Overall Organization of Clementine Components

After purchasing Clementine from Integral Solutions in 1999, SPSS teamed up with Daimler-Benz and NCR Corporation to produce the Cross Industry Standard Process for Data Mining (CRISP-DM). One of the authors of this book contributed to that project. The Clementine package was integrated with CRISP-DM to help guide the modeling process flow. Today, Clementine can be purchased in standalone client mode or with a number of modules, including a client/server version. The modules of version 11.1 include the following components:

The Basic Module

- Classification and Regression Tree node (CART or C&RT)
- Quest node for binary classification
- CHAID node for creating decision trees based on the Chi-square statistic
- K-Means node for clustering
- Generalized Rule Induction node for creating rule sets
- Factor/PCA node for data reduction
- Linear Regression node for ordinary linear regression

Classification Module

- Binary Classifier node
- Neural Net node for classification
- C5.0 node for classification
- Decision List node for creating rule sets of a group of segments
- Time-Series node to create Autoregressive Integrated Moving Average (ARIMA) models
- Feature Selection node for ranking input variables by importance in predicting the target variable
- Logistic Regression node for creating logit models
- Discriminant Analysis node for creating parametric versions of logistic regression models
- Generalized Linear Model (GLM) for GLM modeling
- Self-Learning Response Model node for retraining a model on as few as one new case

Segmentation Module

- Kohonen net for unsupervised clustering using a form of a neural net
- TwoStep node for clustering in two steps in tandem to reduce the number of cases to subclusters and then clustering the subclusters

Association Module

- Apriori node optimized for categorical rule induction
- CARMA node for unsupervised association of cases
- Sequence node to discover association rules in sequential or time-series data