CRM&VIP DATA SCIENTIST

TECHNICAL TEST

USE CASE DEFINITION

Business Problem

For efficiency purposes, Betsson is trying to predict which customers are going to call the Customer Service, based on their past behaviour.

Analytical Problem

For a given day, predict whether a customer will call the Customer Service in the following 14 days.

DATASET DEFINITION

The provided **train and test sets** contain one record per customer and 270 columns, out of which one column is the customer identifier (mk_CurrentCustomer), one column is the day on which the customer dataset was extracted (ScoreDate), one column is the target variable (target) and the remaining 267 columns are customer features. Follows a definition of the target variable and the various features. You are also provided with a separate file containing the customer country.

Target Name	Description
target	Equal to 1 if the customer called the Customer Service in the following 14 days, 0 otherwise.

Feature Name	Description
days_g1/2/3/4/5/6/7/8/9/10	Number of days with activity during week 1/2/3/4/5/6/7/8/9/10.
ro_g1/2/3/4/5/6/7/8/9/10	Number of game rounds during week 1/2/3/4/5/6/7/8/9/10.
to_g1/2/3/4/5/6/7/8/9/10	Total turnover amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
gw_g1/2/3/4/5/6/7/8/9/10	Total game win amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
mar_g1/2/3/4/5/6/7/8/9/10	Total margin calculated as [gw_g]/(1+[to_g]) during week 1/2/3/4/5/6/7/8/9/10.
GOC_ro_g1/2/3/4/5/6/7/8/9/10	Number of GOC game rounds during week 1/2/3/4/5/6/7/8/9/10.
GOC_to_g1/2/3/4/5/6/7/8/9/10	Total GOC turnover amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
GOC_dist_gm_g1/2/3/4/5/6/7/8/9/10	Number of GOC distinct game play days during week 1/2/3/4/5/6/7/8/9/10.

SB_ro_g1/2/3/4/5/6/7/8/9/10	Number of SB game rounds during week 1/2/3/4/5/6/7/8/9/10.
SB_to_g1/2/3/4/5/6/7/8/9/10	Total SB turnover amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
with_cnl_g1/2/3/4/5/6/7/8/9/10	Number of cancelled withdrawals during week 1/2/3/4/5/6/7/8/9/10.
with_cnt_g1/2/3/4/5/6/7/8/9/10	Number of successful withdrawals during week 1/2/3/4/5/6/7/8/9/10.
with_sum_g1/2/3/4/5/6/7/8/9/10	Total successful withdrawal amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
succ_dep_g1/2/3/4/5/6/7/8/9/10	Total successful deposit amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
unsucc_dep_g1/2/3/4/5/6/7/8/9/10	Total unsuccessful deposit amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
succ_dep_cnt_g1/2/3/4/5/6/7/8/9/10	Number of successful deposits during week 1/2/3/4/5/6/7/8/9/10.
unsucc_dep_cnt_g1/2/3/4/5/6/7/8/9/10	Number of unsuccessful deposits during week 1/2/3/4/5/6/7/8/9/10.
pm_sum_g1/2/3/4/5/6/7/8/9/10	Total number of payment methods during week 1/2/3/4/5/6/7/8/9/10.
pm_avg_g1/2/3/4/5/6/7/8/9/10	Average number of daily payment methods during week 1/2/3/4/5/6/7/8/9/10.
ini_bon_g1/2/3/4/5/6/7/8/9/10	Total bonus amount in EUR during week 1/2/3/4/5/6/7/8/9/10.
ini_bon_cnt_g1/2/3/4/5/6/7/8/9/10	Number of bonuses during week 1/2/3/4/5/6/7/8/9/10.
bon_wrt_succdep_g1/2/3/4/5/6/7/8/9/10	Total bonus amount vs total successful deposit amount calculated as

	[ini_bon_g]/(1+[succ_dep_g]) during week 1/2/3/4/5/6/7/8/9/10.
turnover_last_1/2/3/5/10/20/70days	Turnover amount in EUR in the last 1/2/3/5/10/20/70 days.
to_l1_l5/10/20/70	Turnover of the last day vs turnover in the last 5/10/20/70 days.
to_l2_l5/10/20/70	Turnover in the last 2 days vs turnover in the last 5/10/20/70 days.
to_l3_l10/20/70	Turnover in the last 3 days vs turnover in the last 10/20/70 days.
to_l5_l10/20/70	Turnover in the last 5 days vs turnover in the last 10/20/70 days.
dwcanc_last_1/2/3/5day	Total amount of cancelled withdrawals in the last 1/2/3/5 days.
w_canc_count_ratio_1/2/3/5day	Total amount of cancelled withdrawals vs number of cancelled withdrawals in the last 1/2/3/5 days.
SE_total	Number of times the customer was self excluded in the past.
GI_total	Number of times the customer was tagged as gambling issue in the past.
SE_GI_total	Number of times the customer was either self excluded or tagged as gambling issue in the past.
SE_GI_total_70days	Number of times the customer was either self excluded or tagged as gambling issue in the past 70 days.
SE_GI_wrt_days_70days	Number of times the customer was either self excluded or tagged as gambling issue in the

	past 70 days vs total active days in the past 70 days.
SE_GI_max/min/avg/std_datediff	Min/max/avg/std of distance in days between multiple self-exclusion events.
days_since_last_SE_GI_wrt_max/min/avg/std	Days since last self-exclusion/gambling issue event vs min/max/avg/std of distance in days between multiple self-exclusion events.

TECHNICAL REQUIREMENTS

Scope:

Main scope of this test is to build a model to predict the customers who will call the Customer Service in the following 14 days.

Requirements:

- The overall solution can be implemented by using any analytical tool that you are more familiar with (Python 3.7 would be a plus).
- Notebook(s) must be well documented and explain why certain decisions/approaches have been taken.
- Don't focus just on modelling but dedicate some time also to Exploratory Data Analysis, Feature Engineering and Feature Selection.
- Model's performance should be evaluated properly
- A CSV file with the prediction for each customer in the test set. It should have 56,707 rows (including the column names) and should follow the following format:

mk_CurrentCustomer	Prediction
7343754	0
6210071	0
7343754	0
6210071	0
15837704	0
15235670	0
14755861	0
13625765	1
6445975	0
:	:

- A Powerpoint presentation in which you explain:
    - Highlights of the work done.
    - How to measure the success of the use case from a business point of view.

DELIVERY OF THE SOLUTION

The artefacts should be sent via email within 7 days from the communication of these requirements