Case Study Rubric

Case Study Title: *Tomato Trouble: Using Machine Learning to Combat Crop Disease*
Due: TBD
Submission Format: Upload PDF and link to GitHub repo on Canvas

Why am I doing this? This case study is an opportunity to bring together your skills in machine learning, model evaluation, computer vision, and data storytelling. More than just classification, your work could help lay the groundwork for accessible tech that supports sustainable agriculture.

What am I going to do?

At this point, you have likely accumulated a variety of technical and conceptual skills in data science. You will now have the opportunity to combine and express these skills in an independently-driven case study. You will ultimately provide a deliverable that covers all requirements, including significant results and conclusions. The deliverable will include: You will develop a machine learning pipeline to classify images of tomato leaf diseases. The PlantVillage dataset contains labeled images for several common diseases, and your job is to design a prototype tool capable of identifying them.

Your final submission must include:

- **Written Portion (PDF)** – outlining the problem, your approach, results, reflection, and references.

- **GitHub Repository** – containing all code, data, models, and visualizations necessary to replicate your work.

How will I know I have succeeded?

You will meet expectations for this case study when you complete the following components according to the rubric below.

Category	Spec Details

Formatting	● Submit the written portion as a PDF file. ● Submit code created for all portions in a GitHub repository. ○ Include any additional data that was used in GitHub if necessary. ○ The repository should be titled “CS-[insert first & last name]”. ● References should be included on a separate page at the end of the Written Portion PDF. ○ Use IEEE citation style.
Written Portion	● In a small paragraph, summarize the problem presented in the study as well as its importance. ● In a small paragraph, discuss your plan to meet the demands of the deliverable. ○ Include a simple graphic that outlines your analysis plan. Finally, discuss the results of your study in complete sentences as well as the significance of these results in the greater context.
Code	● Include preprocessing for the PlantVillage tomato image dataset (resizing, normalization, augmentation). ● Implement and train four models: ResNet50, VGG19, InceptionV3 (transfer learning), and Support Vector Classifier (SVC). ● Evaluate each model using: Accuracy, Precision, Recall, F1-score, and Confusion Matrix. ○ Include visualizations: training/validation plots and sample image predictions. Write clean, organized, and well-commented code.
References	- include all external references in IEEE citation style. - Place references on a separate page at the end of the PDF. Cite any tools, datasets, libraries, or external content.