

Appendix G

Overview: CART

Appendix G provides a brief overview of the binary recursive partitioning method called the Classification And Regression Tree (CART) analysis that was applied to ozone and meteorology data for several midwestern urban areas (Kenski 2009). This analysis is described in Section 7.3 (“CART Analyses”) of the Attainment Demonstration.

The CART model is a robust, easy-to-use decision tree that automatically sifts large, complex databases, searching for and isolating significant patterns and relationships (Salford systems, 2009). This discovered knowledge is then used to generate reliable, easy-to-grasp predictive models for applications to reveal data relationships that could remain hidden using other analytical tools (Salford systems, 2009).

The CART model has several advantages as a tool for data mining and predictive modeling (Kenski, 2009). The tree produced represents a model or decision tree in which each node (branch) is determined by splitting the dataset on the basis of the one variable that results in the best separation as defined by values of the dependent variable (in this case, ozone concentration). The splitting rule is expressed in natural language – for example, asking the question “is temperature less than 75°F?”. Consequently the output trees are easy to interpret. At every branch, every variable is tested for its usefulness in further splitting. This exhaustive search for splitters can make CART computationally intensive.

References

Kenski, D., 2009: Updated CART Analysis for Ozone through 2008. Draft. LADCO, Feb 26, 3 pp,

Salford Systems, 2009: Product Overview: CART. <http://www.salfordsystems.com/cart.php>