder 6)

7. **Classification: Network Analysis:**
    - **Objective:** Predict the region associated with an input IP address for network analysis.
    - **Approach:** Develop a classification model using datasets containing IP address ranges and associated regions, as well as vulnerabilities and affiliated regions.

- **Output:** The model should accurately predict the region associated with a given IP address, aiding in vulnerability identification.

Challenge 8: (Folder 7)

8. **Steganography Detection:**

    - **Objective:** Develop a specialized ML model to detect hidden messages or data concealed within a digital image.
    - **Approach:** Design a model capable of detecting anomalies or patterns indicative of hidden content within the provided image file.
    - **Output:** The model should detect any hidden materials concealed within the digital image file.

REPORT FORMAT

☯ **INPUT: FILES PROVIDED**

☯ **OUTPUT: CSV REPORT**