

Setting

Online customers often order in a specific online shop only once. One goal of customer relationship management (CRM) is to maximize customer lifetime value, in this case by incentivizing customers to return to the shop. A common method to do so are coupons sent to customers some time after an order. However, a coupon poses a cost of foregone profit to the retailer when it is used. In cases where a customer would have made a follow-up purchase even without the coupon incentive, the coupon value is effectively wasted. For this reason, rather than sending coupons to all customers, only specific, promising ones are targeted.

For this assignment, you are provided with real-world data by an online retailer. Your task is to identify the customers that can be expected to purchase again in the next 90 days based on customer characteristics, order conditions and ordered products. The predicted return customers will not receive a coupon. The shop estimates that sending a coupon to a customer, who does not plan to return, will convince her to place another order in 20% of cases with an average order value of 20 €. Your job is to maximize revenue by providing a list of promising customers to be targeted.

Data

You are provided with two data sets containing 37 predictive variables. Data set known also includes information about one target variable (return_customer) and should be used to build a predictive model. The target values for data set class are not provided and need to be predicted. Be aware that the data has not yet been pre-processed and will need some cleaning, so pay attention to variable types, missing values, and plausibility of values.

Model assessment

You are expected to provide a binary estimate (0/1) if the customer will return naturally within 90 days following the original purchase. The performance of your prediction model will be evaluated by the net revenue gain. In this case, costs and gains are asymmetric. Sending a coupon to a customer that would have returned anyway, i.e. a false negative, entails an effective loss of 10 e. Not sending a coupon to a customer that does not plan to return foregoes an expected profit of 3 e. The resulting cost matrix is depicted in the table below.

		True value	
		non-repurchaser (0)	repurchaser (1)
Prediction	non-repurchaser (0) / coupon	3	-10
	repurchaser (1) / no coupon	0	0

Table 1: Cost matrix for model assessment