



User response prediction in mobile advertising

Faustine Bousquet, Khanh Chuong Duong, Christian Lavergne, Sophie Lèbre,
Anastasia Lieva

► To cite this version:

Faustine Bousquet, Khanh Chuong Duong, Christian Lavergne, Sophie Lèbre, Anastasia Lieva. User response prediction in mobile advertising. ECML PKDD, Sep 2018, Dublin, Ireland. hal-02014821

HAL Id: hal-02014821

<https://hal.science/hal-02014821>

Submitted on 11 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

User response prediction in mobile advertising

Faustine Bousquet^{*,1,2}, Khanh Chuong Duong¹, Christian Lavergne^{2,3}, Sophie Lebre^{2,3}, Anastasia Lieva¹

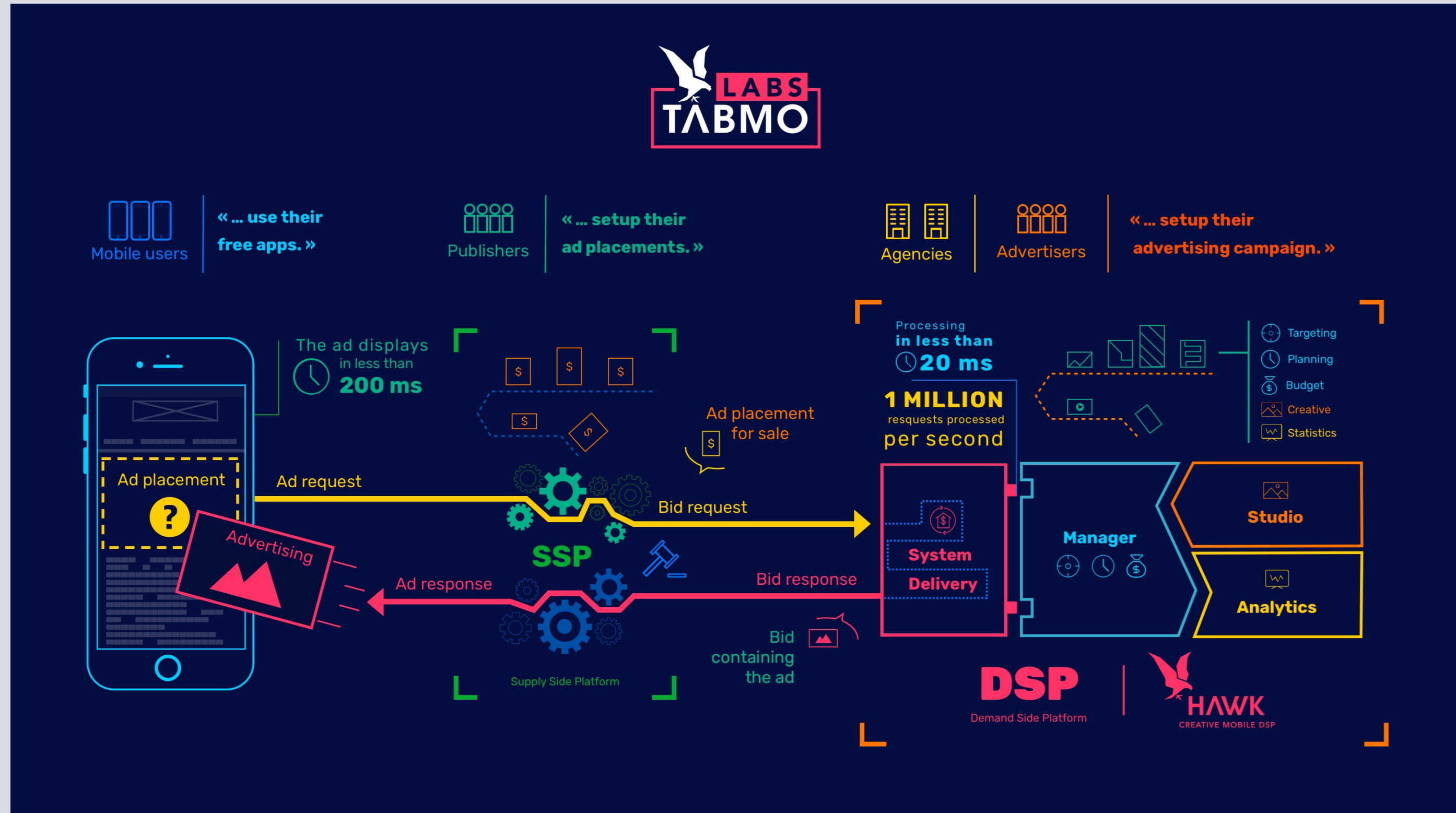
^{*} faustine.bousquet@tabmo.io

¹ TabMo Labs, Montpellier, France

² IMAG Institut Montpellierain Alexander Grothendieck, Université de Montpellier, France

³ Université Paul Valéry, Montpellier 3, France

Mobile Advertising Process

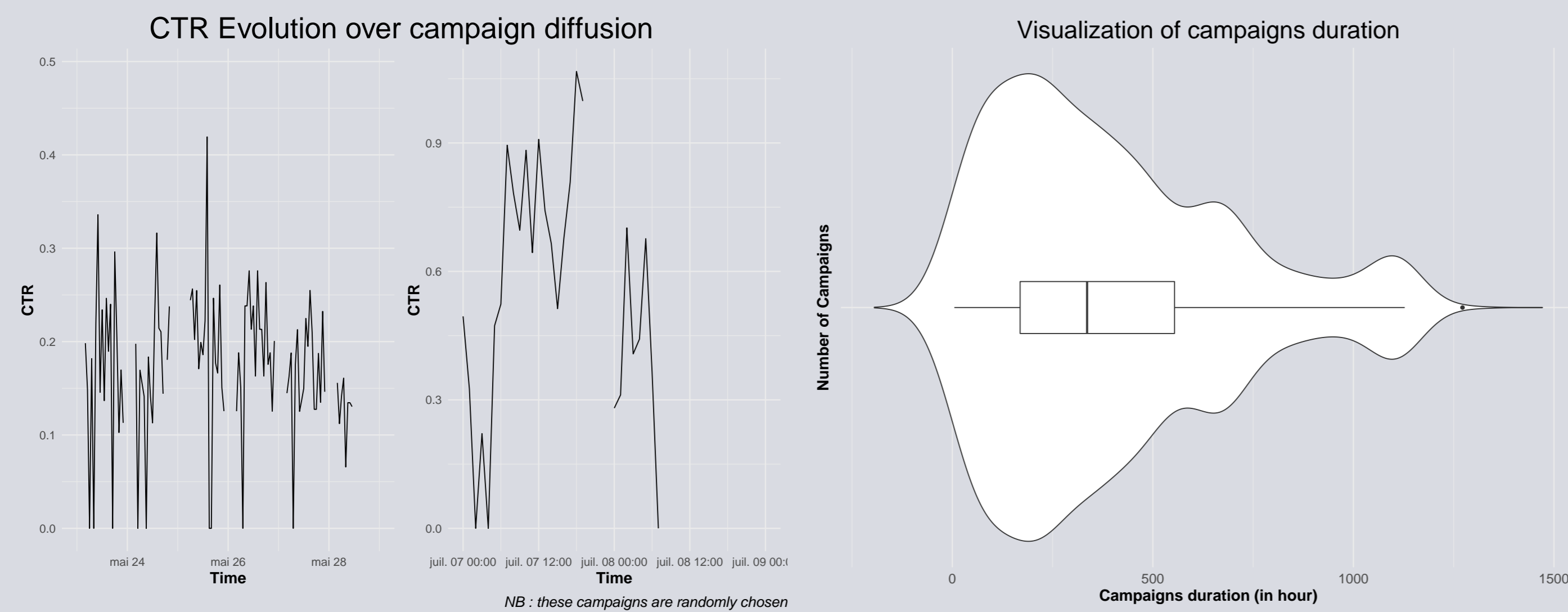


TabMo is an adtech company running Hawk platform. Our product has been built to be the only Creative Mobile Demand side platforms (DSP).

