prudsys® | The Realtime Analytics Company

Revenue maximisation by intelligent couponing

The scenario
Many customers only make an order in an online shop once. There are many reasons why they do not make another order. Online dealers try to counteract this using appropriate customer loyalty measures. For example, a tried-and-tested method is to hand out vouchers some time after an order to encourage the customer to make a follow-on purchase. But from an economic perspective sending vouchers to all customers is not a good solution because customers may make a second purchase without the incentive.

With this scenario in mind, a media dealer has set the task for this year's DATA MINING CUP Competition. In addition to such classical media as books and CDs, the dealer offers in its online and normal shops a variety of book types such as audio book downloads, ebooks and ebook readers. Those customers that could be encouraged to buy again with a voucher are to be identified. This scenario assumes that a follow-on purchase is desired within 90 days of the initial order.

The task
Using the existing characteristics of a customer's initial order, such as order quantity per type of goods, title and delivery weight, a decision must be made on whether to send a voucher worth € 5.00. The customers who receive a voucher should be those who would not have decided to re-order by themselves.

As shown by empirical results, it should be assumed that the voucher initiates a purchase with an average order value of € 20.00 in 10% of those not making a purchase. If a voucher is sent to a customer that would have re-purchased anyway this results in a € 5.00 loss for the dealer. The aim is therefore to maximise revenue by sending the vouchers to selected customers.

Using the data in the *dmc2010_train.txt* file a data mining model is to be generated that is then used to decide which customers from the *dmc2010_class.txt* file will receive a voucher.

The data
A detailed description of the data is found in the *dmc2010_features.pdf* file.

Form of the results provided
The participants can submit an appropriate solution model up to 31st May 2010.

The results are to be submitted in the form of a list as follows:

<Customer number>;<Voucher>

The voucher field must contain a "0" for "no voucher" or a "1" for "send voucher". Unfortunately results that do not contain all the customer numbers from the data that is to be assessed can not be graded.

This list must be sent in the form of a text file (ASCII, no Unicode) to dmc@prudsys.de. The file name should be made up of the team's name, which is the same as the team ID, and the type (text): "*<Team name>.txt*" (i.e. *TU Chemnitz_1.txt*)
The name of the relevant team has already been notified by the team leader in the registration confirmation.

Assessing the results

If a customer is sent a voucher but would have repurchased anyway, this results in a € 5.00 loss in revenue.

In contrast if a customer who would not have repurchased is sent a voucher this increases revenue on average by € 1.50.

This results in the following cost matrix:

		Real	
		Non-repurchasers (0)	Repurchasers (1)
Forecast	No voucher (0)	0	0
	Voucher (1)	1.5	-5

The number of points is therefore calculated as follows:

i ... Customer number

g_i ... Voucher decision

k_i ... Purchase

$$x_i = \begin{cases} 0, & g_i = 0 \\ -5, & g_i = 1 \wedge k_i = 1 \\ 1.5, & g_i = 1 \wedge k_i = 0 \end{cases} \qquad \text{Number of points} = \sum_{i \in \text{customer numbers}} x_i$$

The team with the highest number of points wins the DATA MINING CUP 2010. If several teams have the same number of points, the result is decided by tossing a coin. All decisions are final.