1. Title: Nursery Database

2. Sources:
   (a) Creator: Vladislav Rajkovic et al. (13 experts)
   (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. Olave, V. Rajkovic, M. Bohanec: An application for admission in public school systems. In (I. Th. M. Snellen and W. B. H. J. van de Donk and J.-P. Baquiast, editors) Expert Systems in Public Administration, pages 145-160. Elsevier Science Publishers (North Holland)}, 1989.

   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:

   Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. It was used during several years in 1980's when there was excessive enrollment to these schools in Ljubljana, Slovenia, and the rejected applications frequently needed an objective explanation. The final decision depended on three subproblems: occupation of parents and child's nursery, family structure and financial standing, and social and health picture of the family. The model was developed within expert system shell for decision making DEX (M. Bohanec, V. Rajkovic: Expert system for decision making. Sistemica 1(1), pp. 145-157, 1990.).

   The hierarchical model ranks nursery-school applications according to the following concept structure:

NURSERY	Evaluation of applications for nursery schools
. EMPLOY	Employment of parents and child's nursery
. . parents	Parents' occupation
. . has_nurs	Child's nursery
. STRUCT_FINAN	Family structure and financial standings
. . STRUCTURE	Family structure
. . . form	Form of the family
. . . children	Number of children
. . housing	Housing conditions
. . finance	Financial standing of the family
. SOC_HEALTH	Social and health picture of the family
. . social	Social conditions
. . health	Health conditions

Input attributes are printed in lowercase. Besides the target concept (NURSERY) the model includes four intermediate concepts: EMPLOY, STRUCT_FINAN, STRUCTURE, SOC_HEALTH. 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/nursery.html).

The Nursery Database contains examples with the structural information removed, i.e., directly relates NURSERY to the eight input attributes: parents, has_nurs, form, children, housing, finance, social, health.

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

5. Number of Instances: 12960
   (instances completely cover the attribute space)

6. Number of Attributes: 8

7. Attribute Values:

parents	usual, pretentious, great_pret
has_nurs	proper, less_proper, improper, critical, very_crit
form	complete, completed, incomplete, foster
children	1, 2, 3, more
housing	convenient, less_conv, critical
finance	convenient, inconv
social	non-prob, slightly_prob, problematic
health	recommended, priority, not_recom

8. Missing Attribute Values: none

9. Class Distribution (number of instances per class)

class	N	N[%]
not_recom	4320	(33.333 %)
recommend	2	(0.015 %)
very_recom	328	(2.531 %)
priority	4266	(32.917 %)
spec_prior	4044	(31.204 %)