R Notebook

Library

Description

Decision Trees Algorithms

Pre-work

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.

Choose a dataset from a source in Assignment #1, or another dataset of your choice.

Assignment work

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.

Switch variables* to generate 2 decision trees and compare the results. Create a random forest and analyze the results.

Based on real cases where desicion 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

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)

Exploratory Analysis using R or Python (submit code + errors + analysis as notebook or copy/paste to document)

* **Note:** 1. We are trying to train 2 different decision trees to compare bias and variance - so switch the features used for the first node (split) to force a different decision tree (How did the performance change?) 2. You will create 3 models: 2 x decision trees (to compare variance) and a random forest