CS3 Rubric – Poisonous Plants

DS 4002 – Fall 2025 – Mark Haller
Due: Monday after Presentation week of project cycle 3, via Canvas

Submission format:

- **Upload link to GitHub repo to Canvas AND**

Individual Assignment

Why am I doing this? This is your opportunity to apply image recognition machine learning to the problem of recognizing various poisonous and non-poisonous plant species. As you work through the assignment you will be exposed to ways image recognition is being used to solve pressing problems and improve people's everyday lives. This will also allow you to increase your technical skills as you work through the case study.

What am I going to do? Make sure to access the GitHub repository for this case study, which is github.com/Markmh03/DS-4002-CS3. For this project you will be accessing the database of photos linked at the Google Drive link under the DATA folder in the GitHub repository. After obtaining the photos at the link (warning, the download for the folder can be large, so be prepared for that), you will be running the photos on the model in the ANALYSIS folder. After this, your Deliverables include:

- A Github repository containing all of the work you completed for this project
- Documented code used during the case study
- Link to the data
- List of references - including a few key pieces to get someone started

Tips for success:

- Be bold. This is your chance to pick something you learned and share it.
- Don't overthink it. Image recognition is a complex subject matter but the case study was made to be accessible so do not hesitate to start and complete the project.
- Make sure that you have enough storage in your computer for the photos. It is essential that your model has ample access to the photos for running them.
- Be careful reading through the instructions and GitHub repository,
- Talk to your fellow students. This is a creative assignment, and you are allowed to show ideas to people for comment.

How will I know I have Succeeded? You will meet expectations on the case study when you have followed and completed the criteria below.

Spec Category	Spec Details
Formatting	● Writing portion: Submit this as a pdf document, should be approximately a page in length ● Coding Portion: Include code under a CODE folder in GitHub, make sure to organize the analysis, preprocessing, and EDA into different folders. ● Results: Make sure to organize results into OUTPUT folder ● References: Make sure all references are cited accordingly
README.md	● Goal: This readme is to document the steps that you took to reproduce the case study ● Make sure to document steps as you take them, this reduces the risk of forgetting steps if you wait until the end to document. ● This does not have to be extremely advanced, just enough so that a reasonable person could follow along. ● Recommended to put references in here, although not required
Code	● Goal: be able to identify the basics of the code used and document well for future reproduction ● Make sure to include the following (comment accordingly) - Code used to perform analysis - Code used to perform pre-processing - Confusion matrix - Model output performance - EDA analysis (do different graph than what is in there)
Write-Up	● Goal: Communicate what you did in the case study ● Brief results section to orient people to the case study ● MAKE SURE TO READ THE STEPS ● Brief section on what you learned from this case study (no more than a paragraph) ● Make sure that it is accessible and readable to the average person, should be able to convey (briefly) the data, the model you ran, what it did, and the results and its applications.

Acknowledgements: Special thanks to Gianluca Guadagni for inspiring the set-up for this rubric.