CIS 508 Individual Assignment 1

Paul Bernert

September 2, 2020

Instructions

A recently concluded Kaggle competition had a prize of $60,000 and was about predicting customer satisfaction. Download the training and test sets from the Kaggle site and train and validate a decision tree classification model from the Scikit-learn machine learning library. The training and test sets are also provided.

You can change the default parameters of the decision tree classifier to explore other solutions. Explore at least three other solutions by changing the parameters of the decision tree classifier, such as depth of tree, splitting criterion and maximum number of leaf nodes. Use screenshots of Python code to explain what parameters you changed and why.

In your report, discuss some of the major characteristics of the data-set/problem using the ideas of Exploratory Data Analysis. Also explain why this is a hard problem from a machine learning point of view.

Do Kaggle submissions of your test results from different decision tree models and record all Kaggle scores/rankings in your report.

General Presentation Guidelines - Explain the models you created, the process of creating those models and the results. Since you constructed multiple models, use an Excel table to compare the results from multiple models.

Presentation Length - Maximum 6 minutes. Think through your presentation before you record it. Focus on the interesting aspects of the problem, the data, the modeling approach and the results.

Test