CS3 Rubric – Case Study

DS 4002 – Spring 2024 - Vishal Kamalakrishnan

Individual Assignment

Why am I doing this? This project offers a hands-on opportunity to test your technical and conceptual data science skills. This case study will simulate a real world problem similar to one you will face as a professional Data Scientist. You will synthesize a repository full of key insights to help intelligent transportation systems by the NYSDOT.

What am I going to do? Using the materials found in the GitHub repo, you will download and unzip the data, perform exploratory analysis, and use RESNET to accurately classify emergency vehicles from non-emergency vehicles with an accuracy of 95% or above.

Your final deliverables will be:

- Written report presenting your methodology and findings
- Github repository detailing all scripts, data, and output files

Tips for success:

- Be thorough. You are trying to generate actual, usable insights that will further research on an important and ongoing field.
- Don't overthink it. A clear presentation of fundamentals is more valuable than an unclear presentation of cutting edge techniques.
- Remember that it is always better to ask your peers or instructors for help rather than being stuck on a roadblock for an extended period of time.

How will I know I have Succeeded? You will meet expectations on CS3 Create Case Study when you follow the criteria in the rubric below.

Formatting	Github Repository (submitted via link on Canvas) containing all the materials used in the project. This should include - README.md file - LICENSE.md file (MIT as a default) - SCRIPTS folder (containing all your code) - DATA folder (containing all the data) - OUTPUT folder (containing any generated files/plots) Written Report detailing analysis, methodology, and resutls
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. ● Section 3: Instructions for reproducing your results. In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise, but detailed and precise enough, to make it possible for an interested user to reproduce your results without much difficulty. Keep in mind that the README should be thorough enough so that another student can reproduce the project ● Section 4: References All references should be listed at the end of the document. Use IEEE Documentation style (link)
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	● 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	● Goal: This folder contains all of the data for this project.

	● You should AT LEAST the data include the initial data, and the final data analyzed. If needed, the code in the SCRIPTS folder should be able to get you from the initial piece of data to the final one. N.B. If the initial and final data are the same, then just include that dataset. ● If your data fits in github, place all of it here. ● If your data does not fit in GitHub use a single file explaining the process to obtain the dataset.
OUTPUT	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc. ● The content here can be in progress when MI3 is complete. It should be finished during MI4 though. ● Importantly, any information like tables, figures shown in your presentation should be here. ● **Use informative names for your files**.
REPORT	● Goal: Present findings in a professional and informative manner. Your report should include your analysis, methodology, and results ● Your report should include: o Exploratory plots/findings o Methodology on data analysis and insights o Resnet Model description and detail o Resnet Model results and accuracy o Accuracy plots/metrics ● Create a comprehensive report that details the entire process of data exploration, model creation, analysis, results, and improvements that future studies can make.

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