

Customer targeting

Measuring the propensity to purchase
Creating and interpreting the gain chart

Ricco RAKOTOMALALA

Customer targeting process

Promoting a new product to customers

Goal: Promoting a new product

Direct marketing: seek the most receptive customers (responders, buyers)

- the budget is limited
- do not solicit the hostile customers

Tools:

- customer database
- a target variable which specifies the buyers (positive individuals, +) and the non-buyers (negative, -). *we do not dispose to this variable initially.*
- learning method which enables to assign a score (a probability to be positive, a propensity to purchase) to the individuals
- applying the score to the database - sorting the individuals according to their propensity
- soliciting actually the customers with high propensity
- 2 evaluation criteria (*the baseline is to select at random the individuals*)
 - the rate of return (proportion of + among the individuals targeted)
 - the recall (proportion of + recovered), market share

Note: the approach can be applied to any domains where we want to target a subset of the population (screening campaign in medicine, etc.)

Targeting process

Overall outline

2,000 customers solicited from a **test mailing** (random sample)
 100 customers have responded positively = 100/2,000 → **5%**
 (baseline rate of return)

Title	Insuran	Child	Wages
Mrs	No	2	1408
Mr	No	2	1294
Mrs	No	1	1810
Mrs	Yes	0	1800
Mr	No	5	1770
Mr	No	1	1550
Mrs	Yes	2	1561
Mrs	Yes	2	1561
Mrs	No	1	1660
Mrs	No	2	1408
Mrs	Yes	1	1402
Mrs	No	0	862
Mr	Yes	1	1914
Mrs	No	2	2324
Mrs	No	2	862
Mrs	No	0	892
Mr	No	1	2214
Mrs	No	1	2021
Mr	No	1	1425
Mrs	No	0	1863
Mrs	No	0	1318
Mr	Yes	1	1800
Mrs	No	1	981
Mrs	No	2	2900
Mr	No	0	5400

200,000 customers

Customer database
 (202,000 customers)

Title	Insuran	Child	Wages	Retour
Mrs	No	2	1408	+
Mr	No	2	1294	+
Mrs	No	1	1810	-
Mrs	Yes	0	1800	+
Mr	No	5	1770	+
Mr	No	1	1550	-
Mrs	Yes	2	1561	+

1,000 Test sample

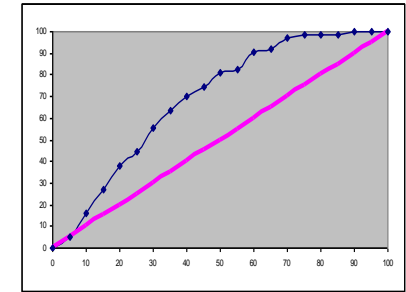
1,000 Train sample

Title	Insuran	Child	Wages	SCORE
Mr	No	0	2185	0.9997
Mrs	No	1	900	0.9992
Mrs	No	2	3000	0.9987
Mr	No	1	1410	0.9976
Mrs	No	2	1600	0.9956
Mrs	No	0	1520	0.9931
Mr	No	0	5400	0.9898
Mrs	No	2	2400	0.9888
Mrs	Yes	3	1237	0.987
Mr	No	2	1572	0.9863
Mrs	No	1	2621	0.9861
Mrs	No	2	1782	0.9855
Mr	No	0	2400	0.9841
Mrs	No	2	1020	0.9836
Mrs	No	0	1812	0.9828
Mrs	No	0	1470	0.9821
Mrs	No	2	1320	0.9799
Mrs	No	1	1080	0.9788

Potential of buyers (+) : **5%** of 200,000 = 10,000 positive customers

Gain chart

Evaluating the performance of the targeting



$$S(R) = \Phi(X)$$

Score function: a binary classifier which enables to assign a score to the individuals

