**88.Bank Marketing**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  | 45211 |

|  |  |
| --- | --- |
| **Abstract**: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). |  |

**Source:**

[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, Guimaraes, Portugal, October, 2011. EUROSIS (<http://hdl.handle.net/1822/14838>)

**Data Set Information:**

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be (or not) subscribed.   
  
There are two datasets:   
1) bank-full.csv with all examples, ordered by date (from May 2008 to November 2010).   
2) bank.csv with 10% of the examples (4521), randomly selected from bank-full.csv.   
The smallest dataset is provided to test more computationally demanding machine learning algorithms (e.g. SVM).   
  
The classification goal is to predict if the client will subscribe a term deposit (variable y).

**Attribute Information:**

For more information, read [Moro et al., 2011].  
  
Input variables:  
# bank client data:  
1 - age (numeric)  
2 - job : type of job (categorical: 'admin.','unknown','unemployed','management','housemaid','entrepreneur','student',  
'blue-collar','self-employed','retired','technician','services')   
3 - marital : marital status (categorical: 'married','divorced','single'; note: 'divorced' means divorced or widowed)  
4 - education (categorical: 'unknown','secondary','primary','tertiary')  
5 - default: has credit in default? (binary: 'yes','no')  
6 - balance: average yearly balance, in euros (numeric)   
7 - housing: has housing loan? (binary: 'yes','no')  
8 - loan: has personal loan? (binary: 'yes','no')  
# related with the last contact of the current campaign:  
9 - contact: contact communication type (categorical: 'unknown','telephone','cellular')   
10 - day: last contact day of the month (numeric)  
11 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')  
12 - duration: last contact duration, in seconds (numeric)  
# other attributes:  
13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)  
14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)  
15 - previous: number of contacts performed before this campaign and for this client (numeric)  
16 - poutcome: outcome of the previous marketing campaign (categorical: 'unknown','other','failure','success')  
  
Output variable (desired target):  
17 - y - has the client subscribed a term deposit? (binary: 'yes','no')

**1. Title: Bank Marketing**

**2. Sources**

**Created by: Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) @ 2012**

**5. Number of Instances: 45211 for bank-full.csv (4521 for bank.csv)**

**6. Number of Attributes: 16 + output attribute.**

**7. Attribute information:**

**For more information, read [Moro et al., 2011].**

**Input variables:**

**# bank client data:**

**1 - age (numeric)**

**2 - job : type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur","student",**

**"blue-collar","self-employed","retired","technician","services")**

**3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means divorced or widowed)**

**4 - education (categorical: "unknown","secondary","primary","tertiary")**

**5 - default: has credit in default? (binary: "yes","no")**

**6 - balance: average yearly balance, in euros (numeric)**

**7 - housing: has housing loan? (binary: "yes","no")**

**8 - loan: has personal loan? (binary: "yes","no")**

**# related with the last contact of the current campaign:**

**9 - contact: contact communication type (categorical: "unknown","telephone","cellular")**

**10 - day: last contact day of the month (numeric)**

**11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")**

**12 - duration: last contact duration, in seconds (numeric)**

**# other attributes:**

**13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)**

**14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)**

**15 - previous: number of contacts performed before this campaign and for this client (numeric)**

**16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")**

**Output variable (desired target):**

**17 - y - has the client subscribed a term deposit? (binary: "yes","no")**

**8. Missing Attribute Values: None**