

# Case Study Rubric

**DS 4002 – Fall 2024**

**Submission format: A GitHub Repository**

## **Individual Assignment**

**General Description:** Submit a link to your case study repository

**Why am I doing this?** This is your opportunity to showcase your technical and conceptual skills to create a basic model regarding an area of rapid technological development. As you work through this assignment, you will gather more knowledge about image data analysis and how complex models can be created to facilitate the interpretation of pictures and videos. You will also be exposed to the ways in which these models are currently being utilized, and how they can easily produce incorrect conclusions that can have broader implications based upon their intended purpose.

**What am I going to do?** You will build a facial recognition model using Keras that will analyze celebrity images and attempt to successfully recognize the individual in a photo. The case study has provided a set of 2500 images of 25 different celebrities (100 images each) which you may use, or you may choose your own celebrities to analyze and download the images yourself. You will use the Keras package in Python to carry out this analysis and compile it into a GitHub repository that includes the results of the model training, as well as a written summary of your research results and any key takeaways.

**Deliverables Include:**

- GitHub repository containing:
  - README.md file that contains a brief description of the project and any references utilized (including both those provided in the case study and gathered by the student)
  - DATA Folder containing the raw images (either provided in the case study or generated by the student) and the additional photos generated by image augmentation/preprocessing
  - SCRIPTS Folder containing all the Python/R code for the analysis and model creation
  - OUTPUT folder containing the results of the training process and any EDA created
  - A one-page written summary of results and key takeaways in PDF format

**Tips for success:**

- Start by learning Python basics or refresh your knowledge of the language
- Research Keras to understand its basic functions and how it will be implemented in the project
  - A link to Keras documentation and Keras Facial Recognition Tutorial have been included with the case study materials, but be sure to look at other resources if necessary
- If you choose to analyze other celebrities and download your own images, be sure to pick 25 different celebrities and gather 100 images of each celebrity to start with a sufficient data set

- Also be sure to organize them in the same way as the provided dataset so that they can easily be iterated through

**How will I know I have Succeeded?** You will meet expectations when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"> <li>● Repository – A GitHub repo (and cloud storage folder if necessary) <ul style="list-style-type: none"> <li>○ Submit a link to the repo containing all necessary materials</li> <li>○ Contents <ul style="list-style-type: none"> <li>▪ README.md</li> <li>▪ DATA Folder</li> <li>▪ SCRIPTS Folder</li> <li>▪ OUTPUT Folder</li> <li>▪ Written Portion</li> </ul> </li> <li>○ Use pdf format when possible and appropriate format for code</li> </ul> </li> </ul>
README.md	<ul style="list-style-type: none"> <li>● Make an H2 (##) section explaining the contents of the repository</li> <li>● Provide a short summary of what is included in the final deliverable</li> <li>● A map of your documentation to show the organization of folders and subfolders and where each file is located</li> <li>● References in IEEE format should be included at the end. Cite all sources provided by the case study and any that may have been acquired by the student.</li> </ul>
DATA Folder	<ul style="list-style-type: none"> <li>● All the images used in the model should be located in this folder</li> <li>● The raw images acquired by the student or provided in the case study should be stored in a subfolder, with further subfolders to organize the images for each celebrity</li> <li>● The images generated from preprocessing/augmentation should be stored in a separate subfolder with the same structure as the raw dataset</li> </ul>
SCRIPTS Folder	<ul style="list-style-type: none"> <li>● This folder will contain all the code that was used in your analysis</li> <li>● Include EDA scripts of image properties/metrics to inspect the overall quality of the dataset</li> <li>● Create the model using Python/Keras and executing the training/testing process using the provided scripts</li> <li>● Label each script with the order in which it should be executed</li> <li>● Include comments and headers to explain each step of your code and any details that readers should make note of</li> </ul>
OUTPUT Folder	<ul style="list-style-type: none"> <li>● This folder will contain all the files and graphs that are generated from the scripts</li> <li>● Clearly label each file to note the step during which it was generated</li> </ul>
Written Summary	<ul style="list-style-type: none"> <li>● A one-page PDF document</li> <li>● Explain how the EDA was conducted and the general process of creating the model and subsequent training/validation</li> </ul>

	<ul style="list-style-type: none"> <li>● Provide a brief interpretation of the results and the potential broader implications</li> <li>● Reflect on what you learned throughout the process, what you would have done differently, and what further steps could be taken to extend the case study</li> </ul>
--	--

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