FRT Case Study Rubric

DS 4002 – Fall 2025 – Atharv Gupta
Due: Monday after Presentation week of project cycle 3, via Canvas

Submission format:

- **Upload link to GitHub repo to Canvas AND**
- **Hard copy turned into Data Science Building (details on Canvas)**

Individual Assignment

Preparatory Assignments – Everything in the course, but especially CS1, CS2.

Why am I doing this? You are doing this case study to learn how a full machine learning pipeline works when you are working with image data. You will see how data choices shape the model, how a CNN learns visual patterns, and how all the steps connect to build a working system. The goal is to help you understand the bigger picture of how image based models are created and evaluated, and to give you confidence in using these ideas in future projects.

- Course Learning Objective: Be comfortable working with image data and replicating a machine learning project

What am I going to do? For this case study you will recreate a full image based machine learning pipeline. You will start by going to the GitHub repository and selecting a small set of images of yourself or your friends to use as your personal dataset. Then you will follow the instructions to prepare the celebrity dataset, explore it, clean it, and place it in the right structure so that the model can read it. After that you will train a CNN to learn the visual patterns in your images and compare them to the celebrity dataset. You will evaluate how well your model works and reflect on what influenced its performance. At the end you will submit a few deliverables that show your final model, your process, and what you learned.

Deliverables include:

- Well commented source code
- Outputs of the CNN model (including confusion matrix)
- Github repository – containing code and necessary data
- List of references – any outside sources used to strengthen the project

Tips for success:

- Enjoy the process, it's fun to work with image data, especially your own
- Use python code to make sure dataset (of your pictures) is resized

- Be familiar with Python as this project was done in Python
- Feel free to collaborate with other students to get their ideas and make any improvements to this CNN project

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	● One GitHub repository (submitted via link on Canvas) It should contain the following ○ README.md ○ SCRIPTS Folder ○ Data Folder ○ Output.md ○ References.md
README.md	● Goal: Clearly explain the GitHub repository Brief summary of what the GitHub contains. It tells a reader what all work you've produced for this case study.
SCRIPTS Folder	● Goal: Contains all the source code for this case study ● Have code files (either python or jupyter) for all aspects of case study - Data Cleaning - EDA Plots - CNN Model - Evaluation Metrics
Data Folder	● Goal: Include data required for this case study ● Have all the data files for this case study (or alternatively directions to acquire that data) - Personal dataset - Celebrity dataset
Output.md	• Goal: Show the output of this case study • Show the confusion matrix for the test set of the CNN Model • Print the test set (of the personal images) with text mentioning what the CNN model predicted and the true value

References.md	• Please include a list of references in IEEE citation style that were not in the already given reference material

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020).