

# Florian Douetteau, Dataiku : « Le GDPR va remodeler les applications Big Data »

Dataiku est une des principales pépites de la French Tech en matière de Big Data. Pour ne pas dire celle qui est la plus fréquemment citée si on se fie aux échanges que nous avons eus sur le salon Big Data Paris, qui a eu lieu les 6 et 7 mars. La société vient d'annoncer la sortie de la version 4 de sa technologie, évolution qui facilite notamment les passages en production des applications de Data Science et inclut des fonctions de traçabilité des données afin d'aider les entreprises à se conformer à leurs obligations réglementaires. Pour Florian Douetteau, ancien d'Exalead qui a cofondé Dataiku en 2013, ces inflexions correspondent aux attentes actuelles du marché.

## **Silicon.fr : Où en sont les entreprises françaises dans leur appropriation des technologies d'intelligence artificielle ?**

**Florian Douetteau :** La majorité d'entre elles sont encore en phase de découverte ou d'apprentissage. On peut estimer que 20 % des entreprises ont toutefois démarré un prototype sur le sujet, une proportion qui monte à 50 % dans notre base de clients. C'est le cas de la plupart des grands acteurs du e-commerce et du retail, qui y ont recours pour optimiser leur ciblage marketing, la personnalisation de leurs offres, leur logistique ou mieux détecter la fraude. Le constat est similaire dans la banque et l'assurance, des acteurs en profonde mutation car ils ont besoin de renouveler leurs offres. C'est plus difficile dans l'industrie où prouver la valeur de ces technologies nécessite de capter des données, de faire des tests, de changer des processus de fabrication... On parle là plutôt de cycles de 5 à 10 ans. Il se passera encore beaucoup de temps avant qu'un moteur d'avion ne soit entièrement vérifié par Machine Learning...

Grâce à notre présence dans 15 pays désormais, nous observons aussi de réelles différences en fonction de la nationalité des entreprises. Avec une appétence plus prononcée en Grande-Bretagne ou aux Etats-Unis. Dans des entreprises du Fortune 500, on voit naître de nouvelles plates-formes d'exploitation de la donnée avec 100 ou 200 utilisateurs. Sur la plateforme Dataiku, plus de 50 clients ont aujourd'hui des applications de Machine Learning en production.

## **Sait-on déjà quel algorithme de Machine Learning est le plus adapté à telle problématique ou le tâtonnement est-il un passage obligé ?**

**F.D. :** Les connaissances sont déjà solides en la matière. On sait, par exemple, que les chaînes de Markov sont efficaces pour lutter contre la fraude aux paiements. Ou que certains types de Deep Learning donnent de bons résultats pour le traitement d'images. Cette grammaire de ce qui fonctionne ou pas émerge sur le marché. Elle s'installe d'autant plus vite que la culture technologique qui entoure le Machine Learning et le Big Data est une culture de partage, impulsée par des sociétés comme Facebook, Google ou AWS. Même si des vendeurs de solutions plus classiques comme IBM ou Palantir ne s'inscrivent évidemment pas dans ce type de démarche.

**Si le domaine de la recherche est fécond, les retours d'expérience des entreprises montrent que le passage des prototypes Big Data à l'application en production reste très difficile...**

**F.D.** : Cela peut effectivement poser problème car on parle de technologies nouvelles. C'est d'ailleurs la raison qui nous a poussés à bâtir une plateforme afin de réduire le coût de cette transition. La logique étant, à l'intérieur de cette plateforme, de procéder à des mises en production agiles. C'est d'autant plus important avec le Machine Learning où le premier modèle qui sera essayé a de bonnes chances de ne pas se révéler efficace avec les données de production. C'est aussi un domaine où il faut procéder à de nombreux tests comparatifs de modèles, ou AB Testing. Si les mises en production sont difficiles, le retour de bâton sera sérieux.

### **Le passage en production soulève également souvent la question du respect des réglementations...**

**F.D.** : On observe effectivement ce phénomène, d'autant plus que notre solution vient souvent compléter un datalake existant, où les données ont été agrégées sans toujours réfléchir aux implications réglementaires. Tous nos clients américains se posent la question de l'auditabilité de leurs processus analytiques, au regard de Sarbanes Oxley ou des législations sur la santé. De ce côté-ci de l'Atlantique, le règlement européen GDPR va amener des contraintes de ce type, en précisant les obligations ambiguës qui pèsent aujourd'hui sur les entreprises et en alourdissant significativement les amendes en cas d'infraction. Par exemple, si on effectue un profilage, il faudra préciser les données qui sont utilisées, y compris celles de tiers. Dans la plupart des cas, cela signifie une reconstruction de bout en bout des processus analytiques.

#### **A lire aussi :**

[Big Data : il est temps de grandir... ou de mourir](#)

[F. Douetteau, Dataiku : « Nul besoin de data scientists superstars pour le Big Data »](#)

[Big Data : Blablacar copilote sa BI avec HP, Tableau et Dataiku](#)