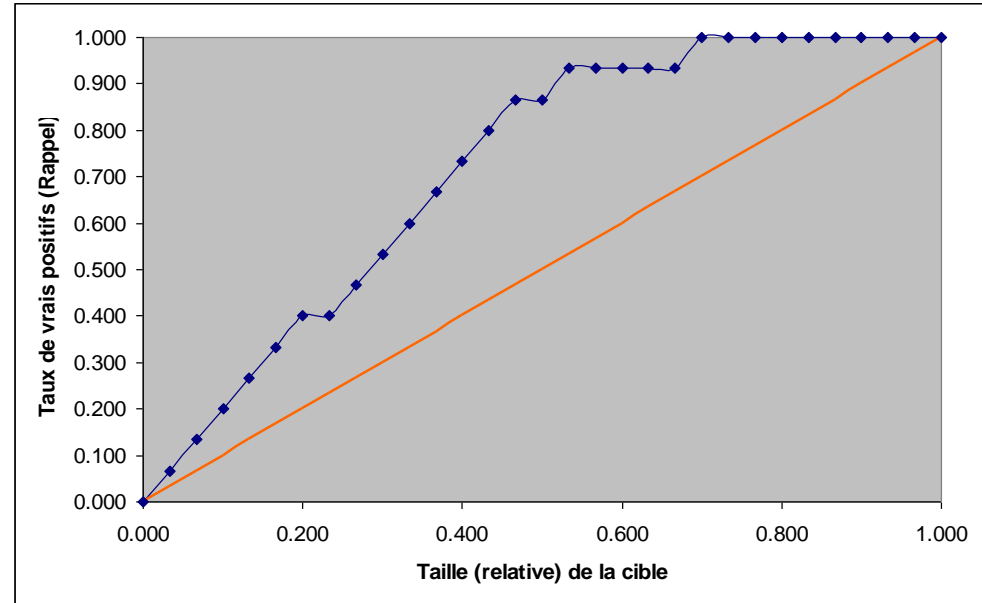
- (1) Applying the score function to the database
- (2) Sorting according to the score
- (3) Targeting the individuals with high score
- (4) Evaluating the performance (**expected buyers** for a number of solicited customers) with the **Gain Chart**

Targeting process

How to build the "Gain chart" (says also "Cumulative lift curve") from a labeled sample?

Responders (+ or -) Sorting in descending order according to the score ("Score" is often the estimation of the probability to be positive. But, it may be any value which reflects the propensity to be positive.)

i	Retour	Score	Taille Cible	Rappel (TVP)
			0.000	0.000
1	positif	1.000	0.033	0.067
2	positif	1.000	0.067	0.133
3	positif	0.999	0.100	0.200
4	positif	0.999	0.133	0.267
5	positif	0.998	0.167	0.333
6	positif	0.992	0.200	0.400
7	néгатif	0.987	0.233	0.400
8	positif	0.987	0.267	0.467
9	positif	0.974	0.300	0.533
10	positif	0.969	0.333	0.600
11	positif	0.953	0.367	0.667
12	positif	0.952	0.400	0.733
13	positif	0.942	0.433	0.800
14	positif	0.825	0.467	0.867
15	néгатif	0.772	0.500	0.867
16	positif	0.590	0.533	0.933
17	néгатif	0.507	0.567	0.933
18	néгатif	0.307	0.600	0.933
19	néгатif	0.294	0.633	0.933
20	néгатif	0.109	0.667	0.933
21	positif	0.073	0.700	1.000
22	néгатif	0.035	0.733	1.000
23	néгатif	0.024	0.767	1.000
24	néгатif	0.016	0.800	1.000
25	néгатif	0.015	0.833	1.000
26	néгатif	0.009	0.867	1.000
27	néгатif	0.004	0.900	1.000
28	néгатif	0.003	0.933	1.000
29	néгатif	0.002	0.967	1.000
30	néгатif	0.000	1.000	1.000



!

