LeTicia Cancel

DATA622 Homework #1

October 2, 2022

Deliverable

1. Essay (minimum 500 word document) Write a short essay explaining your selection of algorithms and how they relate to the data and what you are trying to do
2. Exploratory Analysis using R or Python (submit code + errors + analysis as notebook or copy/paste to document)

Explore how to analyze and predict an outcome based on the data available. This will be an exploratory exercise, so feel free to show errors and warnings that raise during the analysis. Test the code with both datasets selected and compare the results.

Answer questions such as:

1. Are the columns of your data correlated?
2. Are there labels in your data? Did that impact your choice of algorithm?
3. What are the pros and cons of each algorithm you selected?
4. How your choice of algorithm relates to the datasets (was your choice of algorithm impacted by the datasets you chose)?
5. Which result will you trust if you need to make a business decision?
6. Do you think an analysis could be prone to errors when using too much data, or when using the least amount possible?
7. How does the analysis between data sets compare?