1. Title: Car Evaluation Database

2. Sources:
   (a) Creator: Marko Bohanec
   (b) Donors: Marko Bohanec (marko.bohanec@ijs.si)
               Blaz Zupan (blaz.zupan@ijs.si)
   (c) Date: June, 1997

3. Past Usage:

   The hierarchical decision model, from which this dataset is
   derived, was first presented in

   M. Bohanec and V. Rajkovic: Knowledge acquisition and explanation for
   multi-attribute decision making. In 8th Intl Workshop on Expert
   Systems and their Applications, Avignon, France. pages 59-78, 1988.

   Within machine-learning, this dataset was used for the evaluation
   of HINT (Hierarchy INduction Tool), which was proved to be able to
   completely reconstruct the original hierarchical model. This,
   together with a comparison with C4.5, is presented in

   B. Zupan, M. Bohanec, I. Bratko, J. Demsar: Machine learning by
   function decomposition. ICML-97, Nashville, TN. 1997 (to appear)

4. Relevant Information Paragraph:

   Car Evaluation Database was derived from a simple hierarchical
   decision model originally developed for the demonstration of DEX
   (M. Bohanec, V. Rajkovic: Expert system for decision
   making. Sistemica 1(1), pp. 145-157, 1990.). The model evaluates
   cars according to the following concept structure:

   | | |
   |---|---|
   | CAR | car acceptability |
   | . PRICE | overall price |
   | . . buying | buying price |
   | . . maint | price of the maintenance |
   | . TECH | technical characteristics |
   | . . COMFORT | comfort |
   | . . . doors | number of doors |
   | . . . persons | capacity in terms of persons to carry |
   | . . . lug_boot | the size of luggage boot |
   | . . safety | estimated safety of the car |

   Input attributes are printed in lowercase. Besides the target
   concept (CAR), the model includes three intermediate concepts:
   PRICE, TECH, COMFORT. Every concept is in the original model
   related to its lower level descendants by a set of examples (for
   these examples sets see http://www-ai.ijs.si/BlazZupan/car.html).

The Car Evaluation Database contains examples with the structural information removed, i.e., directly relates CAR to the six input attributes: buying, maint, doors, persons, lug_boot, safety.

Because of known underlying concept structure, this database may be particularly useful for testing constructive induction and structure discovery methods.

5. Number of Instances: 1728
   (instances completely cover the attribute space)

6. Number of Attributes: 6

7. Attribute Values:

   | | |
   |---|---|
   | buying | v-high, high, med, low |
   | maint | v-high, high, med, low |
   | doors | 2, 3, 4, 5-more |
   | persons | 2, 4, more |
   | lug_boot | small, med, big |
   | safety | low, med, high |

8. Missing Attribute Values: none

9. Class Distribution (number of instances per class)

   | class | N | N[%] |
   |---|---|---|
   | unacc | 1210 | (70.023 %) |
   | acc | 384 | (22.222 %) |
   | good | 69 | ( 3.993 %) |
   | v-good | 65 | ( 3.762 %) |