Some definitions:

- **Impressions:** the number of times an ad is displayed.
- **Click Through Rate (CTR)** = $\frac{\#Clicks}{\#Impressions}$ (In the following, CTR is calculated by hour)
- **Demand side platforms (DSP):** platform serving advertisers or ad agencies by bidding for their campaigns in multiple ad networks automatically.

Data visualization



Aims and objectives

Our aim: targeting the right person, at the right place and time with the most relevant ad

- **Clustering of mobile campaigns**
Each campaign has its own KPI to optimize. It can be a performance objective (lots of clicks), a branding strategy (lots of impressions) or more complicated goals which are not easy to handle through bid requests. Then, first part of the thesis focuses on obtaining clusters of campaigns.
- **Click prediction models for each cluster**
Clustering approach allows the estimation of a specific type of model for each cluster regarding its own objective. We consider that we know the KPI to optimize and predict. We have to increase the rate of this KPI with an appropriate, scalable and innovative predictive model in real time.

Mathematical approach

Model definition

Impressions and clicks are aggregated by hour. We calculate the corresponding CTR or number of impressions for each time slot. Observations can be described as:

For all,

- $c = 1, \dots, C$ campaigns,
- $j = d_c, \dots, f_c$ days of campaign C ,
- $h = 1, \dots, H$ time slots,
- $t = 1, \dots, T_{jh}$ repetitions of time slot h during day j ,

$$Y_{cjh} = \mu + \beta_h^H + \beta_s^S + \epsilon_{cjh}$$

where μ is a constant, β_h the time slot effect, β_s the day of week effect (assuming the following constraints: $\beta_1^H = \beta_1^S = 0$ for identifiability) and ϵ_{cjh} a gaussian error .

Mixture model

We assume that there are C campaigns which are part of K groups:

$$Z_{kc} = \begin{cases} 1 & \text{if campaign } c \text{ belongs to cluster } k \\ 0 & \text{otherwise} \end{cases}$$

The mixture model can then be written like:

$$f(y_c; \beta, \sigma^2) = \sum_{k=1}^K \lambda_k f_k(y_c; \beta_k, \sigma_k^2)$$

where $f(y_c) = \prod_{j=d_c}^{f_c} \prod_{h=1}^H \prod_{t=1}^{T_{jh}} f(y_{cjh})$ and $P(Z_{kc} = 1) = \lambda_k$, the probability that Y_c belongs to k . We used a classical Expectation-Maximisation (EM) [1] algorithm to estimate this mixture model.

Criteria for choosing number of cluster

- **Bayesian Information Criterion (BIC)** : $BIC(K) = -2 \times \log(Ln(Y; \hat{\beta}_K, \hat{\sigma}_K^2)) + \#parameters \times \log N$ where N is the number of observations.
- **Integrate Classification Likelihood** [2] (ICL) which penalizes the complete log likelihood : $ICL(K) = BIC(K) - 2 \sum_{c=1}^C \sum_{k=1}^K \tau_{kc} \log \tau_{kc}$ where $\tau_{kc} = P(Z_{kc} = 1 | Y_c)$

References

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [2] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated classification likelihood," 1998.

