NexComm 2022 Congress

The Fifteenth International Conference on Communication Theory, Reliability, and Quality of Service

CTRQ 2022

April 24, 2022 to April 28, 2022 - Barcelona, Spain

Explaining Radio Access Network User Dissatisfaction with Multiple Regression Models

Adrien Schaffner
Louise Travé-Massuyès
Simon Pachy
Bertrand Le Marec

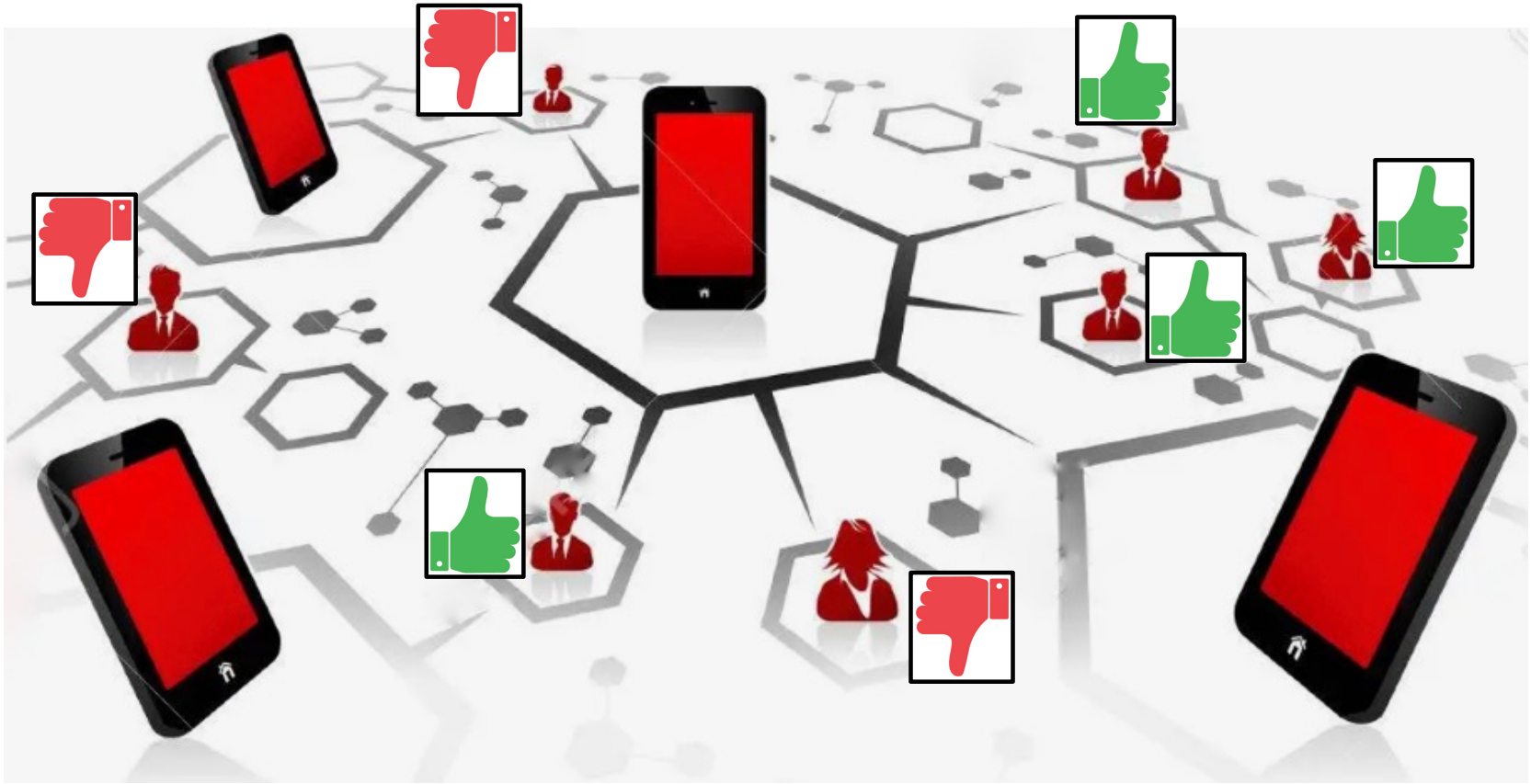
Presenter: Louise Travé-Massuyès, LAAS-CNRS & ANITI, louise.trave@cnrs.fr

LivingObjects is a leading provider of network performance management solutions for operators and service providers. LivingObjects' scalable platforms monitor, manage and optimize mobile, fixed or converged multi-technology network infrastructure as well as the associated deployed services.

