Car Evaluation Data Set

Download: Data Folder, Data Set Description

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:**	412232

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]).

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]

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

Papers That Cite This Data Set[1]:

Qingping Tao Ph. D. MAKING EFFICIENT LEARNING ALGORITHMS WITH EXPONENTIALLY MANY FEATURES. Qingping Tao A DISSERTATION Faculty of The Graduate College University of Nebraska In Partial Fulfillment of Requirements. 2004. [View Context].

Jianbin Tan and David L. Dowe. MML Inference of Decision Graphs with Multi-way Joins and Dynamic Attributes. Australian Conference on Artificial Intelligence. 2003. [View Context].

Daniel J. Lizotte and Omid Madani and Russell Greiner. Budgeted Learning of Naive-Bayes Classifiers. UAI. 2003. [View Context].

Marc Sebban and Richard Nock and Stéphane Lallich. Stopping Criterion for Boosting-Based Data Reduction

Techniques: from Binary to Multiclass Problem. Journal of Machine Learning Research, 3. 2002. [View Context].

Nikunj C. Oza and Stuart J. Russell. Experimental comparisons of online and batch versions of bagging and boosting. KDD. 2001. [View Context].

Marc Sebban and Richard Nock and Jean-Hugues Chauchat and Ricco Rakotomalala. Impact of learning set quality and size on decision tree performances. Int. J. Comput. Syst. Signal, 1. 2000. [View Context].

Iztok Savnik and Peter A. Flach. Discovery of multivalued dependencies from relations. Intell. Data Anal, 4. 2000. [View Context].

Jie Cheng and Russell Greiner. Comparing Bayesian Network Classifiers. UAI. 1999. [View Context].

Hyunwoo Kim and Wei-Yin Loh. Classification Trees with Bivariate Linear Discriminant Node Models. Department of Statistics Department of Statistics University of Tennessee University of Wisconsin. [View Context].

Daniel J. Lizotte. Library Release Form Name of Author. Budgeted Learning of Naive Bayes Classifiers. [View Context].

Nikunj C. Oza and Stuart J. Russell. Online Bagging and Boosting. Computer Science Division University of California. [View Context].

Daniel J. Lizotte and Omid Madani and Russell Greiner. Budgeted Learning, Part II: The Na#ve-Bayes Case. Department of Computing Science University of Alberta. [View Context].

Huan Liu. A Family of Efficient Rule Generators. Department of Information Systems and Computer Science National University of Singapore. [View Context].

Zhiqiang Yang and Sheng Zhong and Rebecca N. Wright. Privacy-Preserving Classification of Customer Data without Loss of Accuracy. Computer Science Department, Stevens Institute of Technology. [View Context].

Jos'e L. Balc'azar. Rules with Bounded Negations and the Coverage Inference Scheme. Dept. LSI, UPC. [View Context].

Shi Zhong and Weiyu Tang and Taghi M. Khoshgoftaar. Boosted Noise Filters for Identifying Mislabeled Data. Department of Computer Science and Engineering Florida Atlantic University. [View Context].

Citation Request:

Please refer to the Machine Learning Repository's citation policy

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info