Overall Organization of SAS-EM Version 5.3 Components

As in SPSS Clementine, the SAS-EM data mining process consists of a process flow diagram, which is a form of a graphical user interface, where you can add nodes, modify nodes, connect nodes with arrows for the direction of flow of the computations, modify nodes, and save the entire workspace as a data mining project. Like in SPSS Clementine, this workspace is designed for use by business analysts (in business, industry, governmental agencies, etc.) with little statistical expertise but who can navigate through the data mining methodology fairly easily. However, you probably can’t give it to your assistant or any person randomly picked off the street and expect that person to make an “intelligent” model or even to get the model to run. At the same time, the quantitative statistical or engineering expert can go “behind the nodes” to customize the analytical processes.

The analytical tools include

• Clustering
• Association and sequence discovery
• Marketbasket analysis
• Path analysis
• Self-organizing maps/Kohonen
• Variable selection (analogous to Feature Selection as termed in STATISTICA Data Miner)
• Decision trees and gradient boosting
• Linear and logistic regression
• Two-stage modeling
• Partial least squares
• Support Vector Machines
• Neural networking

Data preparation tools include

• Outlier detection
• Variable transformations
• Variable clustering
• Interactive binning
• Principal components
• Rule building and induction
• Data imputation
• Random sampling
• Partitioning of data sets (into train, test, and validate data sets)