- ➤ Enabling cross domain end-to-end analysis
- ➤ Implementing best analytics algorithms to fit performance management
- ➤ Leveraging latest visualization features to enhance data browsing
- ➤ Automating recurring performance management tasks

Louise Travé-Massuyès holds a position of *Directrice de Recherche* at *Laboratoire d'Analyse et d'Architecture des Systèmes*, *Centre National de la Recherche Scientifique* (LAAS-CNRS, https://www.laas.fr), Toulouse, France. Her main research interests are in **dynamic systems diagnosis and supervision** with special focus on **model-based reasoning** methods and **data mining**. She is member of the International Federation of Automatic Control IFAC (https://www.ifac-control.org/) Safeprocess Technical Committee, Treasurer of the Society of Automatic Control, Industrial and Production Engineering (https://www.sagip.org). She holds the chair "Synergistic transformations in model-based and data-based diagnosis" in the Artificial and Natural Intelligence Toulouse Institute (ANITI), France (https://aniti.univ-toulouse.fr) and serves as Associate Editor for the well-known Artificial Intelligence Journal (https://www.journals.elsevier.com/artificial-intelligence).

User satisfaction



➤ Essential performance indicator

➤Complaints give a good trend of network quality perceived by customers

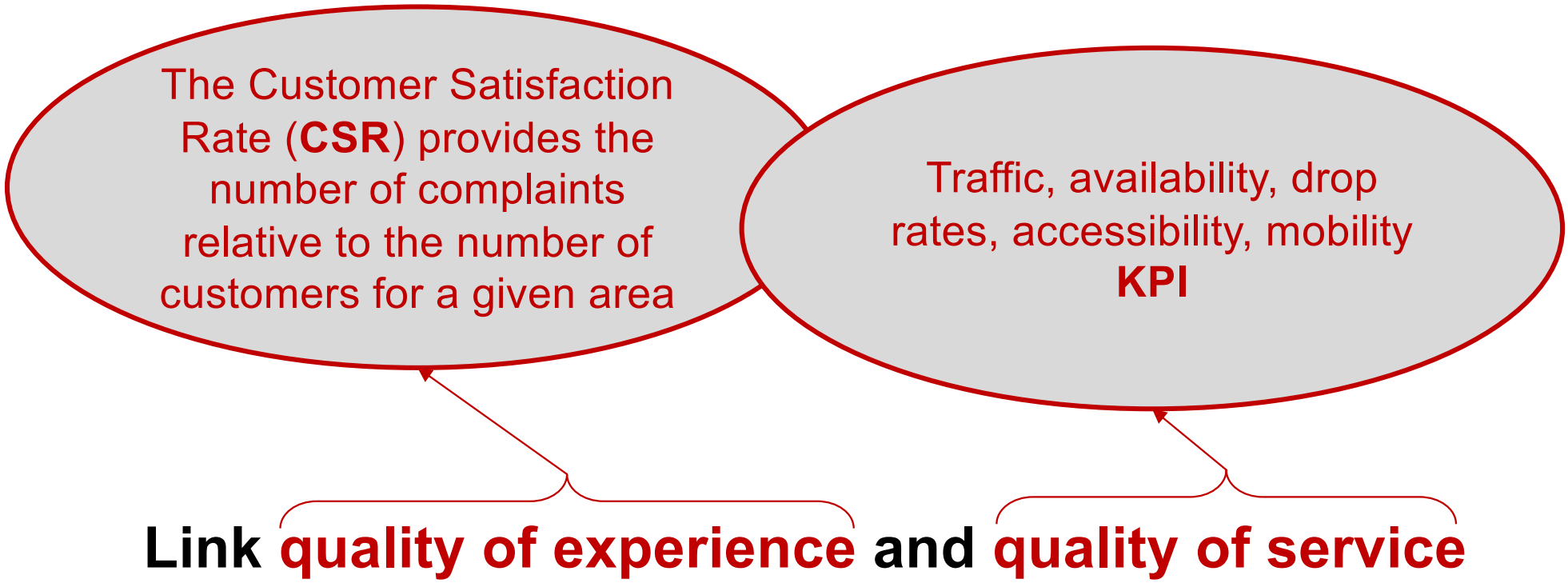
Subjective customer comments ⟷ Objective behavior of the network

