

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 WINE RECOGNITION

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

CODEBOOK	3
Bộ dữ liệu:	3
Codebook	3
Raw data	4
Tidy data	4
Instruction list	4
THAM KHẢO	5

CODEBOOK

Bộ dữ liệu: Wine Recognition

Codebook:

Thông tin	Nội dung
Tên bộ dữ liệu	Wine Recognition
Nguồn thu thập và cách thức thu thập	Dữ liệu này là kết quả phân tích thu thập được của rượu vang được trồng trong cùng một vùng ở Ý nhưng có nguồn gốc từ ba giống cây trồng khác nhau.
Số thuộc tính	13
Kích thước	177
Thông tin tên các thuộc tính	<p>Class:</p> <p>Alcohol: Độ cồn của rượu (3 mức: 1,2,3)</p> <p>Malid acid: Phần trăm axit malid trong rượu (numeric)</p> <p>Ash: Lượng tro trong rượu vang (numeric)</p> <p>Alcalinity of ash: Độ kiềm của tro trong rượu vang (numeric)</p> <p>Magnesium: Lượng magie trong rượu (numeric)</p> <p>Total phenols: Tổng lượng hợp chất phenols trong rượu (numeric)</p> <p>Flavanoids: Tổng lượng hợp chất flavanoids trong rượu (numeric)</p> <p>Nonflavanoid phenols: Tổng lượng hợp chất phi flavanoids trong rượu (numeric)</p> <p>Proanthocyanins: Tổng lượng hợp chất proanthocyanidin trong rượu vang (numeric)</p> <p>Color intensity: Cường độ màu của rượu (numeric)</p> <p>Hue: Màu sắc hoặc tông màu của rượu vang</p> <p>OD280/OD315 of diluted wines: Tỷ lệ hấp thụ của rượu vang ở mức 280nm và 315 nm (numeric)</p> <p>Proline: Lượng proline trong rượu (numeric)</p>
Thông tin tác giả	a) Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of and Food Analysis and Technologies, Via Brigata Salerno,

	16147 Genoa, Italy. b) Stefan Aeberhard, email: stefan@coral.cs.jcu.edu.au c) July 1991
--	---

Raw data:

Raw data gồm file wine.data

Tidy data:

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

Instruction list:

```
rm(list=ls())
myFiles <- list.files(path="wines-data/", pattern="wine.data")
file <- read.csv(paste("wines-data/", myFiles, sep=""), sep=",",header = FALSE)
dataset <- file
variables <- c("Class","Alcohol","Malic acid","Ash","Alcalinity of
ash","Magnesium","Total phenols","Flavanoids","Nonflavanoid
phenols","Proanthocyanins","Color intensity","Hue","OD280/OD315 of diluted
wines","Proline")
colnames(dataset) <- variables
write.csv(dataset, file = "Wine.csv",row.names = FALSE)
```

THAM KHẢO

Wine Recognition Database, <https://archive.ics.uci.edu/ml/datasets/Wine>