Vodafone Data Science Hackathon 2021



The Sons of Analyst

- Luigi Noto
- Bernardo Principi
- Manuel Serra



For each session:

of events
max_time_delta
min_time_delta
mean_time_delta
mean_median_diff

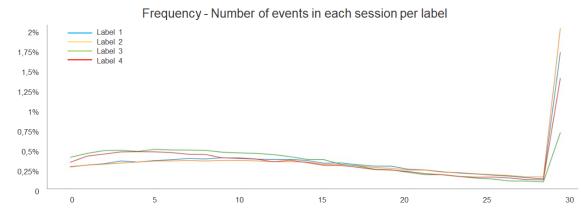
number of events in the session

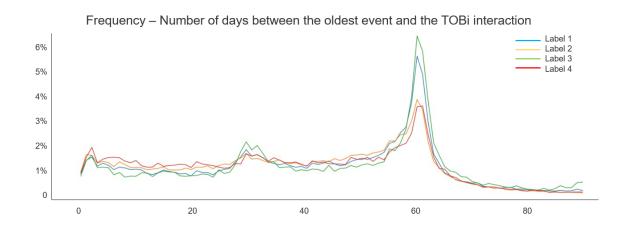
time delta (in days) between the oldest event and the interaction with TOBi

time delta (in days) between the most recent event and the interaction with TOBi

mean of the time delta (in days) between each event and the interaction with TOBi difference between the mean and median of the time delta between each event and

the interaction with TOBi











For each session:

weekend

week_number_0

week_number_17

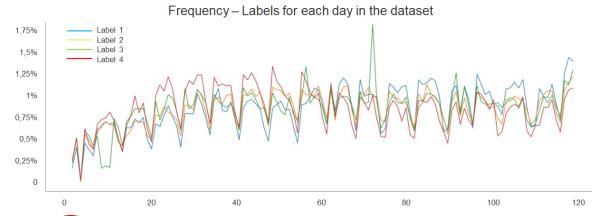
 $oldsymbol{1}$ if the interaction with TOBi occurs in the weekend

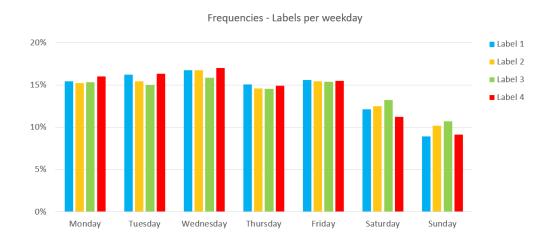
0 otherwise

 $oldsymbol{1}$ if the interaction with TOBi occurs in that specific week

0 otherwise













For each session:

Frequency of each concatenation of categories of the variables

event_category_idx_lv1, ..., event_category_idx_lv4 among the events of the session



session_id	concatenation_1	concatenation_2	concatenation_3	concatenation_4	concatenation_5	
1347460	1	0	0	0	0	
1347460	0	0	1	0	0	
1347460	0	1	0	0	0	
1347460	1	0	0	0	0	
1347460	0	0	0	0	1	•••

aggregate by summing



session_id	concatenation_1	concatenation_2	concatenation_3	concatenation_4	concatenation_5	
1347460	2	1	1	0	1	

compute relative frequencies



session_id	concatenation_1	concatenation_2	concatenation_3	concatenation_4	concatenation_5	
1347460	0,4	0,2	0,2	0	0,2	•••







for each label

For each session:

other_sessions_label_1

other_sessions_label_4

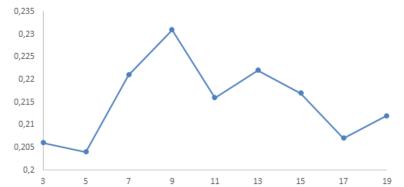
gini_index

customer_label_0

customer_label_8

if at least one of the other sessions of the same user has that specific label
 otherwise

Gini index of the concatenation of values of the variables event_category_idx_lv1, event_category_idx_lv2, event_category_idx_lv3, event_category_idx_lv4 across the events of the session







Model

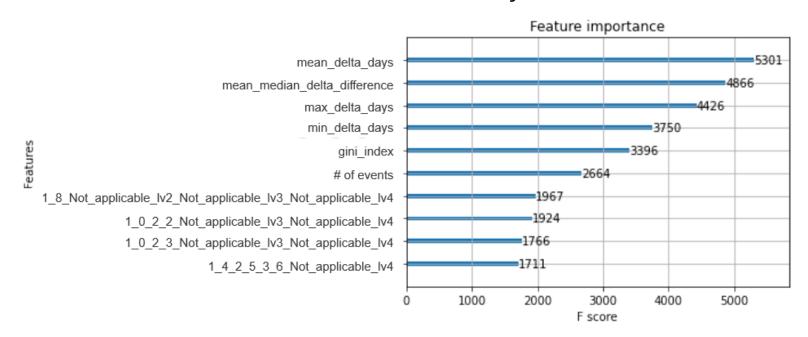


XGBoost

 $max_depth = 8$

eta = 0.5

CV accuracy = 0.54Test set accuracy = 0.54







Thanks for your attention

