

PRO points project

For MIS2502 – Section 3 – Shafer

Overview and Purpose:

The goal of this project is to provide MIS students with additional hands-on experience in data analysis and reinforce the concepts and methods covered in class.

For this project, students should find a new suitable dataset on the internet (that was never used in the class before) and apply Decision Tree analysis to build the prediction of the outcome variable. The process should be very similar to the regular assignment on Decision Trees (i.e., start with the same Jupyter Notebook) but applied to a new dataset.

- Students are to work with CSV data sources. No other data format should be used.
- This is an individual assignment. Not a group project or a team project.
- Students are permitted to reference, edit, and/or adapt Python code used in class. But not the datasets.
- There are thousands of data sets available. Students should try to choose a data set that is unique and distinct from what other students in the class are using. Failure to do original work may result in fewer PRO points, or no PRO points.
- Students *may* wish to start their search for data at <https://opendataphilly.org/>
- Students *may* need to use Excel to construct a suitable categorical outcome variable. If a student does this, it will be important to explain how this was done and why.

Requirements:

To successfully complete this assignment, students must:

1. Describe the Data: Select a dataset and describe the outcome variable and features for prediction. Explain how the outcome variable relates to the features and what insights can be gained from analyzing the data.
2. Find the Best Value for Minimum Split: Use the decision tree algorithm to find the best value for the minimum split for the dataset. Explain your reasoning for selecting the optimal value.
3. Find the Node with the Highest and Lowest Probability: After building the decision tree model, identify the node with the highest and lowest probability. Explain what insights can be gained from analyzing these nodes and how they relate to the outcome variable and features.
4. Provide at least 4 examples of data points and use the tree to predict the outcomes.

CONTINUED

What to submit to Canvas:

- a. Your dataset.
- b. Your Jupyter Notebook.
- c. A Word document with answers to questions 1-4.

When you are ready to request your PRO points, do the following:

1. Make sure that you have uploaded your work to the canvas class server.
2. Log in to the <https://community.mis.temple.edu/> site. Open your Dashboard. Open “Professional Achievement Points” and “Add New”. Select the option to submit a PRO point request relevant to this course, MIS2502. Your PRO points request should include a brief description of your work and your data.

Evaluation:

Students will receive 0, 25, or 50 PRO points based on the instructor’s assessment of their work.

The deadline is December 06, 2023, at 11:59 PM!