Results of simulations

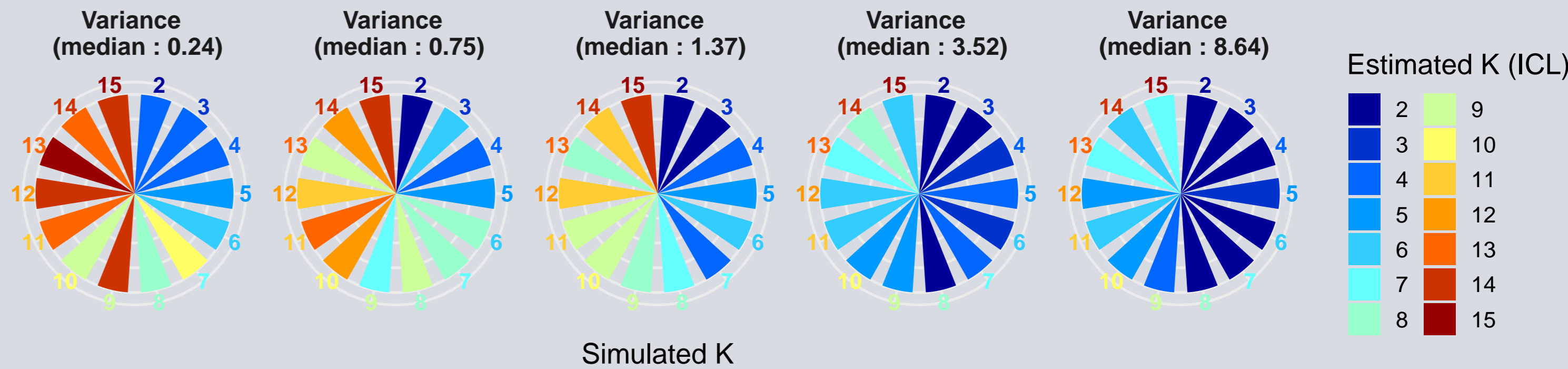
The objective of this design of experiment (DOE) is to evaluate limits of our EM algorithm when noise variance increases.

Simulation settings on 700 campaigns whose CTR is simulated:

- Clusters are equidistributed
- Beta values are estimated on real experiments with $H = 5$ time slot and $S = 7$ day of week. $\beta \in \mathbb{R}^{11}$ and their absolute values vary from 0 to 18.13 with a median value equal to 0.18.

BIC/ICL estimated number of clusters VS simulated number of clusters

Comparison of estimated and simulated number of clusters when noise variance increases



Some explanation : if we take the second piechart whose variance median is 0.75:

- When the number of cluster simulated is $K=2$, the number of cluster estimated is $K=2$.
- When the number of cluster simulated is $K=13$, the number of cluster estimated is $K=9$.

Campaigns confusion matrix (K=4) :

