Case Study Rubric

DS 4002 - Spring 2026
Submission Format: Link to GitHub repository
Individual Assignment

Why am I doing this?
This case study gives you the opportunity to apply DS 4002 skills to a real-world machine learning problem. By building a fruit freshness classification model using transfer learning, you will practice the full applied data science workflow: organizing data, training a model, evaluating performance, and communicating results.

Your role is to act as a junior data scientist developing an automated system to classify fruit freshness. This assignment builds your ability to justify modeling decisions, recognize uncertainty and bias, and explain your work to a broad audience.

What am I going to do?
You will complete a guided machine learning project using real images of fruit at various freshness levels. Using the provided repository and scripts, you will:

1. Organize and preprocess the dataset into training, validation, and test sets.
2. Implement transfer learning using a pretrained ResNet-50 model with a new classification head.
3. Train and fine-tune your model using best practices such as early stopping and normalization.
4. Evaluate and interpret model performance.
5. Identify uncertainty, bias, and limitations within your model and data.
6. Write a short report explaining your workflow, decisions, and conclusions.
7. Submit code that reproduces your results.

This case study simulates how a real data scientist would build, justify, and communicate an end-to-end model.

Deliverables Include:

- GitHub repository including:
    - README.md: A summary of the project, instructions for running the modeling pipeline, and references used.
    - DATA folder: Raw fruit images and any transformed or augmented datasets used during preprocessing.
    - SCRIPTS folder: All Python scripts and Jupyter notebooks used to preprocess images, create train/validation/test splits, train the ResNet-50 model, and generate evaluation metrics.
    - OUTPUT folder: All files produced from training and evaluating the model.
    - One page written summary: A PDF explaining the modeling process, evaluation, and key findings regarding fruit freshness classification.

How will I know I have succeeded?
You will meet expectations when you follow the criteria in the rubric below.

Spec Category	**Spec Details**
Formatting	● Repository – A GitHub repo (and cloud storage folder if necessary) containing all materials ○ Submit a link to the repository. ○ Ensure everything is either contained

	in the repo or linked appropriately if too large. ○ Contents ■ **README.md** ■ **DATA Folder** ■ **SCRIPTS Folder** ■ **OUTPUT Folder** ■ **Written Summary (PDF)** ○ Use PDF format where appropriate ○ Use proper documentation and clear code comments
README.md	● Include a ## Project Overview explaining the purpose of the case study ● Provide a short summary of all contents included in the final deliverable. ● Include a Map of Documentation describing where each major folder and file is located in the repository. ● Include references in proper citation format.
DATA Folder	● All raw fruit images used in model development. ● Any images that were cleaned, resized, or augmented must be stored in a **subfolder**. ● The original dataset (downloaded from its source) should be kept in a clearly labeled subfolder (e.g., `raw/`). ● This folder should represent the complete data pipeline—from raw images to preprocessed inputs ready for modeling.
SCRIPTS Folder	● The folder will contain all code pertaining to analysis ● All EDA scripts that helped to visualize and explore the data should be included ● Create the model using Python and execute all tasks involved in the case study
OUTPUT Folder	● This folder should contain all files and graphs that are generated from scripts ● Label each file according to each step that generated it
Written Summary	● A one page PDF summarizing: ○ The overall modeling process ○ How the data was prepared ○ How the ResNet-50 model was adapted and trained ○ Key evaluation metrics and general performance ○ A concise interpretation of model strengths, limitations, and next steps