

Data mining tool – SPSS Modeler

In this research, SPSS Modeler is employed as data mining tool for analysis. The difference between Modeler and other software is that its data processing is through the use of nodes, which are then connected together to form a stream frame. In addition, data visualization can be presented to users after the mining process has been completed. The data processing in Clementine is performed through the use of nodes, which are then connected together to form a stream frame. In addition, data visualization can be presented to users after the mining process has completed. The nodes can be divided into six categories: the source node, record options node, field options node, graphs node, modeling node, and output node.

Source node: The source selection includes those nodes that can undergo the data connection node via the Open Database Connectivity (ODBC) and relational database management system. It also includes nodes that can input all sorts of common file contents.

Record Options node: This is for the recording and correction of data. These operations are very important in the data investigation stage of data interpretation and data preparation stage, because these operations enable the data to fulfill specific business needs.

Field Options node: The field option node can help the user do modeling and data preparation for the logical data design stage.

Graphs node: This is one of the stages in the data mining process that uses graphs for exploratory data analysis. Another purpose is to examine the new record option's distribution and relationships.

Modeling node: Modeling is the core of the data investigation process. The modeling method of this node enables the user to retrieve new information from the data and to form a forecast model. These modeling methods include derived Machine Learning (ML), Artificial Intelligence (AI), and Statistics, etc. All these methods have their own advantages and are suitable for specific types of problem. The algorithm includes:

- (1) Decision Tree (C5.0, CART).
- (2) Neural Net and RBF Function.
- (3) Association Rule (Apriori, GRI).
- (4) Sequence Detection.
- (5) Clustering Analysis (K-means, Two-step and Kohonen).
- (6) Regression (Linear Regression, Logistic Regression).
- (7) Factor Analysis and PCA.

Output node: This offers one method to achieve data that is related to the users and the model. It can output all types of data in different forms to other software interfaces.

SPSS Modeler provides a different classification of clustering in the modeling node; the data analysis process and the main set of nodes are linked together, to complete the analysis of the data stream processing. Therefore, this study implements ODBC bridge, into Modeler data, in order to establish the analysis process, and employs the SPSS Modeler to analyze data using K-means clustering, followed by application of the Apriori algorithm on each cluster to analyze association rules. These data mining model and processes are summarized in Figure 5.

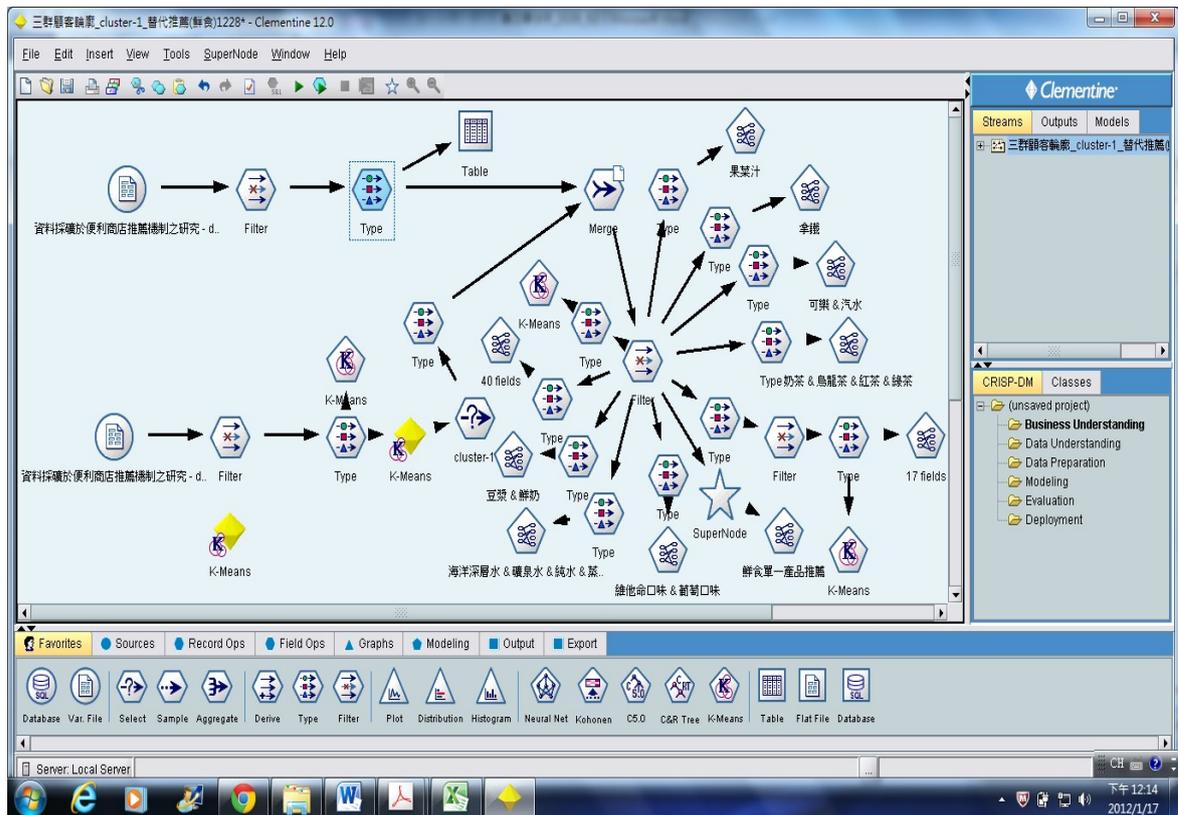


Figure: Complete stream map data nodes of cluster analysis and association rule on this study