Case Study: AI Fruit Detective – Case Study Rubric

DS 4002
Due: TBD
Submission format: PDF Report & link to your GitHub repository (e.g., a fork of the provided starter repo) containing your completed notebooks, submitted via [Canvas/LMS].

Individual Assignment

General Description:
This case study requires you to engage with a practical image classification problem using deep learning. You will work with a provided dataset of fruit images, explore its characteristics, train a Convolutional Neural Network (CNN) using transfer learning, evaluate its performance, and document your findings in a technical report and completed Jupyter Notebooks.

Course Learning Objective:
Apply fundamental machine learning techniques to solve a classification problem, interpret the results, and articulate the process.

What am I going to do?
You will be provided with a GitHub repository containing:

- A dataset of fruit images (or instructions to download it).
- Starter Jupyter Notebooks
- Reference materials to support your understanding of key concepts.

Your tasks will involve:

1. **Setup and Familiarization:** Clone/fork the repository. Review the case study prompt (the Hook document), this rubric, and the provided reference materials. Set up your environment (Google Colab is recommended for GPU access).
2. **Data Exploration:** Execute and understand the 1_Fruit_EDA.ipynb notebook. Document any key observations about the dataset.
3. **Model Training and Tuning:** You will execute your own code to load data, preprocess images, build a CNN using ResNet50 as a base, train it, and potentially perform basic fine-tuning.
4. **Model Evaluation:** Critically assess the performance of your trained model using the metrics generated in the notebook.
5. **Reporting:** Compile a technical report summarizing your methodology, results, and conclusions. Ensure your notebooks are fully executed and outputs are visible.

Deliverables for You:

1. **GitHub Repository:** A link to your GitHub repository containing:
   - A completed 1_Fruit_EDA.ipynb and 2_Fruit_Classifier_CNN.ipynb notebooks with all cells run and outputs visible.

- Your final PDF report.
- (Optional) Any other supporting files you deem necessary.

2. **Technical Report (PDF, approx. 2-4 pages):**
   - **Introduction:** Briefly state the problem (fruit classification) and the objective of the case study.
   - **Data Exploration:** Summarize key characteristics of the fruit image dataset (e.g., number of classes, images per class, any imbalances, image size/brightness variations). Include 1-2 illustrative visualizations from your EDA notebook, with captions.
   - **Methodology:**
     - Briefly describe the image preprocessing steps.
     - Describe the CNN model architecture (mentioning ResNet50 and transfer learning) and the training process (e.g., loss function, optimizer, epochs, fine-tuning if performed).
   - **Results and Discussion:**
     - Present the overall performance of your final model (e.g., accuracy on the test set).
     - Include and *interpret* the confusion matrix. Discuss which fruits the model classifies well and which it struggles with, offering potential reasons.
     - Show training/validation loss/accuracy curves and briefly discuss what they indicate (e.g., overfitting, convergence).
   - **Conclusion and Reflection:** Summarize your main findings. Briefly reflect on what you learned, any challenges encountered, and one or two potential future steps to improve the model.

Tips for Success:

- Allocate sufficient time for each part, especially model training (which can take a while even with a GPU).
- Read the comments in the notebooks thoroughly.
- Don't just run the code; strive to understand the purpose of each step.
- Use the provided reference materials to clarify concepts like CNNs, transfer learning, and evaluation metrics.
- Ensure your report is well-structured, and your visualizations are clear and properly labeled.

How will I know I have Succeeded?
You will meet expectations on this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting & Submission	GitHub repository link submitted. Repository contains clearly named and fully executed Jupyter notebooks and the PDF report.

	PDF report adheres to approximate page length and is professionally formatted.
Report: Introduction	Clearly and concisely introduces the fruit classification problem and the case study's objective.
Report: Data Exploration	Effectively summarizes key characteristics and insights from the dataset exploration. Includes relevant visualizations from the EDA notebook that are well-labeled and support the text.
Report: Methodology	Accurately describes the image preprocessing steps. Clearly outlines the CNN model architecture (including transfer learning aspects) and the main parameters of the training process.
Report: Results & Discussion	Presents key performance metrics (overall accuracy, confusion matrix mandatory). Provides a clear and insightful interpretation of the results, particularly the confusion matrix and training curves. Discusses model strengths/weaknesses with plausible reasoning.
Report: Conclusion & Reflection	Summarizes the main outcomes of the case study. Offers thoughtful reflection on the learning experience, challenges (if any), and suggests at least one specific, relevant idea for future model improvement.
Jupyter Notebooks	1_Fruit_EDA.ipynb: All cells are executed, outputs are visible, and it demonstrates a clear understanding of the dataset. 2_Fruit_Classifier_CNN.ipynb: All cells are executed, showing the complete model training, fine-tuning (if applicable), and evaluation process. Code is understandable (leveraging provided comments).
Overall Quality	Work demonstrates a good understanding of the image classification task and the application of deep learning techniques. Report is well-written, coherent, and free of major errors.