

Customer Quality Prediction

Problem

Suppose you are working for ProcessMiner as a Data Scientist. ProcessMiner is commissioned by an insurance company to develop a tool to optimize their marketing efforts. Your objective is to determine which set of customers the marketing firm should contact to maximize profit.

1. The cost of marketing to a particular customer is \$30. This cost is paid regardless of whether the customer responds to our marketing or not.
2. Only if a customer responds to our marketing, do we earn a profit.
3. Profit does NOT include the marketing cost.
4. Total Profit = Average profit per responding Customer * Number of customers responding – Number of customers to whom you marketed * \$30.

Data

The insurance company has provided you with a historical data set (training.csv). The company has also provided you with a list of potential customers to whom to market (testingCandidate.csv). From this list of potential customers, you need to determine yes/no whether you wish to market to them.

Type	Name	Description
Input Variables	custAge	The age of the customer (in years)
Input Variables	profession	Type of job
Input Variables	marital	Marital status
Input Variables	schooling	Education level
Input Variables	default	Has a previous defaulted account?
Input Variables	housing	Has a housing loan?
Input Variables	loan	Has a personal loan?
Input Variables	contact	Preferred contact type
Input Variables	month	Last contact month
Input Variables	day_of_week	Last contact day of the week
Input Variables	campaign	Number of times the customer was contacted
Input Variables	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)
Input Variables	previous	Number of contacts performed before this campaign and for this client
Input Variables	poutcome	Outcome of the previous marketing campaign
Input Variables	emp.var.rate	Employment variation rate - quarterly indicator
Input Variables	cons.price.idx	Consumer price index - monthly indicator
Input Variables	cons.conf.idx	Consumer confidence index - monthly indicator
Input Variables	euribor3m	Euribor 3 month rate - daily indicator
Input Variables	nr.employed	Number of employees - quarterly indicator
Input Variables	pmonths	Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
Input Variables	pastEmail	Number of previous emails sent to this client
Target Variables	responded	Did the customer respond to the marketing campaign and purchase a policy?
Target Variables	profit	If the customer purchased a policy, how much profit (before marketing costs) did the company make on the policy?

Instructions

Please email back the following files:

1. All associated code files you used to complete your analysis.
2. Please add a column to the testingCandidate.csv file. In this column, for each observation indicate a 1 (yes) or a 0 (no) whether you wish to market to that candidate.
3. Please prepare a presentation to show to ProcessMiner employees onsite. The presentation should be written as if you were presenting your results to a non-technical audience at the client.
4. R or Python is preferred for the codes.
5. If there is any question, email at vnajari@processminer.com.

Your submission will be evaluated based upon:

1. Business understanding

2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Code quality and reusability
7. Presentation

Submission

- Submit all materials within a week in a zipped file.



training.csv



testingCa...idate.csv

Google drive link for dataset: <https://drive.google.com/open?id=0BzrlKrqHScyETThQQjZOcFN0ejQ>