Objectives of the work

➤ Design a model that links the complaint rate with a set of key performance indicators
➤ Provide performance engineers with a better understanding of customer expectations
➤ Help performance engineers act primarily on the indicators that give the most dissatisfaction

Link quality of experience and quality of service

QoE and QoS link



Link quality of experience and quality of service

Problem formulation

A lot of work using Machine Learning, in particular for churn prediction.

Our approach:

➤ Uses solely complaint data to solve problems in maintenance

➤ Handles the KPIs which are the data used on a daily basis by network monitoring operators

➤ Provides the ability to explain predictions: clearly explain this link so that it provides useful information

It is based on simple regression models

Multiple linear regression

A learning method that postulates that a variable y (here y=CSR) is expressed as the weighted sum of other variables $x_j, j = 1,...,p$ (here the KPIs).
The goal is to learn weights $\beta_0, \beta_1, ..., \beta p$ such as:

$$y = \beta_0 + \beta_1 x_1 + ...\beta_p x_p$$

Dataset of n samples: $(x_1^i, x_2^i, \ldots, x_p^i, y^i),\ i = 1, \ldots, n$

$$y^i = \beta_0 + \beta_1 x_1^i + ...\beta_p x_p^i + \epsilon_i,\ i = 1, \ldots, n$$

Solved by least squares minimization or likelihood maximization.

Our case study → $CSR = \beta_0 + \beta_1 KPI_1 + ... + \beta_p KPI_p$

Extract knowledge: determine the KPIs that influence the CSR and to quantify their influence from the coefficients of the regression.

Loss function

$$\mathcal{L} = \min_{\beta_0, \beta_1, \ldots, \beta_p} \frac{1}{2} \sum_{i=1}^{n} (y^i - \hat{y}^i)^2 + \lambda \sum_{j=1}^{p} | \beta_j |$$

$$- \beta_0 - \sum_{j=1}^{p} \beta_j x_j^i)^2$$

Ordinary Least Squares (**OLS**)

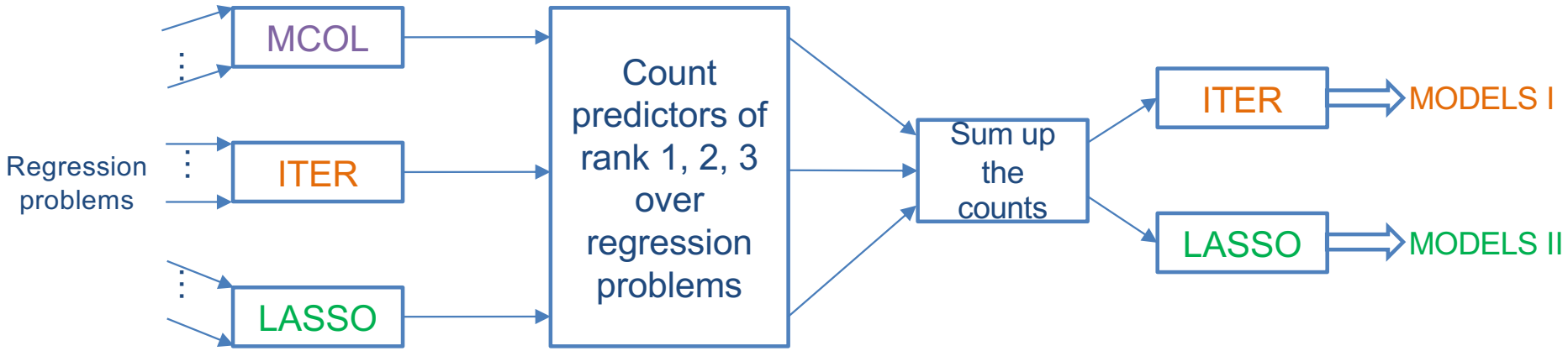
LASSO

L1 regularization term: shrinks the less important feature's coefficients to zero

