

Predictive Modeling on Bank Marketing for Term Deposit

Group Member

Menglan Jiang menglan.jiang@duke.edu
Xinyaun Zhang xy.zhang@duke.edu
Mengrun Li mengrun.li@duke.edu
Yaqian Cheng yaqian.cheng@duke.edu

October 21, 2015

Description

In the paper, we aim to explore the marketing strategy for attracting term deposit from potential clients. The major task is to identify the targeting customers who have the desire to invest in term deposit. Based on the dataset, We have 45211 records each containing the information of a single client. The information varies from demographic information, financial status to credit history. In specific, we have 17 variables of interest, including age, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, poutcome.

The theoretical foundation for our project lies in all kinds of classification and regression methods in statistics. In specific, the candidate methods are Support Vector Machine (SVM), Least Absolute Selection and Shrinkage Operator (LASSO), logistic regression and etc. In classification methods, we will use the SVM to categorize the potential clients into two groups. One who will invest in term deposit, the other will not. In the regression part, we are considering LASSO to pick up significant variables and then do a logistic regression to evaluate the possibility of investing in term deposit. In the end, We will use the R language for majority of the programming.

Appendix

Index	Variable	Description
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; note: divorced means divorced or widowed)
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	balance	number of balance. (numeric)
7	housing	has housing loan? (categorical: no, yes, unknown)
8	loan	has personal loan? (categorical: no, yes, unknown)
9	contact	contact communication type (categorical: cellular, telephone)
10	month	last contact month of year (categorical: jan, feb, mar,..., nov, dec)
11	day of week	last contact day of the week (categorical: mon, tue, wed, thu, fri)
12	duration	last contact duration, in seconds (numeric).
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; 999 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: failure, nonexistent, success)
17	y	has the client subscribed a term deposit? (binary: yes, no)