Artificial Intelligence and Machine Learning Fundamentals

Activity 11: Random Forest Classification for Your Car Rental Company

In this section, we will optimize your classifier so that you satisfy your clients more when selecting future cars for your car fleet. We will be performing random forest and Extreme random forest classification on the car dealership dataset that you worked on in previous activity of this lesson. Suggest further improvements for the model to improve the performance of the classifier:

1. Follow steps 1 to 5 of previous activity.
2. If you are using IPython, your variables may already be accessible in your console.
3. Create a random forest and an extremely randomized trees classifier and train the models.
4. Estimate how well the two models perform on the test data. We can also calculate the accuracy scores.
5. As a first optimization technique, let's see which features more important and which features are less important. Due to randomization, removing the least important features may reduce the random noise in the model.
6. Remove the third feature from the model and retrain the classifier. Compare how well the new models fare compared to the original ones.
7. Tweak the parameterization of the classifiers a bit more.

Note that we reduced the amount of nondeterminism by allowing the maximum number of features to go up to this could eventually lead to some degree of overfitting.