

Projekt zaliczeniowy z przedmiotu Data Mining (mgr sb-ndz)

1. Opis projektu

Dane dotyczą bezpośrednich kampanii marketingowych portugalskiej instytucji bankowej. Kampanie marketingowe opierały się na rozmowach telefonicznych. Często wymagany był więcej niż jeden kontakt z tym samym klientem, aby otrzymać odpowiedź, czy produkt (lokata bankowa) byłby („tak”), czy nie („nie”) subskrybowany.

2. Tytuł

Bank Marketing (w kontekście społeczno-ekonomicznym)

3. Źródło

Stworzony przez: Sérgio Moro (ISCTE-IUL), Paulo Cortez (Univ. Minho) and Paulo Rita (ISCTE-IUL) @ 2014

Link do pliku: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

4. Cel projektu

Celem projektu jest stworzenie modelu predykcyjnego, czy klient zapisze się na lokatę bankową (zmienna y).

5. Liczba obserwacji: 41188

6. Liczba zmiennych: 20 + zmienna celu

7. Informacja o zmiennych

Dane klientów banku:

1 - age (numeric)

2 - job : type of job (categorical: "admin.", "blue-collar", "entrepreneur", "housemaid", "management", "retired", "self-employed", "services", "student", "technician", "unemployed", "unknown")

3 - marital : marital status (categorical: "divorced", "married", "single", "unknown")

4 - education (categorical: "basic.4y", "basic.6y", "basic.9y", "high.school", "illiterate", "professional.course", "university.degree", "unknown")

5 - default: has credit in default? (categorical: "no", "yes", "unknown")

6 - housing: has housing loan? (categorical: "no", "yes", "unknown")

7 - loan: has personal loan? (categorical: "no", "yes", "unknown")

related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: "cellular", "telephone")

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

10 - day_of_week: last contact day of the week (categorical: "mon", "tue", "wed", "thu", "fri")

11 - duration: last contact duration, in seconds (numeric).

Inne zmienne:

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

13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

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

15 - poutcome: outcome of the previous marketing campaign (categorical: "failure", "nonexistent", "success")

Zmienne kontekstu społecznego i gospodarczego:

16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)

17 - cons.price.idx: consumer price index - monthly indicator (numeric)

18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)

19 - euribor3m: euribor 3 month rate - daily indicator (numeric)

20 - nr.employed: number of employees - quarterly indicator (numeric)

Zmienna celu:

21 - y - czy klient zapisał lokatę terminową? (binary: "yes", "no")

8. Brakujące wartości

Występują braki danych oznaczone za pomocą etykiety „unknown”.

Wykonują:

Andrzej Kiewlak

Uładzislau Darhevich

Krzysztof Wolański