N	30
N(positif)	15

Relative cumulative number of cases = i / N

TPR (true positive rate) = $N(+ \text{ among the "i" first cases}) / N(+)$

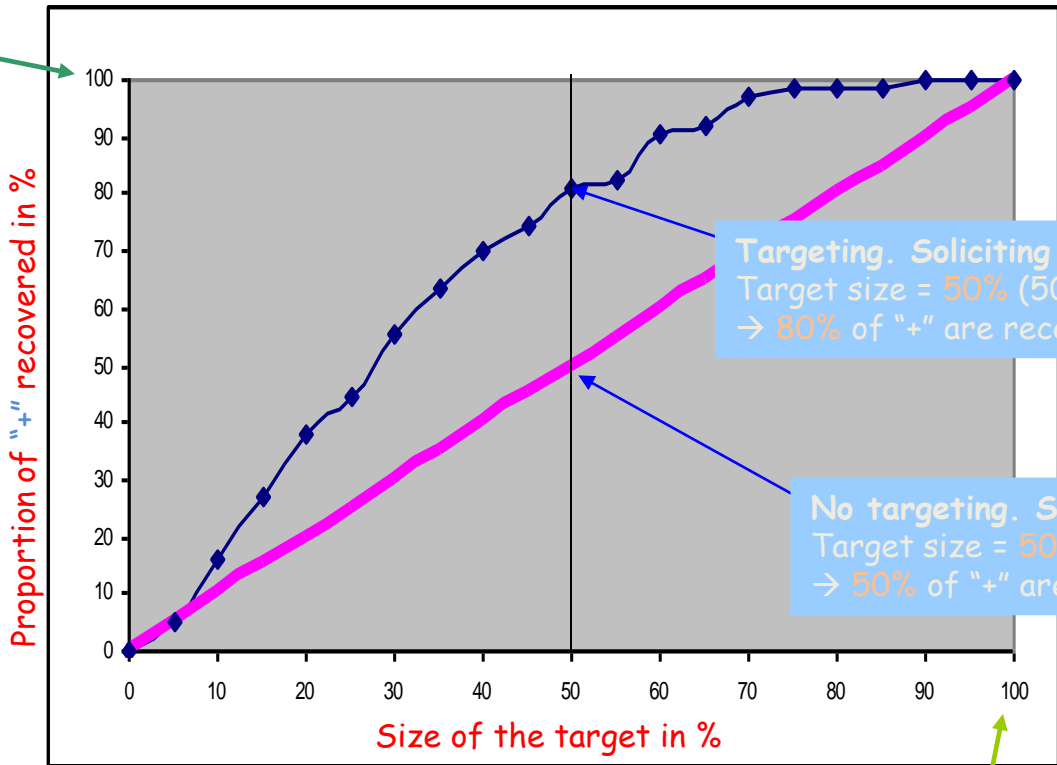
Targeting process

How to interpret the Gain chart on the test sample?

1,000 cases in the test sample
50 (5%) are positive

The dataset is sorted in descending order according to the score.

100 % of "+" = 50 cases



Targeting. Soliciting in priority the cases with high score
Target size = 50% (500 first cases of the sample)
→ 80% of "+" are recovered (40 cases "+")

No targeting. Select cases at random.
Target size = 50% (500 cases of the sample)
→ 50% of "+" are recovered (25 cases "+")

