Discussions

a. The default neural network (multilayer perceptron) produced the best total profit. The tree model is the best in terms of average profit for each customer in the retention program (n = 1,672) but its total profit is about $5,000 less than that of the neural network model.

b. Auto-Neural and SVM, again, do not perform well.

c. The preceding exercise follows the Arens-Wegman approach to use the entire data set in the comparison of the models. The approach may be useful in certain applications, but a caution is that if the so-called best model overfits the data, then the resulting profit may be misleading.

OVERSAMPLING AND RARE EVENT DETECTION

In this section, we will first explore the distribution of the target variable to motivate the need of oversampling technique in the detection of rare event. We will then discuss the profit matrix, which is unique in SAS-EM among the data mining packages we have known. The process needs the addition of the Sample node to the previous flow, as shown in Figure D.39. In the first two nodes, we will have to make changes in the settings for rare event detection.

1. Explore the Target Variable:
   Right-click on the Data Source node and then select Edit Variables. In the pop-up window, click on the Role twice to sort the Target to the top of the column (see Figure D.40).
   Highlight the Target and then click on the Explore button (at the lower-right corner of the window) as shown in Figure D.40 to view the distribution of the variable. The result is shown in Figure D.41.
   The bar chart indicates that there are about 15% unsatisfied customers.