Case Study Rubric – Classifying Snake Species in the DMV Area

Goal: The goal of this case study is for you to **preprocess a real-world image dataset**, **train a convolutional neural network (CNN)** to classify snake species, and **evaluate the model's performance**. You will experience the typical workflow of a computer vision classification project — from messy real-world data to building an initial, working model.

Tasks

You will complete the following tasks:

1. **Data Preprocessing**
    - Reorganize image files into species-labeled subfolders using the provided annotations.
    - Resize all images to a uniform size (e.g., 224x224 pixels).
    - Apply basic data augmentation (e.g., rotation, flipping, brightness adjustment).
2. **Model Training**
    - Build a simple CNN model (or fine-tune a ResNet-34, if you feel confident).
    - Use TensorFlow and Keras to implement the model.
    - Train your model using the preprocessed dataset.
3. **Model Evaluation**
    - Evaluate the model using:
        - Accuracy
        - Precision
        - Recall
    - Generate and visualize a confusion matrix.
    - Briefly explain (3–5 sentences) the main sources of model errors you observe.
4. **Recommendations**
    - Write 3–5 sentences suggesting one way you would improve the model if you had more time or resources (e.g., more data, different architectures, advanced augmentations).

Deliverable

Upload a PDF containing:

- A short summary of your preprocessing steps (around 1 paragraph)
- A description of your CNN model architecture (around 1 paragraph)
- Evaluation results (metrics + confusion matrix plot)
- Brief explanation of errors and one improvement suggestion

Your code should be in a clean, commented Jupyter Notebook (.ipynb) and uploaded to the GitHub repository alongside your report.

Submission Checklist

- ☐ Data reorganized and preprocessed
- ☐ CNN model built and trained
- ☐ Evaluation metrics reported
- ☐ Confusion matrix created and explained
- ☐ Final reflection and improvement recommendation written
- ☐ Code and PDF report uploaded to GitHub repository