A fusion regression model

Extract knowledge: determine the KPIs that influence the CSR and to quantify their influence from the coefficients of the regression:

- variable selection
- weigths learning



MCOL method

Multicollinearity analysis with OLS

➤ Multicolinearity: some explanatory variables in the model measure the « same thing »

➤ Multicolinearity ⇒ strong correlation

➤ Can be assessed by the *Variance Inflation Factor* (VIF): ratio of the overall model variance to the variance of a model that includes only that single explanatory variable

➤ No multicollinearity : all VIF(x_j) equal to 1

Preprocessing + OLS

Computation of VIF(x_j), $j = 1, \ldots, p$

If VIF(x_j) ≥ α, then eliminate x_j

ITER method

Iterative reduction via p-value with OLS

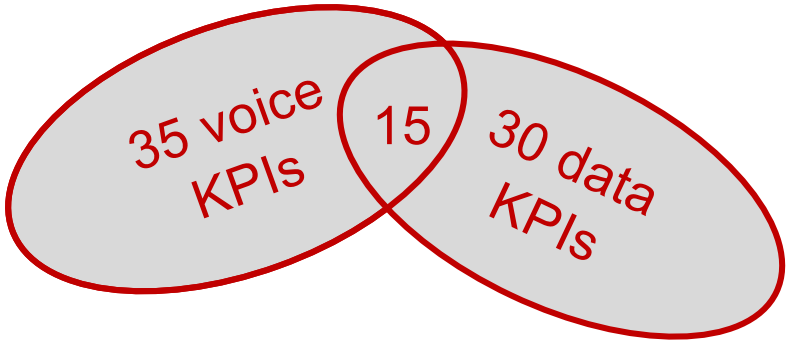
➤ a ***p-value*** *p* for each variable can be obtained
➤ the *p-value p* tests the null hypothesis that the coefficient is equal to zero
➤ p ≤ 0,05 indicates that one can reject the null hypothesis, i.e., the predictor is a meaningful addition to the model as it changes the model prediction

1- Train a model with all variables,
2- Remove the one with the highest p-value.
3- Iterate

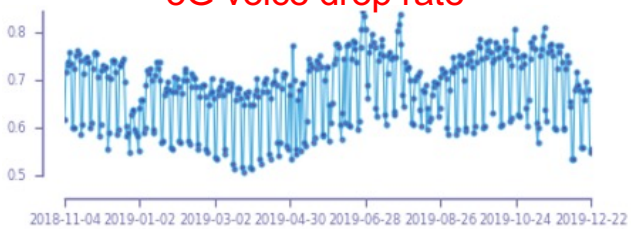
Data

At the level of the 93 French departments covering a full year at week granularity:

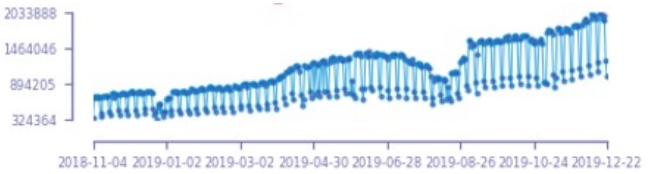
➤ mixture of signals for both 2G, 3G, and 4G for six classes : **traffic**, **availability**, **drop rates**, **accessibility**, **performance**, and **mobility**



