**Họ tên: Ngô Đức Vũ**

**MSSV:20520950**

**BÁO CÁO THỰC HÀNH TUẦN 2**

Bộ dữ liệu: Wines Data Set

Nguồn: <https://archive.ics.uci.edu/ml/datasets/Wine>

Codebook: Được mô tả trong file wine.pdf

**Bảng CODE Book mô tả bộ dữ liệu:**

|  |  |
| --- | --- |
| Thông tin | Nội dung |
| Tên bộ dữ liệu | Wine recognition data |
| Nguồn thu thập và cách thu thập dữ liệu | Cách thu thập: Những dữ liệu này là kết quả phân tích hóa học của các loại rượu vang được trồng ở cùng một vùng ở Ý nhưng có nguồn gốc từ ba giống rượu khác nhau. Phân tích xác định số lượng của 13 thành phần được tìm thấy trong mỗi loại trong số ba loại rượu vang. |
| Số thuộc tính | 14 |
| Thông tin và tên các thuộc tính | Các thuộc tính được liệt kê theo danh sách bên dưới:  1)Type: Loại (1, 2, 3)  2) Alcohol: Nồng độ cồn (numeric)  3) Malic acid: Nồng độ axit malic (numeric)  4) Ash: tro (numeric)  5) Alcalinity of ash: Độ kiềm của tro (numeric)  6) Magnesium: Độ magie (numeric)  7) Total phenols: Độ cồn (numeric)  8) Flavanoids: (numeric)  9) Nonflavanoid phenols: (numeric)  10) Proanthocyanins: (numeric)  11)Color intensity: Cường độ màu (numeric)  12)Hue: (numeric)  13)OD280/OD315 of diluted wines: (numeric)  14)Proline: (numeric) |
| Thông tin tác giả: | (a) Tác giả: 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.  (b) Nhà tài trợ: Stefan Aeberhard, email: stefan@coral.cs.jcu.edu.au  (c) Thời gian: Tháng 7 năm 1991  (d) Được cập nhật ngày 21 tháng 9 năm 1998 bởi C.Blake : Thêm thông tin thuộc tính |

* Raw data: wine.data
* Tidy data: lưu thành file wine.csv
* Instruction list: wine.R

rm(list=ls())

#Đọc và lưu file dữ liệu vào biến wine, ngăn cách bởi dấu phẩy “,”

wine <- read.csv("wine/wine.data", sep=",")

##Tạo biến chứa header

variables <-  c('Type', 'Alcohol', 'Malic', 'Ash',

                'Alcalinity', 'Magnesium', 'Phenols', 'Flavanoids',

                'Nonflavanoids','Proanthocyanins', 'Color', 'Hue',

                'Dilution', 'Proline')

#Nối header với biến dữ liệu

colnames(wine) <- variables

#Xuất dữ liệu ra file iris.csv

write.csv(wine,"wine.csv",row.names = FALSE)