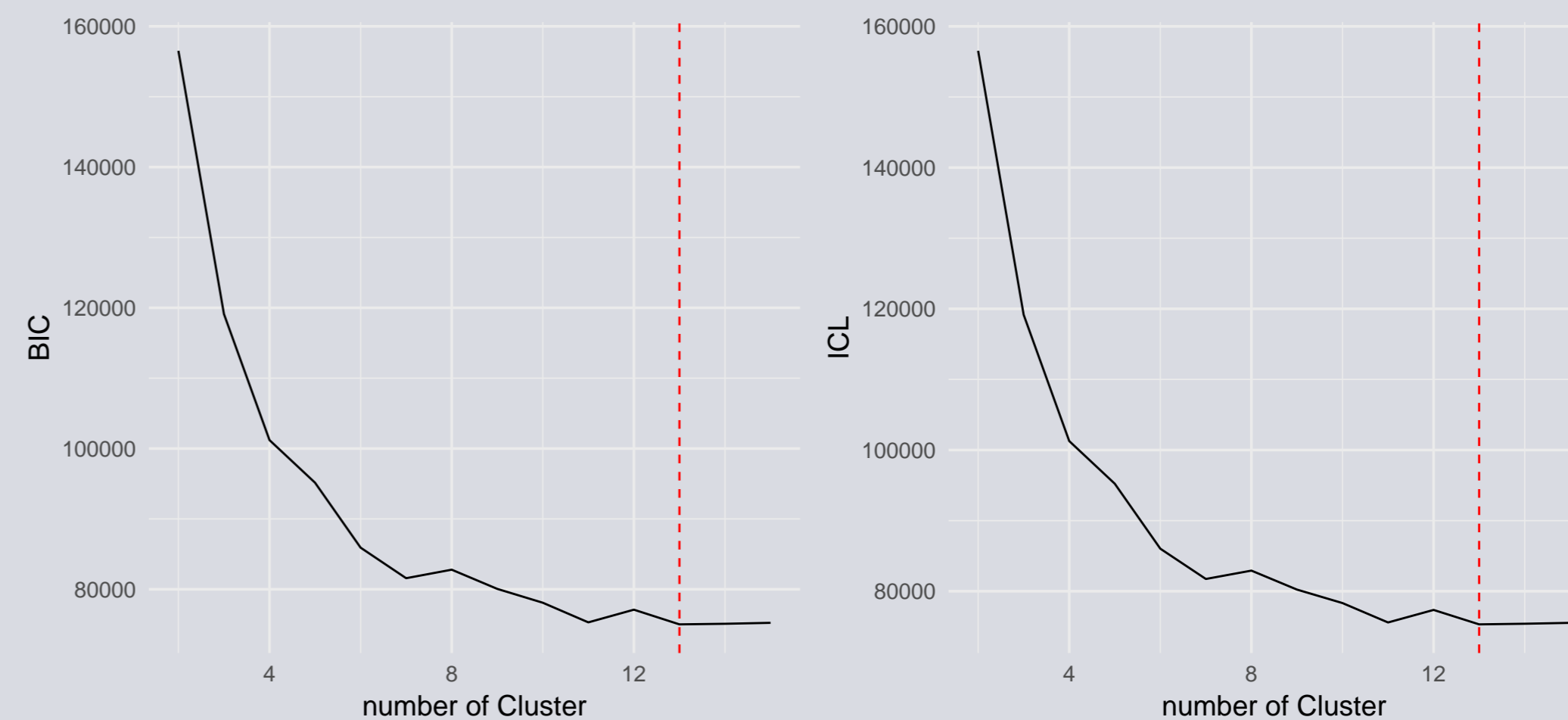
Variance (median : 0.24)					Variance (median : 0.75)					Variance (median : 1.37)					Variance (median : 3.52)					Variance (median : 8.64)				
1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4	
1	8	0	227	0	1	16	0	218	0	1	31	2	214	0	1	33	6	214	1	1	40	10	209	2
2	0	4	0	99	2	0	8	2	92	2	0	9	0	83	2	1	12	0	72	2	0	7	0	47
3	109	0	9	0	3	101	0	16	0	3	86	0	18	0	3	83	0	17	0	3	76	0	20	0
4	0	232	0	12	4	0	228	0	19	4	0	225	4	28	4	0	218	5	38	4	1	219	7	62

First Clustering result

First results are on CTR metric. We worked with 700 campaigns which started and ended between May the 10th and July the 10th. Our model included 2 temporal variables:

