

# Utilisation d’algorithme scalable pour la prédiction des taux de déperditions de clients en télécommunication

El Hassane Nassif \*, Hicham Hajji \*,  
Rizlane Guati \*\*

\* Institut Agronomique et Vétérinaire Hassan II , Rabat  
nassif.hassane@gmail.com, h.hajji@iav.ac.ma  
<http://www.iav.ac.ma>

\*\* Ecole Nationale de Commerce et de Gestion de Casablanca  
rguati@gmail.com  
<https://www.encgcasa.ac.ma/>

## 1 Présentation de la thèse

Le Churn est un terme utilisé dans le domaine des télécommunications pour indiquer la résiliation du contrat qui lie le client à l’entreprise Telecom. Ce phénomène également appelé attrition est un problème-clé pour les opérateurs et remet en question leur performance sur le marché (Lemmens et Gupta (2017)). Étant donné que la survie de l’entreprise dépend fortement de sa capacité à retenir ses clients existants, il était nécessaire pour les opérateurs de mettre en place une stratégie de rétention prédictive. En mettant en place un processus de détection des clients dont la probabilité de Churn est élevée. De nombreuses approches de prédictions (linéaires et non linéaires) ont été suivies pour aborder ce sujet, mais sans donner des résultats suffisants. Cela est dû principalement à la limitation de la taille de la population d’apprentissage et/ou le nombre des variables descriptives. En effet, ces deux facteurs ont tendance à limiter le pouvoir prédictif des algorithmes et par conséquent la qualité des modèles qui s’en dégagent. Nous proposons dans travail d’implémenter une approche d’apprentissage automatique ensembliste basée sur le XGBoost (Chen et Guestrin (2016)) et distribuée sur un environnement Spark 2.1.0 avec Dataframe (Armbrust et al. (2015)). Aussi, nous montrons que ce couplage donne de meilleurs résultats et s’adapte mieux à la contrainte du volume des données d’apprentissage. D’autres parts, la construction du modèle XGBoost dans notre cas de prédiction des taux de déperditions a permis de trouver de résultats très satisfaisants.

## Références

Armbrust, M., R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi, et al. (2015). Spark sql : Relational data processing in spark. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 1383–1394. ACM.

## Prédiction du churn et Machine Learning

Chen, T. et C. Guestrin (2016). Xgboost : A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794. ACM.

Lemmens, A. et S. Gupta (2017). Managing churn to maximize profits.