

# Fruit Freshness Classification Case Study Rubric

DS 4002 – Fall 2024

Colette D'Costa

**Submission Format:** Presentation and GitHub Repository

## Purpose

This project focuses on applying machine learning techniques to a real-world sustainability problem: food waste reduction in dining halls. The goal is to classify fruits as "fresh" or "spoiled" using image data and to analyze the model's effectiveness. Through this assignment, you will strengthen your skills in exploratory data analysis, computer vision, and binary classification modeling. You will also gain experience in evaluating the trade-offs and limitations of model design choices.

The deliverable will contribute to a better understanding of how machine learning can be used to inform operational decisions, offering a framework that can be extended to other areas of food sustainability.

## Task

1. **Explore the Dataset:**
  - Analyze the provided fruit image dataset, containing labeled images of fresh and spoiled fruits.
  - Understand patterns in the data through color histograms and exploratory visualizations.
2. **Preprocess the Images:**
  - Resize and normalize the images to a consistent format for modeling (224x224 pixels).
3. **Build and Train the Model:**
  - Use transfer learning with ResNet50 to classify images as "fresh" or "spoiled."
4. **Evaluate the Model:**
  - Measure model accuracy using training and validation data.
  - Create a confusion matrix to analyze the model's classification performance.
5. **Document Findings:**
  - Compile results, visualizations, and conclusions into a well-documented GitHub repository.

## Tips for Success

- **Visualizations:** Create clear, informative visualizations to support your analysis and demonstrate trends in the data.
- **Documentation:** Ensure all scripts and processes are well-commented and include clear instructions for reproducibility.
- **Focus on Impact:** Highlight practical implications and limitations of the model, providing actionable recommendations.
- **Be Concise:** Present results in a clear and professional manner, avoiding unnecessary complexity.

## How Will I Know I Succeeded?

By the end of this project, you should deliver:

- A well-documented GitHub repository with clean data, scripts, and visualizations.
- Outputs that include insights into the model's performance and its applicability to food waste reduction.

Spec Category	Details
Repository and Submission	<p>Provide a GitHub repository with:</p> <ul style="list-style-type: none"><li>• Preprocessed and original datasets.</li><li>• Scripts for data exploration, preprocessing, and modeling.</li><li>• Outputs, including exploratory visualizations, confusion matrices, and accuracy plots.</li><li>• A README file with clear instructions.</li></ul>
Exploratory Data Analysis	<ul style="list-style-type: none"><li>• Execute scripts to generate initial visualizations (e.g., histograms, color intensity distributions).</li><li>• Highlight trends or anomalies in the dataset that inform model design.</li></ul>
Preprocessing	<ul style="list-style-type: none"><li>• Transform images into consistent dimensions (224x224).</li><li>• Apply normalization techniques to standardize image data.</li></ul>
Modeling	<ul style="list-style-type: none"><li>• Implement a ResNet50 transfer learning model.</li><li>• Clearly document the choice of model parameters, including binary classification focus.</li></ul>
Evaluation	<ul style="list-style-type: none"><li>• Produce visualizations such as training/validation accuracy and loss plots.</li><li>• Generate a confusion matrix to assess true/false positives and negatives.</li></ul>