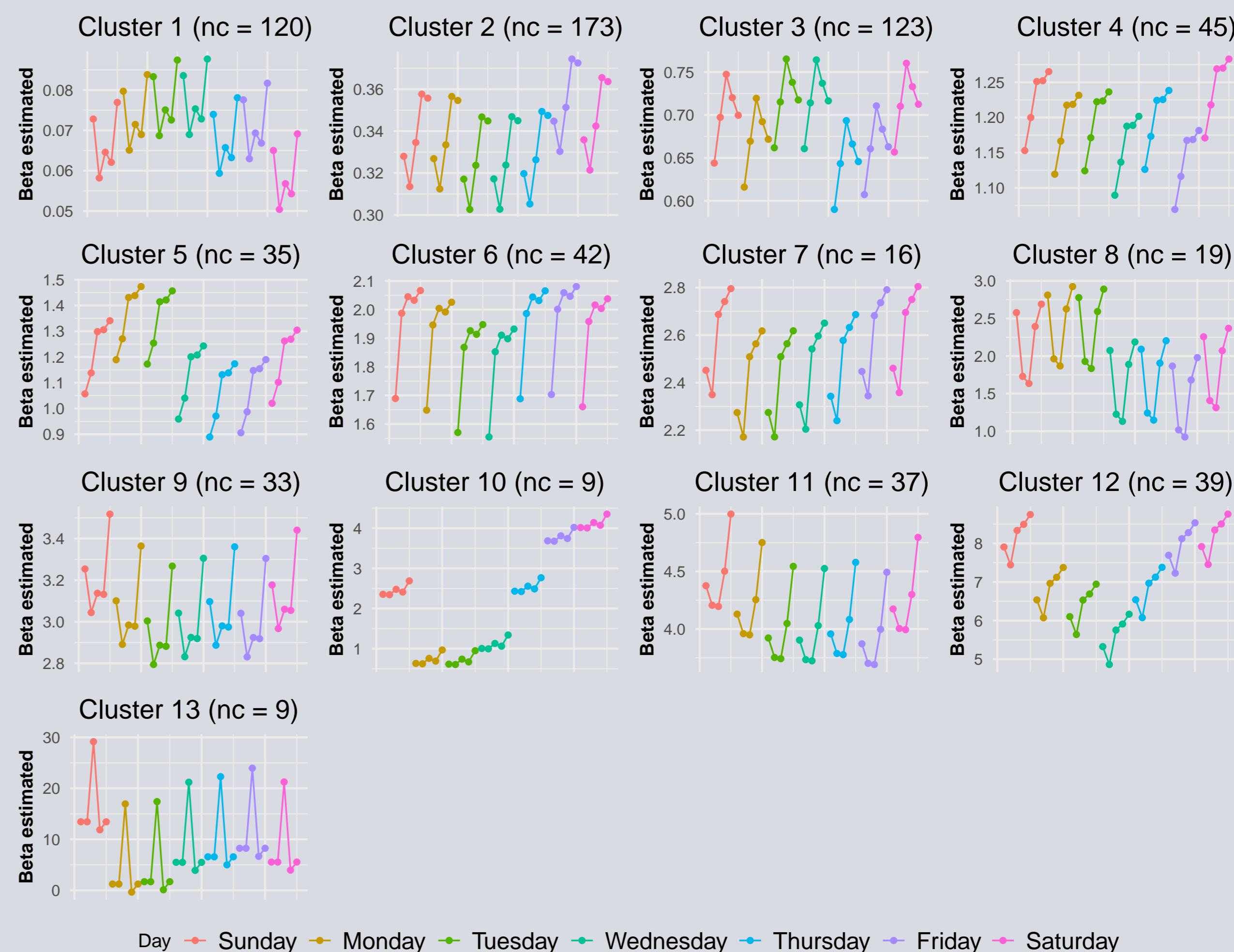
- day of week (cardinality $S = 7$)
- time of the day into buckets (cardinality $H = 5$)

Optimal number of clusters by BIC/ICL criteria:



Inferred profiles:

- Beta values are very different from one cluster to another.
- Same observation about clusters size : they include from 9 to 123 campaigns.
- Time slot and day of week effect seem to be significant.



Conclusions

- First results are encouraging. We obtain different cluster profiles with only few temporal variables. Evaluating algorithm, its limit and statistical results still remain in progress and are part of the challenges of this thesis. Also, we should work with domain expert to validate pertinence of clusters.
- Context variables on ads (size, type) and device (OS, model and so on) will enrich the model. Hopefully, these new variables and their interactions will lead to more and more homogeneous clusters.
- Once the clustering objective achieved, the goal will be to determine an adaptive predictive model for each cluster. The model should be able to scale the large amount of request received per second (around 1 million).