

Traceis™ Data Exploration Studio k-nearest neighbors

1 Select the prediction step

In this example, the body fat dataset was used. First, click on the prediction step

2 Select the kNN tab

The screenshot displays the Traceis Data Exploration Studio interface. On the left, a vertical navigation pane shows a list of steps: Definition, Preparation, Prepare, Implementation, Tables and graphs, Statistics, Grouping, Prediction, and Deployment. The 'Prediction' step is highlighted with a blue bar and a large white circle containing the number '1'. In the main workspace, the 'kNN' tab is selected, and a large white circle containing the number '2' is positioned over the top of the configuration area. The configuration area includes a list of variables (Density, Percent body fat, Age(years), Weight(lbs), Height(inches), Neck(cm), Chest(cm), Abdomen(cm)) with 'Percent body fat' selected as the response variable. Below this, there are options for similarity methods (Euclidean, Jaccard), defining 'k' (Manually set k to 10, or Automatically set k with a range from 1 to 100), and cross-validation percentage (10%). At the bottom, there are fields for prediction and residual names, and buttons for 'Save model' and 'Build model'. A 'Selected Items' list at the bottom right shows the variables: Density, Percent body fat, Age(years), Weight(lbs), Height(inches), Neck(cm), Chest(cm), Abdomen(cm), Hip(cm), Thigh(cm), Knee(cm), Ankle(cm), Biceps(cm), R.

MAKING SENSE OF DATA

1 Select descriptors

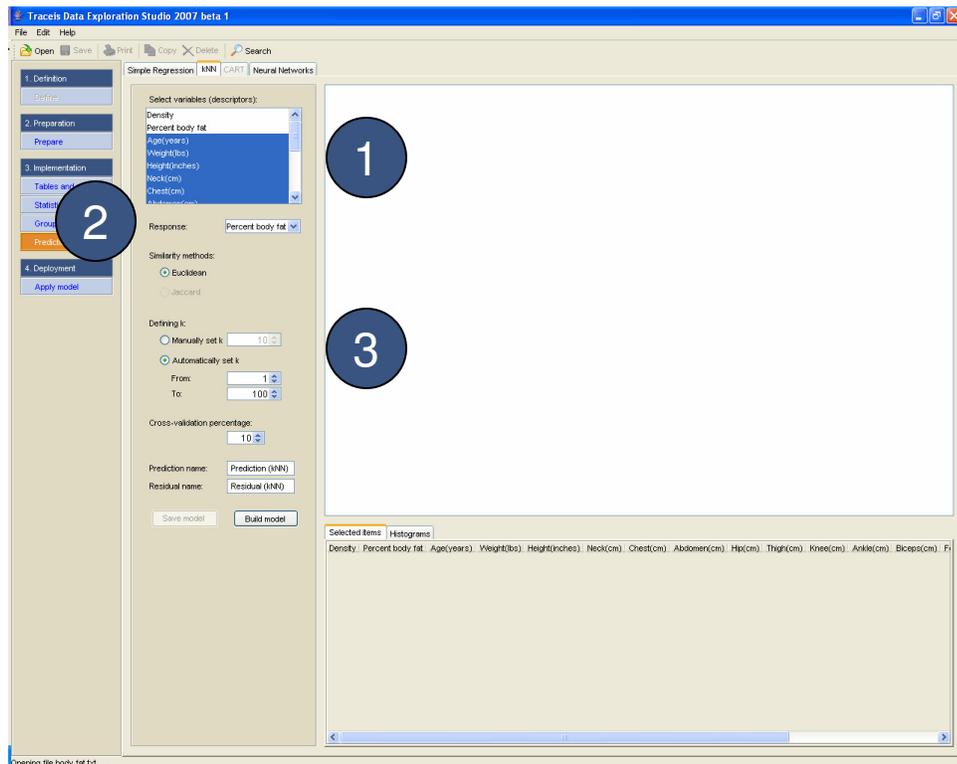
In this example, all descriptors were selected except density and percent body fat.

2 Select response

In this example, percent body fat was selected as the response value.

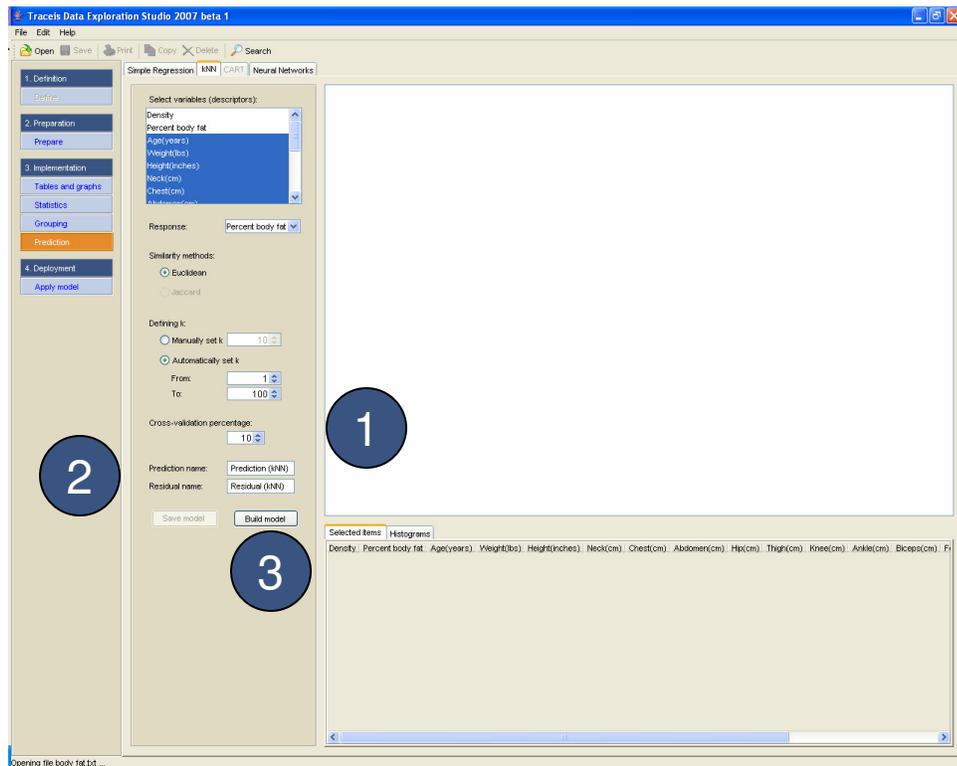
3 Select k

A value for k must be set, either manually or by selecting the best value from a specified range. In this example, values for k between 1 and 100 are tested and the best value chosen.



MAKING SENSE OF DATA

- 1** Enter the cross-validation percentage
In this example, a model is to be built using a 10% cross-validation.
- 2** Enter names for the prediction and residual variables
- 3** Display the model



MAKING SENSE OF DATA

1 View model summary

The value of k used to build the model is shown

2 View model assessment

An assessment of the model is presented. The assessment of a model built where the response is continuous is shown. When the model is built using a categorical response, a contingency table of actual values versus predicted values is presented.