100 % of the target = 1,000 cases



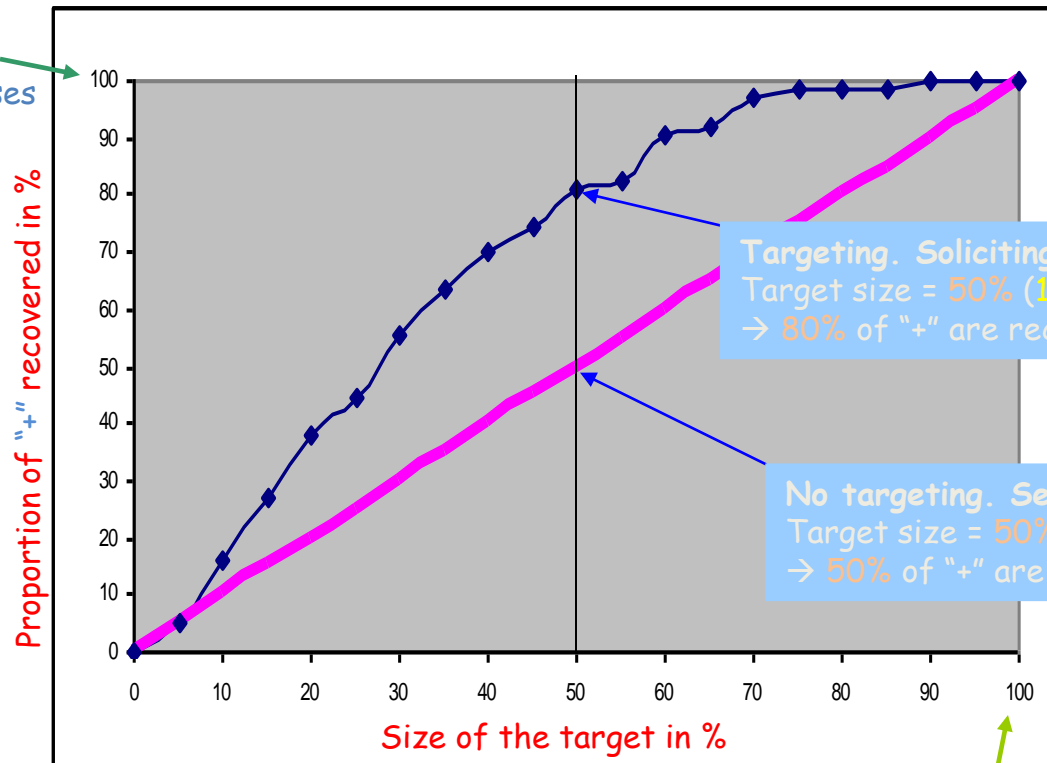
Targeting process

How to transpose the reading of the gain chart on the customer database?

200,000 cases in the customer database
We do not know who are positive
But we expect that ~5% are positive i.e. ~10,000 cases

The dataset is sorted in descending order according to the score.

100 % of "+"
= 10,000 cases



Targeting. Soliciting in priority the cases with high score
Target size = 50% (100,000 first cases of the database)
→ 80% of "+" are recovered (8,000 cases "+")

No targeting. Select cases at random.
Target size = 50% (100,000 cases of the database)
→ 50% of "+" are recovered (5,000 cases "+")

100 % of the target = 200,000 cases

Targeting process

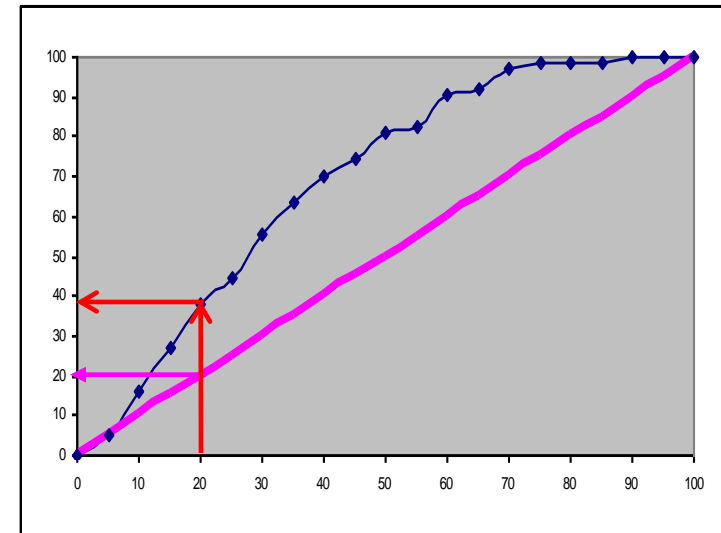
By fixing the target size (costs), how many positive instances (benefit) will be obtained?