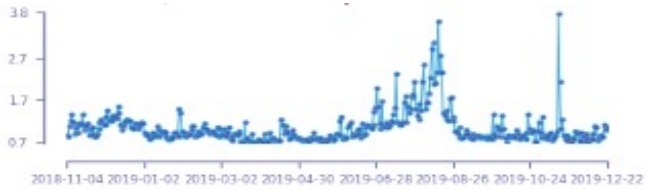
3G voice drop rate



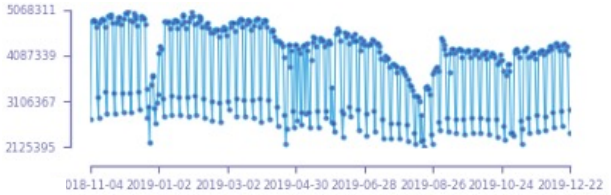
4G voice traffic



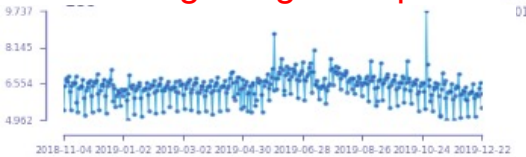
3G voice failure rate



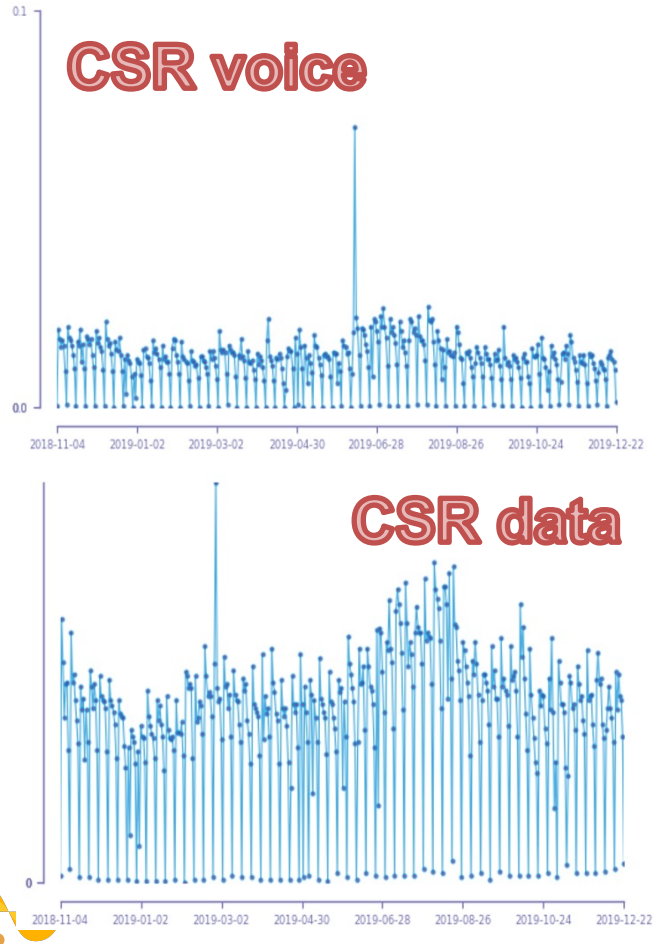
3G voice traffic



3G signaling attemps



Data



