4523

Assignment: Identify a dataset

Team: 404 Brain Not Found

Dataset Description:

This dataset consists of direct marketing campaigns by a Portuguese banking institution using phone calls. The campaigns aimed to sell subscriptions to a bank term deposit (see variable y). It contains customers' geodemographics and behavioral data (Appendix.1).

Data Shape: (41188, 21)

Dataset Link:

https://drive.google.com/file/d/1XwPmv2pzMPo-t8yzG_Q4NkSp-ZHV1wLU/view?usp=sharing

Research Goal:

Through customer demographics and bank attributions to conduct an exploratory data analysis, and finally predict whether the customer will subscribe to the term deposit, and come up with a marketing segmentation strategy, as well as targeting specific segments through cluster analysis.

- Goal I: Exploratory
    - Find the underlying relationships and correlations between features and the outcome - data preprocessing & visualization
- Goal II: Descriptive Analysis
    - Find the attribute that has the strongest effects on subscription decision - logistics regression, decision tree
- Goal III: Application
    - Help the company to set a classifier to target specific customers to make direct campaigns. Under such methodology, we could help the company to optimize the outcome - clustering analysis, factor analysis

Citation

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank
Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

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.

Appendix

Data Dictionary

Column	Variable	Class
age	age of customer	
job	type of job	categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown"
marital	marital status	categorical: "divorced","married","single","unknown"; note: "divorced" means divorced or widowed
education	highest degree of customer	categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","unknown"
default	has credit in default?	categorical: "no","yes","unknown"
housing	has housing loan?	categorical: "no","yes","unknown"
loan	has personal loan?	categorical: "no","yes","unknown"
contact	contact communication type	categorical: "cellular","telephone"
month	last contact month of year	categorical: "jan", "feb", "mar", ..., "nov", "dec"
day_of_week	last contact day of the week	categorical: "mon","tue","wed","thu","fri"
campaign	number of contacts performed during this campaign and for this client	numeric, includes last contact
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
previous	number of contacts performed before this campaign and for this client	numeric
poutcome	outcome of the previous marketing campaign	categorical: "failure","nonexistent","success"
emp.var.rate	employment variation rate - quarterly indicator	numeric
cons.price.idx	consumer price index - monthly indicator	numeric
cons.conf.idx	consumer confidence index - monthly indicator	numeric
euribor3m	euribor 3 month rate - daily indicator	numeric
nr.employed	number of employees - quarterly indicator	numeric
y	has the client subscribed a term deposit?	binary: "yes","no"