We specify the budget of the campaign
e.g. 40,000 prospects

!
 We found 1,800
additional buyers

38% of "+" are recovered
i.e. $0.38 \times 10,000 = 3,800$ "+"

At random, 20% of "+" recovered
i.e. $0.20 \times 10,000 = 2,000$ "+"



Budget: 40,000 mailing (20% of the database)

Conclusion:

Rate of return: $3,800 / 40,000 = 9,5\% \rightarrow 5\%$ if we select the customers at random

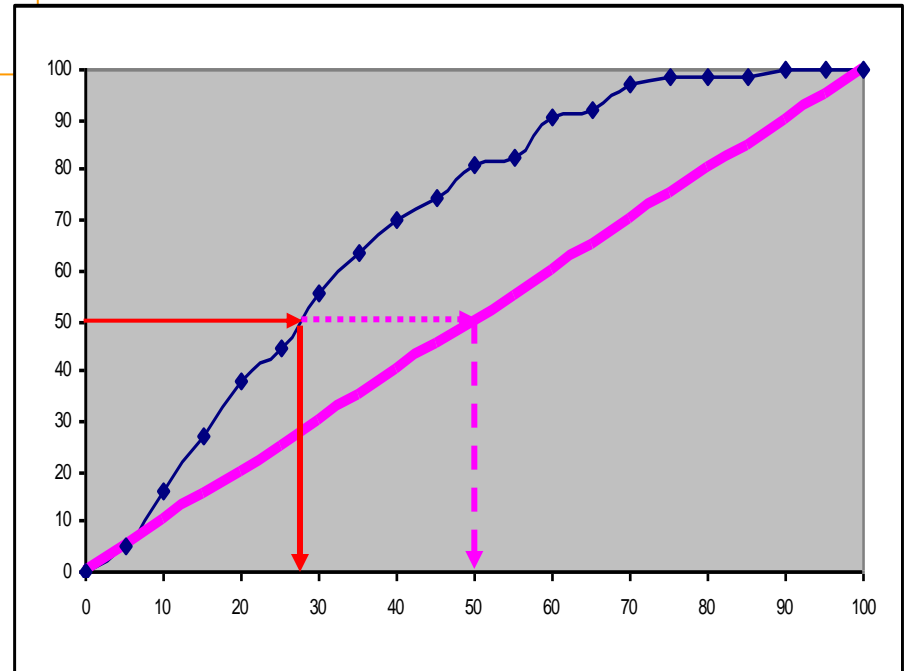
Market share: $3,800 / 10,000 = 38\% \rightarrow$ it remains 6,200 unsolicited buyers

Targeting process

By fixing the objective, how many customers must be solicited?

We specify the number of buyers we must obtain
e.g. 5,000 buyers

5,000 buyers
i.e. 50% of potential buyers = 5,000 / 10,000



We must send mails to 27% of the customers with the higher scores
i.e. $0.27 \times 200,000 = 54,000$ individuals

At random, we must send 100,000 mails to obtain this objective

We save 46,000 mails

Conclusion:

Rate of return : $5,000 / 54,000 = 9,25\% \rightarrow 5\%$ if we select the customers at random

