**The overview**

An online sales company wants to predict if the user will come back to its website after 10 days based on the actions done within the first 3 days after registration. The model will be used in several marketing campaigns.

The dataset - train.csv

The test dataset - will not be available till the end of the exam - test.csv

**Description:**

Each row represents a user who registered and did some actions.

**Target:**

**came\_back:** 0 - user didn’t come back after 10 days, 1 - came back.

**Features:**

**n\_visits -** number of visits of the website during the first 3 days.

**n\_actions -** number of actions/button clicks in the website during the first 3 days.

**unique\_actions** - number of unique actions in the website during the first 3 days.

**products\_tried** - number of products viewed during the first 3 days.

**products\_added** - number of products added to cart during the first 3 days.

**unique\_products\_category\_tried** - number of unique products’ categories viewed during the first 3 days.

**Unique\_products\_category\_added** - number of unique products’ categories added to cart during the first 3 days.

**saved\_to\_cart** - if the user actually saved the purchases for later purchase during the first 3 days.

**comments\_writen** - how many unique comments the user did during the first 3 days.

**counts\_of\_buys** - how many times the user actually paid during the first 3 days.

**Allowed Models:**

* KNN
* Decision Trees
* Regressions
* Clustering

**Grading:**

Grading is based on the performance of the model on the test dataset. Keep in mind there could be 1- 4% accuracy fluctuations. The student provides the whole notebook with the final model, feature engineering and scaling, otherwise the results might differ from the original one.

The student receives the maximum grade if the accuracy metrics score is higher than **75%**. If the score is between 60% and 75% then the student gets half of the grade, otherwise zero.