Dog Imaging Case Study Rubric

DS4002 - Fall 2024 – Caroline Hagood
Due: December 9
Submission format: Link to Github Repository uploaded to Canvas

Individual Assignment

Why am I doing this? This case study allows you to use your data science and machine learning knowledge by using an image analysis model to accurately identify different dog breeds. In completing this assignment, you will build upon your skills to process and use image data in a CNN model.

What am I going to do? The link for the Github repository is https://github.com/carolinehagood/CS3-DS4002. You will download the Stanford Dogs Data from the link in the Github. From the 200 breed folders of images, select five breed folders to upload to import into your notebook and process into rbg image data. You will then create a combined dataset which you will build your CNN model on. Your model will include test and train sets and will use accuracy as a main criterion. Finally, you will run your final model and evaluate the accuracy level, aiming for at least 80%, and produce a confusion matrix of the results.

Final deliverables include:

- The image data for the five chosen dog breeds
- Well documented source code
- A confusion matrix and final accuracy values
- A Github repository containing all material used and produced

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	• One Github repository (submitted via link on Canvas) • The top-level page of the repository should include -A README.md file -A LICENSE.md file -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. • This should give a brief introduction to what you produced for the case study but does not need to be very detailed.
Scripts folder	• Goal: Well documented Colab Python notebook file that contains code used to create your CNN model and results. • The code must show how you produced: -The CNN model -The accuracy results -The confusion matrix
Data Folder	• Goal: This contains the data used for model analysis, NOT the original dataset • A single or multiple csv file of your cleaned and processed image data
Output Folder	• Goal: This folder contains the final results of the model. • Must include: -Confusion matrix -Test and train accuracy values
REFERENCES.md	• Markdown file citing any resources used to build model