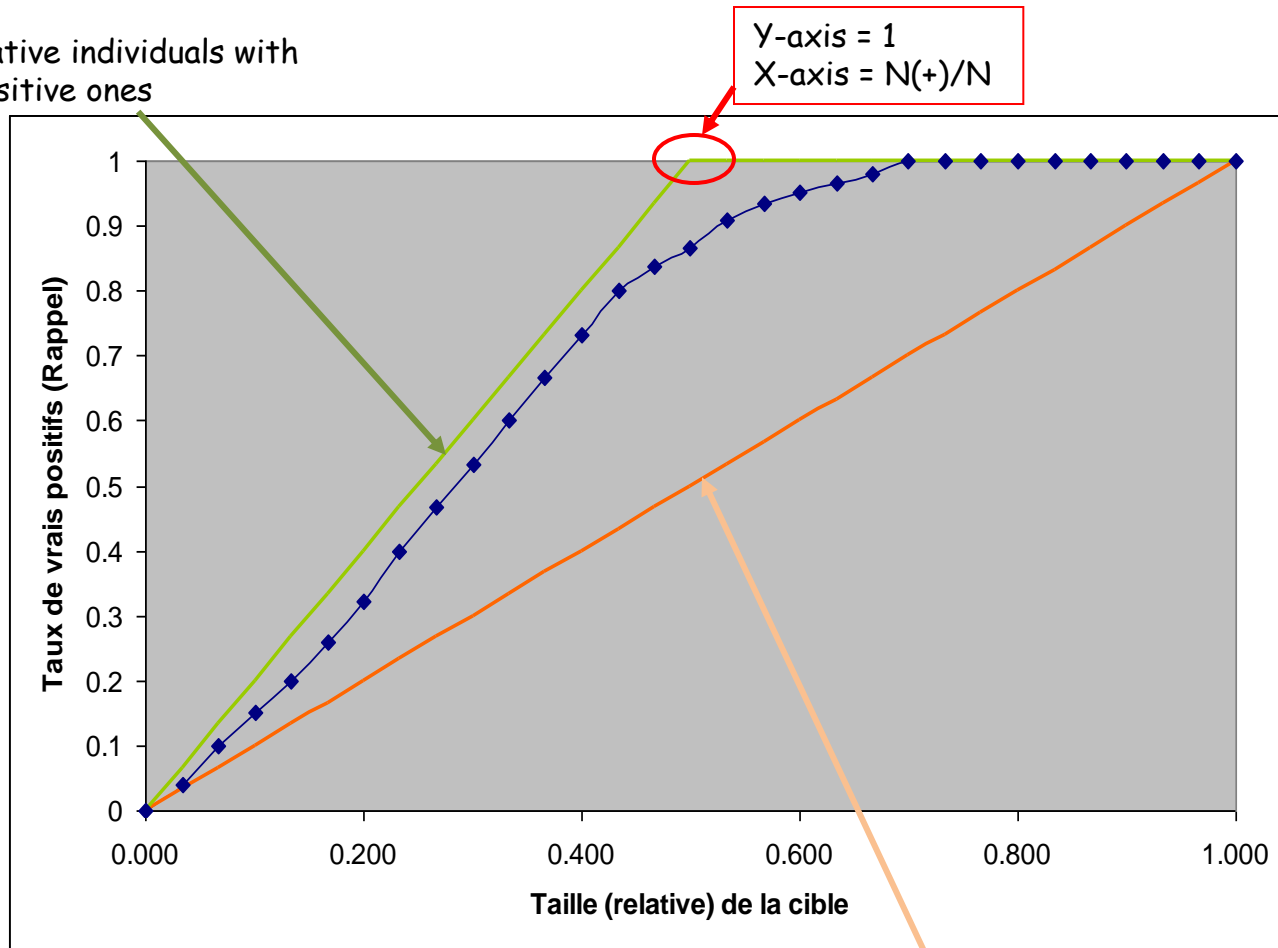
Market share: $5,000 / 10,000 = 50\%$ \rightarrow this is a given in this context

Conclusion

No targeting (selecting cases at random) and perfect targeting (all the positives have higher score than the negatives)

Perfect targeting

i.e. there are no negative individuals with higher score than positive ones



Y-axis = 1
X-axis = $N(+)/N$

Targeting at random

i.e. The score is not efficient and may be considered as a random value

References

Microsoft, “[Lift chart \(Analysis Services – Data Mining\)](#)”, SQL Server 2014.

H. Hamilton, “[Cumulative Gains and Lift Charts](#)”, in CS 831 – Knowledge Discovery in Databases, 2012.

M. Vuk, T. Curk, “[ROC Curve, Lift Chart and Calibration Plot](#)”, in Metodoloski zvezki, 3(1), 89-108, 2006.

S. Sayad, “[Model Evaluation – Classification](#)”, in Introduction to Data Mining, 2012.