

# ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH



**UIT**  
Trường Đại học  
Công nghệ Thông tin

**Khoa Khoa học  
và Kỹ thuật Thông tin**

## CODEBOOK MÔ TẢ BỘ DỮ LIỆU CAR EVALUATION

Môn học: Thu thập và tiền xử lý dữ liệu - DS103.N21

Tên: Nguyễn Thị Huyền Trang

MSSV: 21520488

**TP HỒ CHÍ MINH, 2023**

# MỤC LỤC

<b>CODEBOOK</b> .....	3
<b>Bộ dữ liệu:</b> .....	3
<b>Codebook</b> .....	3
<b>Raw data</b> .....	3
<b>Tidy data</b> .....	3
<b>Instruction list</b> .....	3
<b>THAM KHẢO</b> .....	5

# CODEBOOK

**Bộ dữ liệu:** Car Evaluation

## Codebook:

Thông tin	Nội dung
Tên bộ dữ liệu	Car Evaluation
Nguồn thu thập và cách thức thu thập	Bộ dữ liệu được thu thập bởi Marko Bohanec và Blaz Zupan từ trường đại học Ljubljana, Slovenia. Họ thu thập dữ liệu bằng cách sử dụng kết hợp kiến thức chuyên môn và khảo sát những người bán xe.
Số thuộc tính	6
Thông tin tên các thuộc tính	Buying: Giá mua xe Maint: Chi phí duy trì Doors: Số lượng cửa Persons: Số người có thể chở được Lug_boot: Kích thước chỗ để hành lý Safety: Ước tính an toàn của xe ô tô Class: Phân loại mỗi xe ô tô
Thông tin tác giả	a) Người tạo: Marko Bohanec (b) Người đóng góp: Marko Bohanec (marko.bohanec@ijs.si) Blaz Zupan (blaz.zupan@ijs.si) (c) Thời gian: Tháng 6, 1997

## Raw data:

Raw data gồm file car.data

## Tidy data:

Xử lý file car.data, dữ liệu sau khi xử lý được lưu dưới file car.csv

## Instruction list:

```
rm(list=ls())  
myFiles <- list.files(path="car-data/", pattern="car.data")  
file <- read.csv(paste("car-data/", myFiles, sep=""), sep=";", header = FALSE)  
dataset <- file
```

```
variables <- c("buying", "maint", "doors", "persons","lug_boot","safety",  
"class")  
colnames(dataset) <- variables  
write.csv(dataset, file = "car.csv",row.names = TRUE)
```

# THAM KHẢO

Car Evaluation Database, <https://archive.ics.uci.edu/ml/machine-learning-databases/car/>