CS-Canada Rubric

DS 5000 – Spring 2023 – Michael Vaden
Due: May 20, 5pm
Submission format: Link to github (or similar) repository (collab assignments)

Individual Assignment

General Description: Submit to collab assignments a link to your case study repository

Preparatory Assignments – Class discussion on time series forecasting. Study on Measuring Errors in Fourier Analysis. Decision Tree and Random Forest Codeathon. Deep-Learning Codeathon. Scikit Learn Tutorial.

Why am I doing this? We read and produce solutions to case studies to practice thinking like a data scientist. In the case of Canada's serious global warming problem, we approach a problem in the real world and apply in-depth analysis to better our own understanding of data analysis techniques, as well as real applications of data science. Additionally, in the case of global warming, there are tons of resources available that explain the effects of the unfortunate phenomenon and provide extensive analysis. By approaching the problem, you can grow as a student by juxtaposing your project with real-world analysis.

- Course Learning Objective: logic and problem solving
- Course Learning Objective: applied thinking
- Course Learning Objective: comparing results and analysis
- Course Learning Objective: presentation of results

What am I going to do? You will first read the one-page prompt for the Canada temperature rising case study. You will be given an objective with context and a deliverable. Take some time to reflect on the assignment. Then, consider previous assignments that you have completed that may be helpful for this prompt, and prepare to research as necessary. Make the structure of your deliverable by creating a repository with files listed down below in the table. As you create these files, add them to your existing structured repository. Develop your models and think about what error metric you want to use to compare them. Research a time-series specific model to implement and consider how it performs compared to traditional models you have prior experience with. After analyzing the results of your initial models on precipitation and temperature, tune and prepare two of these models as recommendations for Toronto to better understand its global warming issue. Produce results from these models for the years 2010-2019 to show their effectiveness, and BONUS: attempt to predict temperatures and precipitation for 2020-2029. Make sure to include your error metrics, as your chosen metric needs to answer the question of "How effective are your recommended models at predicting temperature and precipitation in Toronto".

Tips for success:

- Consider previous models you have implemented. Which are most appropriate for time-series data?
- Do some research. There are many models that are used for time series data. Don't just pick the first one you come across.

- Think about what your results mean in a real-world context. Global warming is an important issue- if you were the government, what would you want a data scientist to tell you?

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

Spec Category	Spec Details

Formatting	• Repository – A new github repository
	  • Create a new github repo for this assignment containing
	  • README.md
	  • LICENSE
	  • SRC folder
	  • PLOTS folder
	  • RELEVANT DOCS folder

README.md	• Goal: This file is what will be assessed for the assignment. • Structure this file in such a way to be easily readable by an individual who has read the prompt • Include your two (or more) model recommendations • Include your choice of error metric for your model results • Include references • markdown format
SRC	• Goal: Show off your code that produces your two recommended models • Fine tune your models • Calculate your chosen error metric for each model • Clearly explain why you chose the error metric that you did • Make sure that your code is neat, well-named, and organized. It should be replicable • Write a short conclusion explaining your results.
PLOTS	• Goal: Show plots that were produced in your code by your various models • Plots should be clearly labeled and legible • Include a plot for each of your models
LICENSE	• Goal: Explain to readers the terms under which they may use and share your work. • The MIT license is the default recommendation
RELEVANT DOCS	• Goal: Include any resources or documents that you came across during your research • Aim to provide context on the models you created or the prompt as a whole • Cite sources as necessary

Acknowledgements: Special thanks to Professor Alonzi for previous examples of similar rubrics.