

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 BANK MARKETING

Họ và tên : Lưu Quang Tiến Hoàng

MSSV : 20521342

Lớp : DS103.M21

Yêu cầu trích dẫn

**Bộ dữ liệu này được công bố để nghiên cứu. Các chi tiết được mô tả ở [Moro et al., 2011].
Hãy ghi chú này nếu bạn định sử dụng cơ sở dữ liệu này:**

[Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology*.

In P. Novais et al. (Eds.), *Proceedings of the European Simulation and Modelling Conference - ESM'2011*, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.

Available at: [pdf] <http://hdl.handle.net/1822/14838>

[bib] <http://www3.dsi.uminho.pt/pcortez/bib/2011-esm-1.txt>.

Tên bộ dữ liệu:

Bank Marketing

Nguồn thu thập và cách thu thập:

Tạo bởi: Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) @ 2012.

Thông tin liên quan:

Dữ liệu có liên quan đến các chiến dịch tiếp thị trực tiếp của một tổ chức ngân hàng Bồ Đào Nha.

Các chiến dịch tiếp thị dựa trên các cuộc gọi điện thoại. Thông thường, cần có nhiều hơn một liên hệ với cùng một khách hàng, để truy cập xem sản phẩm (tiền gửi có kỳ hạn ngân hàng) sẽ được (hoặc không) đăng ký.

Về bộ dữ liệu bộ dữ liệu:

bank-full.csv với tất cả các ví dụ, được sắp xếp theo ngày (từ tháng 5 năm 2008 đến tháng 11 năm 2010).

Mục tiêu phân loại là để dự đoán xem liệu khách hàng có đăng ký một khoản tiền gửi có kỳ hạn hay không (biến y).

Kích thước dữ liệu:

File raw data:

Số thuộc tính: 17

Kích thước: 4503KB

File tidy data:

Số thuộc tính: 16 + thuộc tính đầu ra

Tidydata được lưu thành file: tidy_bank-full.csv

Kích thước: 4547KB

Số dòng dữ liệu: 45211

Thông tin các thuộc tính:

Để biết thêm thông tin, hãy đọc [Moro et al., 2011].

Biến đầu vào:

dữ liệu khách hàng ngân hàng:

age (numeric): tuổi

job: công việc.

Phân loại:

"admin": quản trị viên

"unknown": không xác định

"unemployed": thất nghiệp

"management": quản lý

"housemaid": người giúp việc

"entrepreneur": doanh nhân

"student": học sinh, sinh viên

"blue-collar": nhân viên kinh doanh

"self-employed": tự kinh doanh

"retired": đã nghỉ hưu

"technician": nhân viên kỹ thuật

"services": dịch vụ

marital: tình trạng hôn nhân

Phân loại:

"married": đã kết hôn

"divorced": đã ly hôn (nghĩa là đã ly hôn hoặc góa phụ/quả phụ)

"single": độc thân

education: Trình độ học vấn

Phân loại:

"unknown": không xác định

"secondary": tốt nghiệp cấp 2

"primary": tốt nghiệp cấp 3

"tertiary": tốt nghiệp đại học

default: có tín dụng trong tình trạng vỡ nợ không? ("yes", "no")

balance (numeric): Số dư trung bình hằng năm, theo đơn vị Euro housing: Có khoản vay nhà đất không? ("yes", "no")

loan: Có khoản vay cá nhân không? ("yes", "no")

contact: cách thức liên lạc

Phân loại:

"unknown": không xác định

"telephone": số điện thoại để bàn

"cellular": số điện thoại di động

day (numeric): Ngày liên hệ cuối cùng trong tháng

month: tháng liên hệ cuối cùng trong năm

Phân loại:

"jan": tháng 1

"feb": tháng 2

"mar": tháng 3

...

"nov": tháng 11

"dec": tháng 12

duration (numeric): thời gian liên hệ cuối cùng, tính theo giây

Các thuộc tính khác:

campaign: số lượng liên hệ được thực hiện trong chiến dịch này và cho khách hàng này (, bao gồm liên hệ cuối cùng)

pdays: số ngày trôi qua sau khi khách hàng được liên lạc lần cuối từ một chiến dịch trước đó (numeric)

“-1”: có nghĩa là khách hàng trước đây không được liên lạc previous: số lượng liên hệ được thực hiện trước chiến dịch này và cho khách hàng này (numeric)

poutcome: kết quả của chiến dịch tiếp thị trước đó

Phân loại:

"unknown": Không xác định

"other": Kết quả khác

"failure": Thất bại

"success": Thành công

Biến đầu ra (mục tiêu mong muốn):

– y – Khách hàng đã đăng ký tiền gửi có kỳ hạn chưa? ("yes", "no")

Missing Attribute Values:

Không có

Instruction list:

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

```
data=read.csv('dataset/bank-full.csv', sep=';')
```

```
write.csv(data, 'tidy_bank-full.csv', row.names = FALSE)
```

--- HẾT ---

