

Sample Train, Test, Evaluate for Binary Classification: Automobile Price Data

Summary

This experiment demonstrates how we can build a binary classification model to predict income levels of adult individuals. The process includes training, testing and evaluating the model on the Automobile Price data.

Description

Binary Classification: Income Level Prediction

In this sample experiment we will train a binary classifier on the Automobile datasheet, to predict the price of an individual Automobile. We will show how you can perform basic data processing operations, split the dataset into training and test sets, train the model, score the test dataset, and evaluate the predictions.

Creating the Experiment

1. Drag and drop the **Automobile price data (raw)dataset** module into your experiment's workspace.
2. Add a **Clean Missing Data** module, and use the default settings, to replace missing values with zeros. Connect the dataset module output to the input port.
3. Add a **Project Columns** module, and connect the output of **Clean Missing Data** module to the input port.
4. Use the column selector to only include these columns: **Make, fuel-type, Body-style, num-of-cylinders, Horsepower, Peak-rpm, and Price**.
5. Add a **Split** module to create the testing and test sets. Set the *Fraction of rows in the first output dataset* to 0.7. This means that 70% of the data will be output to the left port and the rest to the right port of this module. We will use the left dataset for training and the right one for testing.
6. Add a **Two-Class Boosted Decision Tree** module to initialize a boosted decision tree classifier.
7. Add a **Train Model** module and connect the classifier (step 5) and the training set (left output port of the **Split** module) to the left and right input ports respectively. This module will perform the training of the classifier.
8. Add a **Score Model** module and connect the trained model and the test set (right port of the **Split** module). This module will make the predictions. You can

click on its output port to see the actual predictions and the positive class probabilities.

9. Add an **Evaluate Model** module and connect the scored dataset to the left input port. To see the evaluation results, click on the output port of the **Evaluate Model** module and select *Visualize*.

