

Data 622 - Homework 2

Leticia Salazar

October 29, 2023

Pre-work

1. Read this blog: <https://decizone.com/blog/the-good-the-bad-the-ugly-of-using-decision-trees> which shows some of the issues with decision trees.
2. Choose a dataset from a source in Assignment #1, or another dataset of your choice.

Assignment work

1. Based on the latest topics presented, choose a dataset of your choice and create a Decision Tree where you can solve a classification problem and predict the outcome of a particular feature or detail of the data used.
2. Switch variables* to generate 2 decision trees and compare the results. Create a random forest for regression and analyze the results.
3. Based on real cases where decision trees went wrong, and 'the bad & ugly' aspects of decision trees (<https://decizone.com/blog/the-good-the-bad-the-ugly-of-using-decision-trees>), how can you change this perception when using the decision tree you created to solve a real problem?

Deliverable

1. Essay (minimum 500 word document) Write a short essay explaining your analysis, and how you would address the concerns in the blog (listed in pre-work)
2. Exploratory Analysis using R or Python (submit code + errors + analysis as notebook or copy/paste to document)

Due Date: Sunday October 29th, end of day * We are trying to just train 2 different decision trees - so switch the feature used for the first node (split) to force a different decision tree.