Book Cover Case Study Rubric

DS 4002 – Fall 2024 - Victoria DaRosa
Due: Dec 9
Submission format: Upload GitHub repository link (submitted to Canvas)

Individual Assignment

Why am I doing this? This case study is meant to introduce you to the basics of image analysis. It will give you the chance to practice using new technologies before going out into the real world. Throughout the process you will be exposed to innovative tools and see how image analysis can be used for real life applications. It will also provide you with the opportunity to document your process and analysis.

What am I going to do? The Github repository for this case study can be found here. You will obtain Image URLs of book covers, along with the prominent genre the book falls under. In the Github you will find example code on how to scrape image URLs, along with the book genre, off of multiple Goodreads webpages. In addition to the images found in the output of this code, you will find supplementary images of book covers to add to each genre to enhance the dataset. After collecting all of your images, you will split the data into a training and testing set to create a neural network model to classify books covers into their genre. The script should record the accuracy and loss values for both training and validation sets during each epoch and output the final validation accuracy and loss. Submit your work in a Github repository.

Final Deliverables Should Include:

- A dataset including the image URLs and their genre
- Output graphs to display the accuracy and loss of your model
- A data dictionary
- Well documented source code
- A Github repository storing all materials used and produced

Tips for success:

- Research the different components of creating a convolutional neural network to understand what goes into creating a quality model.
    - What may have to experiment with your model configuration to create the most accurate model.
- Ensure you have a quality dataset. Don't use old images whose styles may be out of date or add the same cover under multiple genres.
    - Collecting modern book covers and only classifying them under one genre will set your model up for success.
- Don't get stressed! Learning to create CNN models for image classification can get overwhelming at first.
    - Utilize resources online or take a step back to help you understand the problem at hand.

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) ● Title the Github repository "CS3_GenreClassification_<YourName> ● The repository should contain the following folders: ○ A README.md file (which auto displays) ○ A SCRIPTS folder ○ A DATA folder ○ A REFERENCES 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. ● Provide a brief summary of the purpose of the case study, the code you wrote, and your findings from the CNN model.
SCRIPTS folder	● Goal: This folder contains all the source code for your project. ○ Make sure to clearly state what the script was used for in the script name and header. ● Add detailed comments to each script so that someone who never worked on the script before would understand what is happening.
DATA folder	● Goal: This folder contains all of the data related to this project. ● Include the dataset that was used to train your model. ● Include any outputted graphs or data created from the model. ○ Clearly label if the data within the folder is part of the initial dataset or if the data is the result of code. ● The final data dictionary will also be added in this folder.
REFERENCES folder	● Goal: Cite additional resources related to this project. ● Any additional resources that were used to help you create your model should be included here. ○ Cite resources using IEEE Documentation style.

Acknowledgements: Special thanks to Professor Alonzi for the structure of this rubric!