

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU
LAB2

CODE BOOK
DATASET CAR EVALUATION
DATABASE

Họ và tên : Lưu Quang Tiến Hoàng
MSSV : 20521342
Lớp : DS103.M21

Tên bộ dữ liệu:

Car Evaluation Database

Nguồn thu thập và cách thu thập:

Creator: Marko Bohanec

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

Blaz Zupan (blaz.zupan@ijs.si)

Date: June, 1997

Thông tin liên quan:

Cơ sở dữ liệu đánh giá xe hơi có nguồn gốc từ một mô hình quyết định phân cấp đơn giản ban đầu được phát triển để trình diễn DEX (M. Bohanec, V. Rajkovic: Expert system for decision making. Sistemica 1(1), pp. 145-157, 1990.).

Mô hình đánh giá xe theo cấu trúc khái niệm sau:

CAR: tính chất có thể chấp nhận xe

PRICE: giá tổng thể

buying: giá mua

maint: giá bảo trì

TECH: đặc điểm kỹ thuật

COMFORT: tiện nghi

doors: số lượng cửa

persons: số lượng người xe có thể chở

lug_boot: kích thước của cốp xe . .

safely: ước tính an toàn của xe

Thuộc tính đầu vào được in bằng chữ thường. Bên cạnh mục tiêu Khái niệm (CAR), mô hình bao gồm ba khái niệm trung gian: PRICE, TECH, COMFORT. Mọi khái niệm đều có trong mô hình ban đầu liên quan đến hậu duệ cấp thấp hơn của nó bằng một tập hợp các ví dụ (đối với các bộ ví dụ này xem <http://www-ai.ijs.si/BlazZupan/car.html>).

Cơ sở dữ liệu đánh giá xe hơi chứa các ví dụ với cấu trúc thông tin bị xóa, tức là, liên quan trực tiếp đến CAR với sáu thuộc tính đầu vào: buying, maint, doors, persons, lug_boot, safety.

Do cấu trúc khái niệm cơ bản đã biết, cơ sở dữ liệu này có thể đặc biệt hữu ích để thử nghiệm các phương pháp cảm ứng và khám phá cấu trúc mang tính xây dựng.

Kích thước dữ liệu:

File raw data:

Số thuộc tính: 7

Kích thước: 51KB

File tidy data:

Số thuộc tính: 7

Tidydata được lưu thành file: tidy_car.csv

Kích thước: 77KB

Số dòng dữ liệu: 1728

Thông tin các thuộc tính:

buying: giá mua

Phân loại:

“v-high”: Rất cao

“high”: Cao

“med”: Trung bình

“low”: Thấp

maint: giá bảo trì

Phân loại:

“v-high”: Rất cao

“high”: Cao

“med”: Trung bình

“low”: Thấp

doors: số cánh cửa của xe

Phân loại:

“2”: 2 cánh

“3”: 3 cánh

“4”: 4 cánh

“5-more”: 5 cánh hoặc hơn

persons: Số người xe có thể chở

Phân loại:

“2”: 2 chỗ

“4”: 4 chỗ

lug_boot: kích thước của cốp xe

Phân loại:

“small”: Nhỏ

“med”: Vừa, trung bình

“big”: To, lớn

safety: ước tính an toàn của xe

“high”: Cao

“med”: Trung bình

“low”: Thấp

Phân phối lớp (số lượng phiên bản trên mỗi lớp)

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

Missing Attribute Values:

Không có

Instruction list:

```
rm(list=ls())
```

```
mydata <- read.csv('dataset/car.data', header = FALSE)
```

```
vab<-c('buying', 'maint', 'doors', 'person', 'lug_boot', 'safety', 'class')
```

```
colnames(mydata)<-vab
```

```
write.csv(mydata,"tidy_car.csv",row.names = FALSE)
```

--- HẾT ---