Coursework

One of the main objectives of this module is to help you gain hands-on experience in communicating insightful and impactful findings to stakeholders. In this coursework, you will use the tools and techniques you learned throughout this module to train a few machine learning models on a dataset that you feel passionate about, select the techniques that best suits your needs, and communicate insights you found from your modeling exercise.

After going through some guided steps, you will have insights that either explain or predict your outcome variable. As a main deliverable, you will submit a report that helps you focus on highlighting your analytical skills and thought process.

You are expected to leverage a wide variety of tools, but these report should focus on present findings, insights and next steps. You should include the visuals from your code output, but this report is intended as a summary of your findings, not as a code review. Optionally, you can submit your code as a python notebook or as a print out in the appendix of your document.

The grading will center around 5 main points:

1. Does the report include a section describing the data?
2. Does the report include a paragraph detailing the main objective(s) of this analysis?
3. Does the report include a section with variations of machine learning models and specifies which one is the model that best suits the main objective(s) of this analysis?
4. Does the report include a clear and well presented section with key findings related to the main objective(s) of the analysis?
5. Does the report highlight possible flaws in the model and a plan of action to revisit this analysis with additional data or different predictive modeling techniques?

Before you begin, you will need to choose a data set that you feel passionate about. You can brainstorm with your peers about great public data sets using the forum in this module.

Please also make sure that you can print your report into a pdf file.

As a suggested first step, spend some time finding a data set that you are really passionate about. This can be a data set similar to the data you have available at work or data you have always wanted to analyse. For some people this will be sports data sets, while some other folks prefer to focus on data from a datathon or data for good.

Required

Once you have selected a data set, you will produce the deliverables listed below. Treat this exercise as an opportunity to produce analysis that are ready to highlight your analytical skills for a senior audience, for example, the Chief Data Officer, or the Head of Analytics at your company.

Sections required in your report:

A. Main objective of the analysis that specifies whether your model will be focused on prediction or interpretation.
B. Brief description of the data set you chose and a summary of its attributes.
C. Brief summary of data exploration and actions taken for data cleaning and feature engineering.
D. Summary of training two machine learning models.
E. A paragraph explaining which of your models you recommend as a final model that best fits your needs in terms of accuracy and explainability.
F. Summary Key Findings and Insights, which walks your reader through the main drivers of your model and insights from your data derived from your models.
G. Suggestions for next steps in analyzing this data, which may include suggesting revisiting this model adding specific data features to achieve a better explanation, a better prediction, etc.

FAQs:

Q1: Do I have to come up with my own data set?

Ans: You are highly encouraged to find a data set you feel really passionate about. This will help you showcase analytical work that truly matches your skills. But if you prefer, you can use some of the data sets from this module.

Q2: Is it OK to choose the same data set as someone else?

Ans: Yes, more than one person can analyse the same data set. Most likely your insights will be different from your peers and you will still be able to showcase your own talent as a unique solution.

Q3: Do I have to train more than two machine learning models?

Ans: It is highly recommended that you try at least two machine learning models to highlight which tool or technique improved your prediction or interpretation.

Q4: Is this an individual assignment?

Ans: You can ask for help or assistance on technical issues and general direction of your analysis, but the interpretation of the analytical output and the writing of the report should be your own.