

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 WINE RECOGNITION DATA

Họ và tên : Lưu Quang Tiến Hoàng

MSSV : 20521342

Lớp : DS103.M21

Tên bộ dữ liệu:

Wine Recognition Data

Updated Sept 21, 1998 by C.Blake : Added attribute information

Nguồn thu thập và cách thu thập:

Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.

Stefan Aeberhard, email: stefan@coral.cs.jcu.edu.au

July 1991

Thông tin liên quan:

Những dữ liệu này là kết quả của một phân tích hóa học về rượu vang được lên men trong cùng một khu vực ở Ý nhưng có nguồn gốc từ ba giống khác nhau.

Phân tích đã xác định số lượng của 13 thành phần được tìm thấy trong mỗi loại rượu vang.

Tác giả của bộ dữ liệu nghĩ rằng bộ dữ liệu ban đầu có khoảng 30 biến, nhưng vì một số lý do, họ chỉ thể hiện phiên bản 13 chiều.

Họ đã có một danh sách về 30 biến hoặc hơn, nhưng:

Họ đã mất nó

Họ sẽ không biết 13 biến nào được bao gồm trong tập hợp.

Kích thước dữ liệu:**File raw data:**

Số thuộc tính: 13

Kích thước: 11KB

File tidy data:

Số thuộc tính: 13

Tidydata được lưu thành file: tidy_wine.csv

Kích thước: 12KB

Số dòng dữ liệu: 178

Thông tin các thuộc tính:

Các thuộc tính là (donated by Riccardo Leardi, riclea@anchem.unige.it):

- 1) **Alcohol:** còn gọi là rượu, đồ uống có cồn
- 2) **Malic acid:** Axit malic là chất lỏng, tan nhiều trong nước. Có mùi thơm của táo. Thành phần chính làm nên vị chua của táo là axit này.
- 3) **Ash:** Hàm lượng tro
- 4) **Alcalinity of ash:** Độ kiềm trong hàm lượng tro
- 5) **Magnesium:** Magiê
- 6) **Total phenols:** Hàm lượng phenols
- 7) **Flavanoids:** chất chuyển hóa trung gian của thực vật.
- 8) **Nonflavanoid phenols:**
- 9) **Proanthocyanins:** một loại polyphenol được tìm thấy trong nhiều loại thực vật, chẳng hạn như nam việt quất, việt quất và hạt nho.
- 10) **Color intensity**
- 11) **Hue**
- 12) **OD280/OD315 of diluted wines**
- 13) **Proline**

Đối với mỗi thuộc tính:

Tất cả các thuộc tính là liên tục

Không có số liệu thống kê có sẵn, nhưng đề nghị tiêu chuẩn hóa các biến cho một số mục đích sử dụng nhất định (ví dụ: đối với chúng tôi với các phân loại không phải là bất biến quy mô)

LƯU Ý: Thuộc tính 1 là định danh lớp (1-3)

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

class 1: 59

class 2: 71

class 3: 48

Missing Attribute Values:

Không có

Instruction list:

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

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

```
vab=c('Alcohol','Malic acid','Ash','Alcalinity of ash','Magnesium','Total  
phenols','Flavanoids','Nonflavanoid phenols','Proanthocyanins','Color  
intensity','Hue','OD280/OD315 of diluted wines','Proline','Class')
```

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

```
write.csv(mydata,'tidy_wine.csv', row.names = FALSE)
```

--- HẾT ---