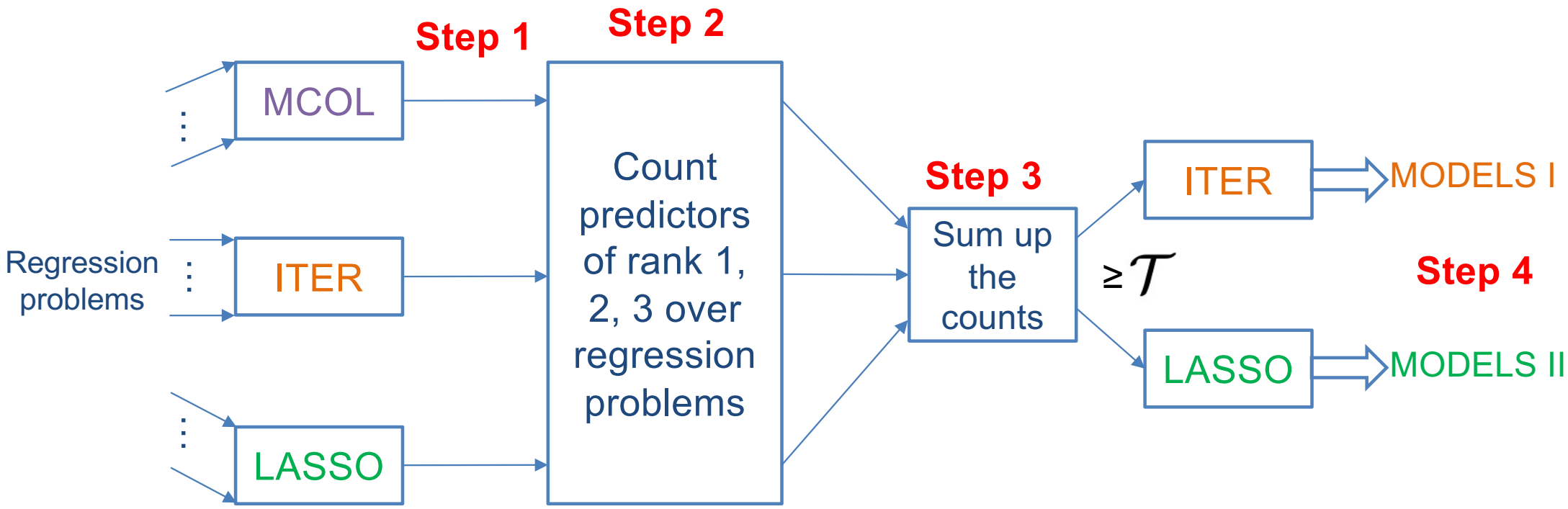
Preprocessing

Some variables are removed:

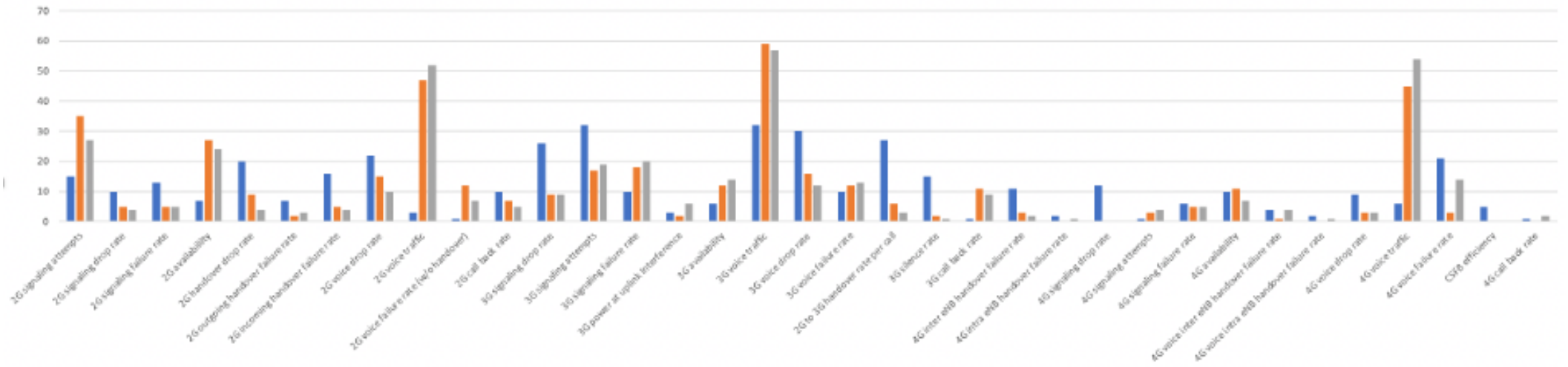
➤ strongly correlated variables
➤ low variance through the dataset

Variables are scaled so that they could be ranked according to the magnitude of their corresponding weights in the regressions.

