DS 4002 - Case Study by Henry Duke

Due: TBD

Submission Format: Turn in a Github Repo containing the following four sections. Section 1 - Data: A folder containing a zip file with all of the snake images. Section 2 - Scripts: A folder containing all of the scripts of code you used for this project. Section 3 - Output: A screenshot of the final confusion matrix and statistics. Section 4 - Reflection: A pdf document reflecting on your process and describing your findings.

Why am I doing this?: Reproduction of results is a huge facet of data science. The concept of not reinventing the wheel every time you take upon a task is vital to efficient learning of certain material and techniques. Reproduction of this process will give you the opportunity to hone skills in image processing models as well as neural networks. Both skills are highly transferable and lucrative in a society that is quickly trying to harness the power of AI.

How will I know I've succeeded? : Follow each section of the rubric to successfully complete and format into a github repository to successfully complete this case study.

Category	**Details**
Data	Format: A folder within your github repo that contains zip files with images of snake Within this zip file, images of snakes should be sorted with the following folder organization: Snake.zip — Train — - Venomous - Nonvenomous Test — - Venomous - Nonvenomous

Scripts	Format: A github folder containing all of the scripts you used. Run the python code from the following scripts: 1. Preprocess_and_EDA a. In this script you will preprocess your image zip file data. b. Once processed, you will complete multiple EDA checkpoints to ensure that your image data is sufficient to be run within the convolutional neural network model. 2. Snake_Model a. In this script you will process the image data once again. b. Once processed, you will import the pre-trained neural network from MobileNetV2 and ensure that the top layer is built for binary classification. c. Create a confusion matrix from the model run with the test data. d. Find the following model statistics after running with the test data: Accuracy, Recall, F1-Score, Precision. 3. Any other scripts you used or needed to complete analysis (Optional)
Output	Format: A github folder containing screenshots of your results Run the python scripts and find the following: 1. Model Results a. A screenshot with the confusion matrix followed by the accuracy, recall, F1-score, and precision. 2. EDA Plots (Optional)

	a. This step is to help you to ensure that the data is properly processed and adequate to use in the model.
Reflection	Format: Upload a 1-2 paragraph pdf reflecting on your process Reflection should include: - Description of the snakes used in the data. - Analysis of the model results. - Skills that you learned. - Challenges that you encountered. - Feedback on the quality of the case study.
References	If you used any external references not provided by the case study, create a section at the end of your reflection. All references should be in IEEE format.