Case Study Create

DS 4002 – Spring 2024

Submission format:

- **Upload link to github repo**

Individual Assignment

General Description: Submit a link to your case study GitHub Repository

Why am I doing this? This is your opportunity to showcase the technical and conceptual data science skills you have learned this semester to create a basic model for object detection and image recognition. As you work through this assignment, you will learn more about image recognition and how different models can be utilized to detect different and multiple objects within an image. You will also learn how these models are currently being employed in our everyday lives and how we can create detection models to help make this process more efficient and useful for society.

What am I going to do? In this case study, you will create an image recognition and object detection model using You Only Look Once (YOLO) that will analyze and identify different fruits within the same image. The case study has provided a dataset with 21,122 images of fruits taken in different combinations, from different angles, and in different lighting conditions which includes a test and train data set. You will use the YOLO package in Python to complete your analysis and modeling and compile your resources and results into a GitHub repository that also includes a written summary of your results and what you learned.

Deliverables include:

- GitHub repository containing:
  - README.md file that contains a brief description of the project and any references used
  - DATA folder containing the image dataset
  - SCRIPTS folder containing all of the Python code for the analysis and model creation
  - OUTPUT folder with the results of the training process and any EDA created
  - A one-page summary of your results and what you learned as a PDF

Tips for success:

- Be bold. This is your chance to pick something you learned and share it.
- Refresh yourself on Python language basics
- Research YOLO and understand how it works and how you can implement it to create your model
  - A link to YOLO documentation is included in the resources

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

Formatting	● One Github Repository ● Contents: ○ A README.md file (which auto displays) ○ A LICENSE.md file (use MIT as default) ○ A SCRIPTS folder ○ A DATA folder ○ AN OUTPUT folder
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ● Use markdown headers to divide content. ● Make an H2 (##) section explaining the contents of the repository ● Section 1: Software and platform section ○ The type(s) of software you used for the project. ○ The names of any add-on packages that need to be installed with the software. ○ The platform (e.g., Windows, Mac, or Linux) you used. ● Section 2: A Map of your documentation. In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder.
LICENSE.md	● Goal: This file explains to a visitor the terms under which they may use and cite your repository. ● Select an appropriate license from the GitHub options list on repository creation. ● Usually, the MIT license is appropriate.
SCRIPTS folder	● Goal: This folder contains all the source code for your project. ● Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results. ● All script files should include header comments at the beginning of a script to provide information that anyone working with or executing the script should be aware of. **Throughout all your scripts, you should include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is**.
DATA folder	● Goal: This folder contains all of the data for this project. ● All of the images used in the model should be located here
OUTPUT folder	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc.

	• Any information like tables, figures shown in your presentation should be here. • **Use informative names for your files**.
Written Summary	• One-page PDF document • Explain how you conducted your EDA, model, and testing • Provide an interpretation of the results as well as a reflection of what you learned and how you would alter this approach to extend or improve the case study
References	• All references should be listed at the end of the document • Use IEEE Documentation style (link)

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