MNIST Classification Case Study Rubric

DS 4002 – Spring 2025 – Emily McMahon
Due: TBD
Submission format: Upload link to GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?
This case study guides you through a fundamental aspect of data science: evaluating various machine learning models using the MNIST handwritten digit dataset. As you work through the material, you'll gain insight into the critical role model selection plays in achieving accurate classification, particularly in real-world applications and industries such as banking, postal services, and government data entry where precision is critical.

What am I going to do?
The GitHub Repository for this case is found at: https://github.com/feu6sx/DS4002-CS.git. There, you will find an example of how to load in the MNIST handwritten dataset from Keras and prepare it for use in a machine learning model. Next, you'll have the opportunity to apply the prepared dataset to a variety of machine learning models and evaluate their performance through comparison.

Your final deliverables should include:
- For each model:
  - A confusion heatmap visualizing the TP, TN, FP, and FN calculated.
  - An accuracy score representing how well the model performed.
- An organized source code showing all analytical work.
- A one-page written summary of your findings in PDF format.
- A GitHub Repository containing all materials needed to recreate your work.

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

Formatting	• A new GitHub Repository containing the following: o README.md o LICENSE.md o Source Code File o Written Summary o REFERENCES.md
README.md	• Purpose ○ This document is designed to help anyone who visits your repository quickly understand its structure and

	purpose. It acts as a guide to help users navigate your project with ease. ● How to Organize Your Content ○ Use markdown headers to clearly separate each section of the document. ● Repository Overview ○ Provide a brief description of the contents of your repository. This section should give users a high-level understanding of what to expect in the project. ● Software and Platform Details ○ List the software tools used in the project. ○ Mention any additional packages or libraries that need to be installed. ○ Specify the operating system(s) the project was developed and tested on (e.g., Windows, macOS, or Linux). ● Project Structure ○ Include a directory tree or outline showing how your project folders and files are organized. Be sure to list all key files and indicate which folder each one is located in. ● Reproducing Results ○ Provide clear, step-by-step instructions that someone else can follow to replicate your results.
LICENSE.md	● Purpose ○ This file informs visitors about the terms and conditions for using, modifying, and citing the contents of your repository. ● When creating your repository, choose a suitable open-source license from GitHub's available options. ● In most cases, the MIT License is a good default choice.
Source Code File	● A well-documented Jupyter Notebook containing the code used to perform your images and statistical analyses. Your source code should include the deliverables for this project: ○ Confusion Matrix for each model ○ Accuracy score
REFERENCES.md	● All references must be compiled at the end of the document and formatted according to the IEEE citation style.

Acknowledgements: Thank you to Professor Rasero for the basic outline of this rubric.