Fusion method

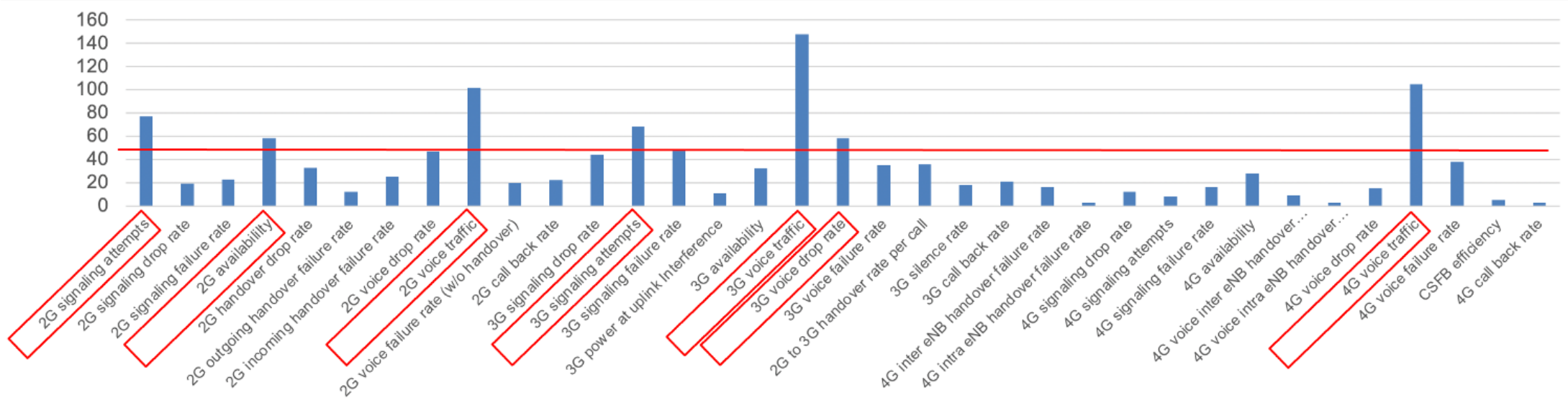


Steps 1-2 (Voice performance problem)



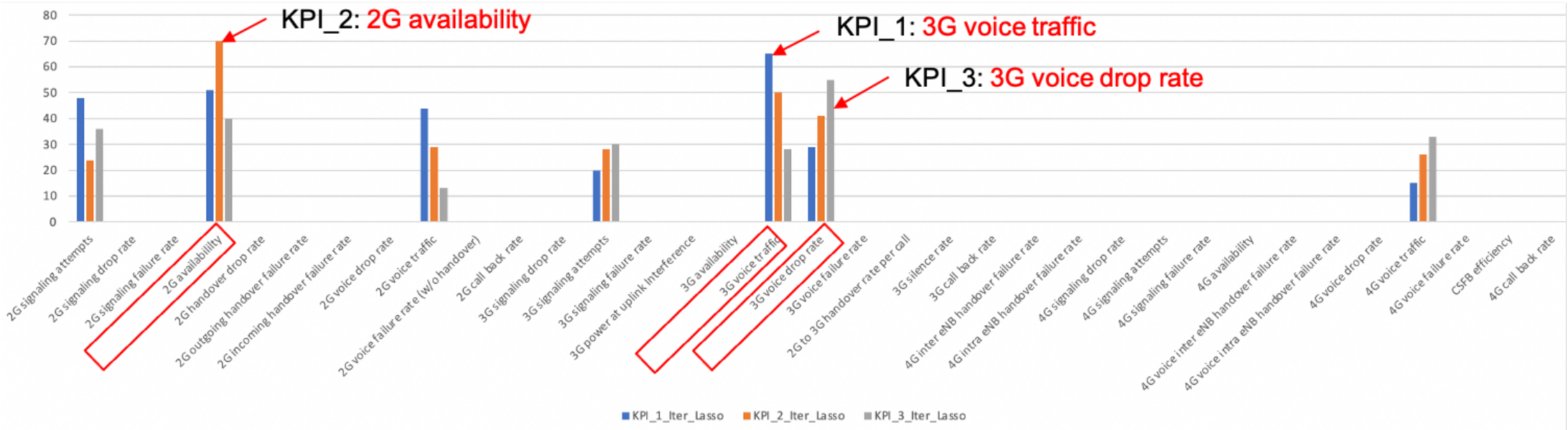
Count of the number of times an explanatory KPI is ranked 1, 2, or 3 by MCOL (blue), ITER (orange), LASSO (grey)

Step 3 (Voice performance problem)



Sum of the counts of the number of times an explanatory KPI is ranked 1, 2, or 3 by MCOL, ITER, LASSO. KPIs framed in red count above the threshold.

Stand out KPIs (Voice performance problem)



KPIs ranked 1, 2, and 3 over ITER and LASSO and over all the French departments.

