

JONQUILLE - UPMC

13 juin 2016

Edouard, Mickael, Remi, Thomas

Master DAC - Université Pierre et Marie Curie - MLIA - LIP6



Learning pipeline

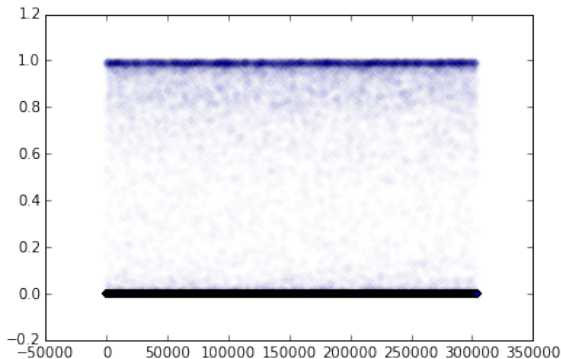
- 1 Understand data and prediction goal
- 2 Extract features
- 3 Train Xgboost
- 4 Train other models
- 5 ~~Go to sleep~~ Drink Café
- 6 Aggregate models

Results

Model	Score logloss
Mean	0.188734
Mean by SCID	0.143391
One Xgboost	0.010896
Ensemble Xgboost	0.010724

1.	nbQuotesByUser
2.	max_std_all_features_byUser
3.	mean_std_all_features_byUser
4.	std_broker_AllDriversNbConvictions
5.	mean_converted_byBroker
6.	std_broker_countSocioDemographicId
7.	nbUsersBySCID
8.	std_broker_DeysSineCarPurchase

Some quick observations



	>.90	>.95	>.99	Trainset
0	75.59%	79.42%	92.36%	89.34%
1	0.24%	0.20%	7.60%	10.24%
2	0.51%	0.31%	0.05%	0.33%