

PRO points project

For MIS2502 Section 004

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.

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.
 - This data is looking at covid causes and death were I made death into more than 50 deaths =1 and less than 0 = no death =0. This is show per day. It predicts this based on number of tested number of positives and number of hospitalized
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.
 - I used a minimum split of 4000 because it had a test accuracy of 96.85% which is high compared to the train accuracy. If you increase the minimum split the accuracy will decrease
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.
 - The least likely to have more than 50 die would be number hospilized is les than 224.5 and less than 195.5 and less than 176.5, and 157.5 which is node 4 at 1
 - The most likely to have 50 die is more than 224.5, less than 15409 more than 1462, and more than 2620 which is node 17 at .994
4. Provide at least 4 examples of data points and use the tree to predict the outcomes.
 1. 500 hospitalized is node 14 which is .786 have less than 50 deaths
 2. 2000 hospitlized is node 16 with 3.04 having more than 50 deaths
 3. 10 hospitalized is node 4 with 100% change of less than 50 death
 4. 40000 hospitalized node 20 with .0895 change of having more than 50 deaths

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 April 30, 2023, at 11:59 PM!