Aggregated interpretation

- **3G voice traffic** → network unavailability, loss of coverage, and network engineering issues
- **2G availability** → network maintenance processes
- **3G voice drop rate** → network call drops

Accessibility or mobility issues are less impacting than call drops or traffic issues

Priority actions

- optimize the maintenance process to reduce unavailability periods
- modify network parameter settings, optimize site engineering, or build new sites to improve the call drop rate

Interpretation by department

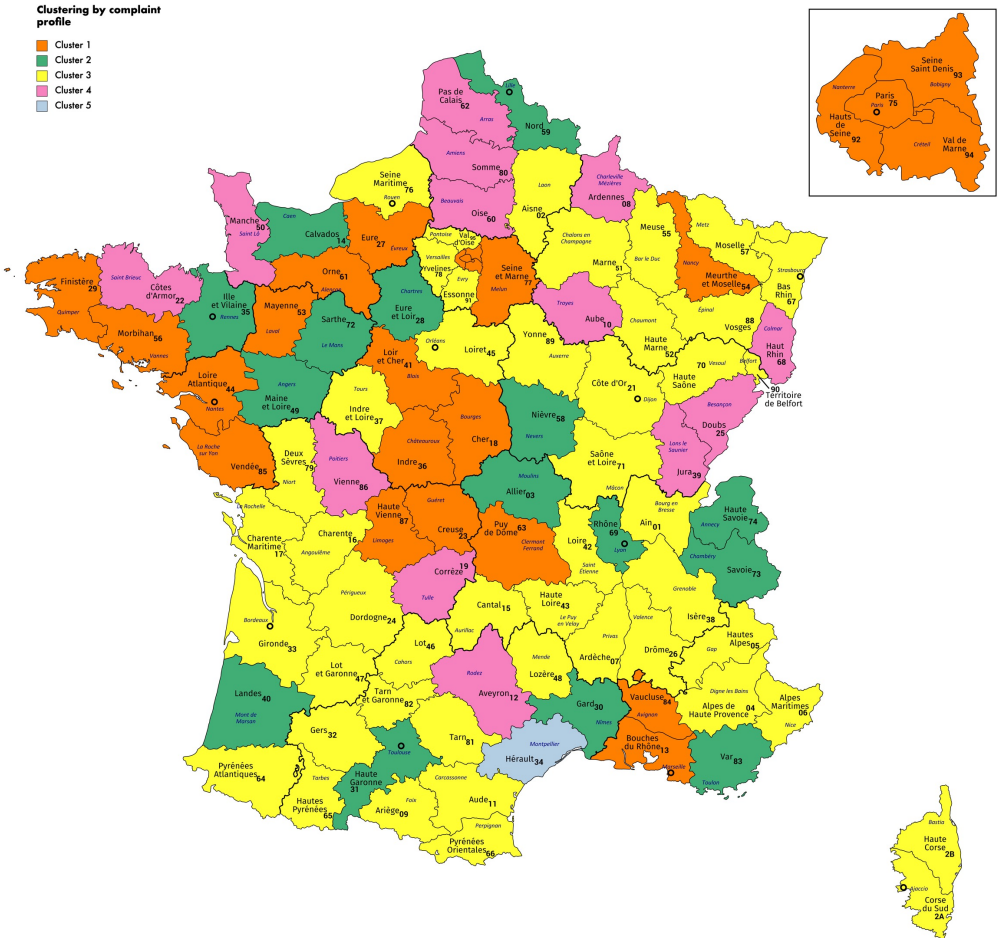
Final ITER models for each French department

⇩

Cluster the weigths

⇩

Department profile



Conclusion

➤ The work leverages the benefits of three methods to select relevant explanatory variables and deduce a robust regression model

➤ Method tested on telecom data : model that links the complaint rate with a set of objective performance indicators

➤ Profiling French departments

➤ Top KPIs on a global scale

Perspective work:

➤ Map the top KPIs to actual actions

NexComm 2022 Congress

The Fifteenth International Conference on Communication Theory, Reliability, and Quality of Service

CTRQ 2022

April 24, 2022 to April 28, 2022 - Barcelona, Spain

Explaining Radio Access Network User Dissatisfaction with Multiple Regression Models

Adrien Schaffner
Louise Travé-Massuyès
Simon Pachy
Bertrand Le Marec

Presenter: Louise Travé-Massuyès, LAAS-CNRS & ANITI, louise.trave@cnrs.fr