

# Artificial Intelligence and Machine Learning Fundamentals

## Activity 10: Car Data Classification

In this section, we will discuss how to build a reliable decision tree model that's capable of aiding your company in finding cars that clients are likely to buy. We will be assuming that you are employed by a car rental agency who's focusing on building a lasting relationship with its clients. Your task is to build a decision tree model that classifies cars into one of four categories: unacceptable, acceptable, good, and very good.

The dataset for this activity can be accessed here:

<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>. Click the Data Folder link to download the dataset. Then, click the Dataset Description link to access the description of the attributes.

Let's evaluate the utility of your decision tree model:

1. Download the car data file from here:  
<https://archive.ics.uci.edu/ml/machinelearning-databases/car/car.data>. Add a header line to the front of the CSV file so that you can reference it in Python easily. We simply call the label Class. We named the six features after their descriptions in <https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.names>.
2. Load the dataset into Python and check if it has loaded properly. It's time to separate the training and testing data with the cross-validation (in newer versions, this is model-selection) feature of scikit-learn. We will use 10% test data.

Note that the `train_test_split` method will be available in the `model_selection` module, and not in the `cross_validation` module, starting from scikit-learn 0.20. In previous versions, `model_selection` already contains the `train_test_split` method.

Build the decision tree classifier.

3. Check the score of our model based on the test data.
4. Create a deeper evaluation of the model based on the `classification_report` feature.