|  |  |
| --- | --- |
| **Car *Evaluation Data Set***  https://archive.ics.uci.edu/ml/assets/MLimages/Large19.jpg  **Abstract:** Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 1728 | **Area:** | N/A |
| **Attribute Characteristics:** | Categorical | **Number of Attributes:** | 6 | **Date Donated** | 1997-06-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | No | **Number of Web Hits:** | 1016787 |

**Source:**

Creator:   
  
Marko Bohanec   
  
Donors:   
  
1. Marko Bohanec (marko.bohanec **'@'** ijs.si)   
2. Blaz Zupan (blaz.zupan **'@'** ijs.si)

**Data Set Information:**

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 [[Web Link]](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.

**Attribute Information:**

Class Values:   
  
unacc, acc, good, vgood   
  
Attributes:   
  
buying: vhigh, high, med, low.   
maint: vhigh, high, med, low.   
doors: 2, 3, 4, 5more.   
persons: 2, 4, more.   
lug\_boot: small, med, big.   
safety: low, med, high.

**Relevant Papers:**

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.   
[[Web Link]](http://rexa.info/paper/5e2ae6fa6748dfe24067bb2b59823f2df3f7ed73)   
  
B. Zupan, M. Bohanec, I. Bratko, J. Demsar: Machine learning by function decomposition. ICML-97, Nashville, TN. 1997 (to appear)   
[[Web Link]](http://rexa.info/paper/0f23f96c4a89bbb221a151f5db381924c17a6eaa)

**Papers That Cite This Data Set:**

Qingping Tao Ph. D. [MAKING EFFICIENT LEARNING ALGORITHMS WITH EXPONENTIALLY MANY FEATURES](http://rexa.info/paper/8e674b6eff0f726ba6fff46ef6eaff968dc89f39). Qingping Tao A DISSERTATION Faculty of The Graduate College University of Nebraska In Partial Fulfillment of Requirements. 2004. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#8e674b6eff0f726ba6fff46ef6eaff968dc89f39)].  
  
Jianbin Tan and David L. Dowe. [MML Inference of Decision Graphs with Multi-way Joins and Dynamic Attributes](http://rexa.info/paper/14f025e969e3a0418fd852ee46e54039ab3f216a). Australian Conference on Artificial Intelligence. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#14f025e969e3a0418fd852ee46e54039ab3f216a)].  
  
Daniel J. Lizotte and Omid Madani and Russell Greiner. [Budgeted Learning of Naive-Bayes Classifiers](http://rexa.info/paper/f14d3edaeac2280dc4e49948d9d0fc1159bd05ca). UAI. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#f14d3edaeac2280dc4e49948d9d0fc1159bd05ca)].  
  
Marc Sebban and Richard Nock and Stéphane Lallich. [Stopping Criterion for Boosting-Based Data Reduction Techniques: from Binary to Multiclass Problem](http://rexa.info/paper/1b77c2b6fd8a261af286cf411879f9f520824bd6). Journal of Machine Learning Research, 3. 2002. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#1b77c2b6fd8a261af286cf411879f9f520824bd6)].  
  
Nikunj C. Oza and Stuart J. Russell. [Experimental comparisons of online and batch versions of bagging and boosting](http://rexa.info/paper/5193dfc0a9d39b5f86fe360d6beff81aa9b7390e). KDD. 2001. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#5193dfc0a9d39b5f86fe360d6beff81aa9b7390e)].  
  
Marc Sebban and Richard Nock and Jean-Hugues Chauchat and Ricco Rakotomalala. [Impact of learning set quality and size on decision tree performances](http://rexa.info/paper/d2ad474fa4c9b346e9ac8f41900cea7d4917c7ac). Int. J. Comput. Syst. Signal, 1. 2000. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#d2ad474fa4c9b346e9ac8f41900cea7d4917c7ac)].  
  
Iztok Savnik and Peter A. Flach. [Discovery of multivalued dependencies from relations](http://rexa.info/paper/bc80295973a43d3806ff4dfe83e5724260301c33). Intell. Data Anal, 4. 2000. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#bc80295973a43d3806ff4dfe83e5724260301c33)].  
  
Jie Cheng and Russell Greiner. [Comparing Bayesian Network Classifiers](http://rexa.info/paper/47354ca48da5014e0a8f5e4da7f3a7e9aaa6e9e5). UAI. 1999. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#47354ca48da5014e0a8f5e4da7f3a7e9aaa6e9e5)].  
  
Hyunwoo Kim and Wei-Yin Loh. [Classification Trees with Bivariate Linear Discriminant Node Models](http://rexa.info/paper/4fe77f0de67f4dda7e7174b944840d4d49fe15ac). Department of Statistics Department of Statistics University of Tennessee University of Wisconsin. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#4fe77f0de67f4dda7e7174b944840d4d49fe15ac)].  
  
Daniel J. Lizotte. [Library Release Form Name of Author](http://rexa.info/paper/217beab6a7a7b64dc929c3c5fdb42e812f8b2431). Budgeted Learning of Naive Bayes Classifiers. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#217beab6a7a7b64dc929c3c5fdb42e812f8b2431)].  
  
Nikunj C. Oza and Stuart J. Russell. [Online Bagging and Boosting](http://rexa.info/paper/3e9ebff12a232c9f091156827e92c55d259b95f3). Computer Science Division University of California. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#3e9ebff12a232c9f091156827e92c55d259b95f3)].  
  
Daniel J. Lizotte and Omid Madani and Russell Greiner. [Budgeted Learning, Part II: The Na#ve-Bayes Case](http://rexa.info/paper/7e787ada7263180d2a9bad6a3c490e7f8b0d4cd0). Department of Computing Science University of Alberta. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#7e787ada7263180d2a9bad6a3c490e7f8b0d4cd0)].  
  
Huan Liu. [A Family of Efficient Rule Generators](http://rexa.info/paper/8f5ae7219e74a85e3f722b58b3fedb30eab7a1d7). Department of Information Systems and Computer Science National University of Singapore. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#8f5ae7219e74a85e3f722b58b3fedb30eab7a1d7)].  
  
Zhiqiang Yang and Sheng Zhong and Rebecca N. Wright. [Privacy-Preserving Classification of Customer Data without Loss of Accuracy](http://rexa.info/paper/6aae20aa5fd96e903634bb73244782652cd4e947). Computer Science Department, Stevens Institute of Technology. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#6aae20aa5fd96e903634bb73244782652cd4e947)].  
  
Jos'e L. Balc'azar. [Rules with Bounded Negations and the Coverage Inference Scheme](http://rexa.info/paper/80d25f1152aed1f55fc47bd0f450312debe7617e). Dept. LSI, UPC. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#80d25f1152aed1f55fc47bd0f450312debe7617e)].  
  
Shi Zhong and Weiyu Tang and Taghi M. Khoshgoftaar. [Boosted Noise Filters for Identifying Mislabeled Data](http://rexa.info/paper/e2e72927eb590e2b7daf9095e42d6ed43ce21e68). Department of Computer Science and Engineering Florida Atlantic University. [[View Context](https://archive.ics.uci.edu/ml/support/car+evaluation#e2e72927eb590e2b7daf9095e42d6ed43ce21e68)].

**Data set Description:**

Sources:

(a) Creator: Marko Bohanec

(b) Donors: Marko Bohanec (marko.bohanec@ijs.si)

Blaz Zupan (blaz.zupan@ijs.si)

(c) Date: June, 1997

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)

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.

Number of Instances: 1728

(instances completely cover the attribute space)

Number of Attributes: 6

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

Missing Attribute Values: none

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 %)