Default of Credit Card Clients

Donated on 1/25/2016

This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods.

Dataset Characteristics
Multivariate

Subject Area
Business

Associated Tasks
Classification

Feature Type
Integer, Real

Instances
30000

Features
23

Dataset Information

Additional Information

This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients. Because the real probability of default is unknown, this study presented the novel Sorting Smoothing Method to estimate the real probability of default. With the real probability of default as the response variable (Y), and the predictive probability of default as the independent variable (X), the simple linear regression result (Y = A + BX) shows that the forecasting model produced by artificial neural network has the highest coefficient of determination; its regression intercept (A) is close to zero, and regression coefficient (B) to one. Therefore, among the six data mining techniques, artificial neural network is the only one that can accurately estimate the real probability of default.

SHOW LESS

Has Missing Values?

No

Introductory Paper

[The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients]

By I. Yeh, Che-hui Lien. 2009
Published in Expert systems with applications

Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
ID	ID	Integer				no
X1	Feature	Integer		LIMIT_BAL		no
X2	Feature	Integer	Sex	SEX		no
X3	Feature	Integer	Education Level	EDUCATION		no
X4	Feature	Integer	Marital Status	MARRIAGE		no
X5	Feature	Integer	Age	AGE		no
X6	Feature	Integer		PAY_0		no
X7	Feature	Integer		PAY_2		no
X8	Feature	Integer		PAY_3		no
X9	Feature	Integer		PAY_4		no
X10	Feature	Integer		PAY_5		no
X11	Feature	Integer		PAY_6		no
X12	Feature	Integer		BILL_AMT1		no

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
X13	Feature	Integer		BILL_AMT2		no
X14	Feature	Integer		BILL_AMT3		no
X15	Feature	Integer		BILL_AMT4		no
X16	Feature	Integer		BILL_AMT5		no
X17	Feature	Integer		BILL_AMT6		no
X18	Feature	Integer		PAY_AMT1		no
X19	Feature	Integer		PAY_AMT2		no
X20	Feature	Integer		PAY_AMT3		no
X21	Feature	Integer		PAY_AMT4		no
X22	Feature	Integer		PAY_AMT5		no
X23	Feature	Integer		PAY_AMT6		no
Y	Target	Binary		default payment next month		no

Rows per page 25 0 to 25 of 25

Additional Variable Information

This research employed a binary variable, default payment (Yes = 1, No = 0), as the response variable. This study reviewed the literature and used the following 23 variables as explanatory variables:
X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
X2: Gender (1 = male; 2 = female).
X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
X4: Marital status (1 = married; 2 = single; 3 = others).

X5: Age (year).
X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .;X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.
X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .;X23 = amount paid in April, 2005.

SHOW LESS

Dataset Files

File	Size
default of credit card clients.xls	5.3 MB

Papers Citing this Dataset

SORT BY YEAR, DESC

Credit Default Mining Using Combined Machine Learning and Heuristic Approach
By Sheikh Islam, William Eberle, Sheikh Ghafoor. 2018
Published in ArXiv.

Dancing in the Dark: Private Multi-Party Machine Learning in an Untrusted Setting
By Clement Fung, Jamie Koerner, Stewart Grant, Ivan Beschastnikh. 2018
Published in ArXiv.

Rows per page 5 0 to 3 of 3

Reviews

There are no reviews for this dataset yet.

LOGIN TO WRITE A REVIEW

DOWNLOAD (5.3 MB)

IMPORT IN PYTHON

CITE

Creators

I-Cheng Yeh

DOI

10.24432/C55S3H

License

THE PROJECT

About Us

CML

National Science Foundation

NAVIGATION

Home

View Datasets

Donate a Dataset

LOGISTICS

Contact

Privacy Notice

Feature Request or Bug Report