# Comment se structure le secteur du machine learning ?

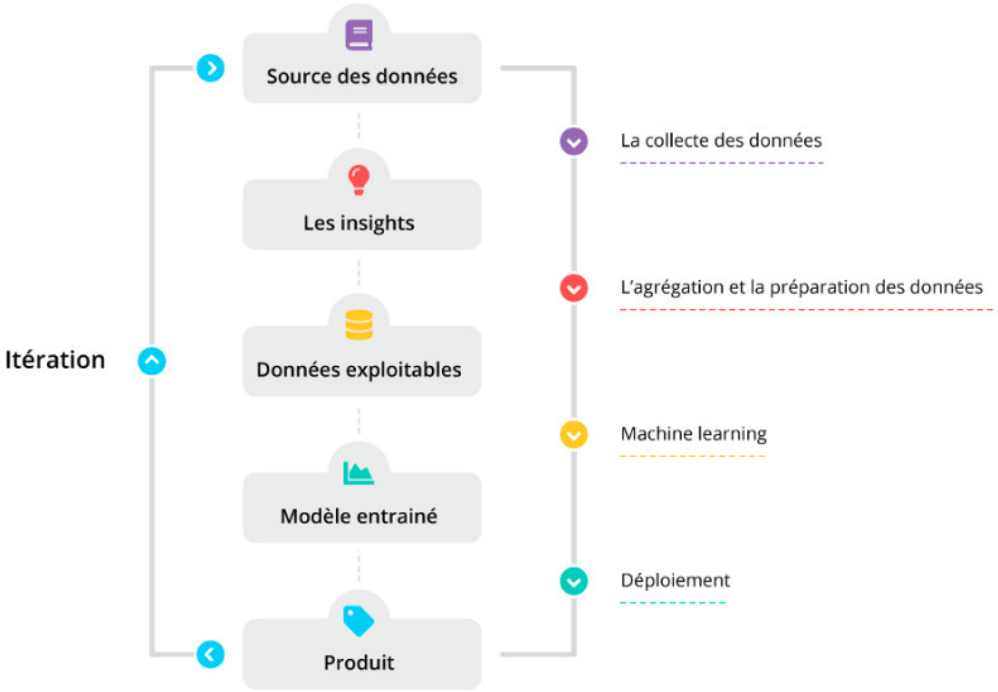
*Paramètres, technologies et cas d'utilisation, orchestrateurs ou encore produits.... Eytan Messika, Venture Catalyst chez OneRagtime, fait le point sur les différents acteurs du machine learning.*

Temps de lecture : minute

---

*20 juillet 2018*

Savez-vous véritablement comme s'organise le secteur du machine learning ? Entre les sources de données, les insights, les frameworks, les outils d'acquisition de données ou encore les infrastructures, il est en effet très facile de s'y perdre. Eytan Messika, Venture Catalyst chez OneRagtime, fait un état des lieux des différents acteurs du machine learning.




## 1. Les paramètres :

*Data :*

- Les sources de données: Les sources pour récolter de la donnée sont multiples. Les

données peuvent être publiques (accessible sur un site web en ligne) ou bien privées (les données d'un patient par exemple). Les sources peuvent elles aussi être de nature publique (réseaux sociaux, API publiques, open source) ou bien privées (CRM ou ERP, Iot network, Cloud privé, Data lake…). Plus, la source est maîtrisée en interne, plus la donnée crée un avantage compétitif parce qu'elle ne pourra pas être utilisée par un concurrent.

- Les insights : Ce sont les données réelles récoltées. De façon pragmatique, il en existe de toutes sortes et de toutes structures. Voici un graphique qui en fait une liste pour bien comprendre.

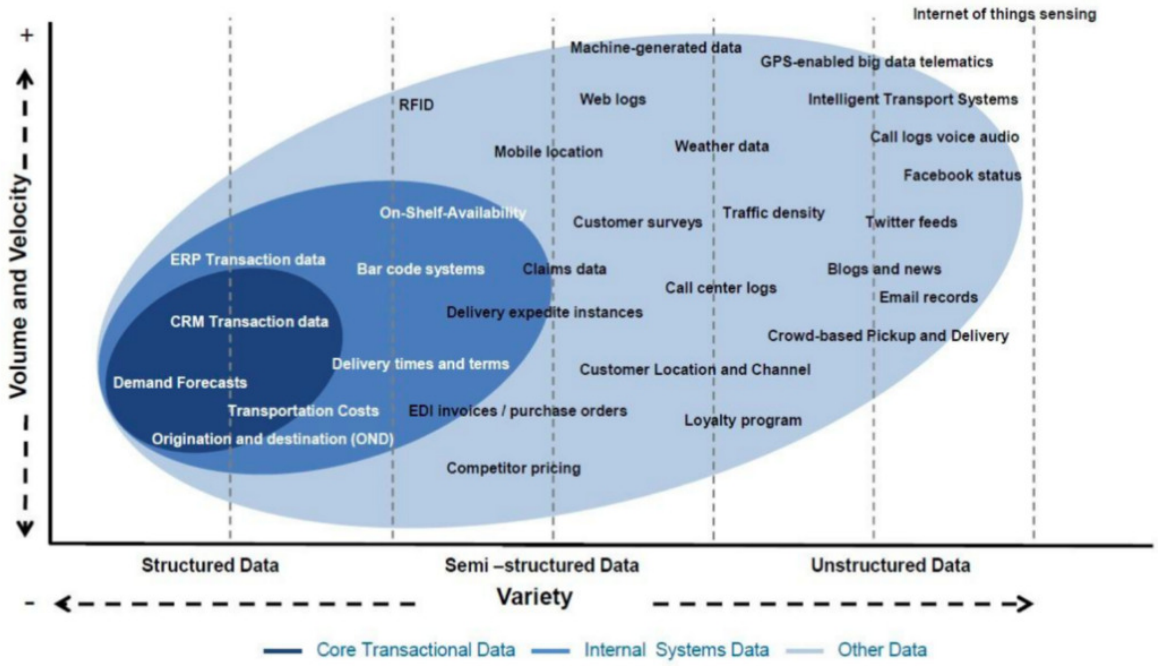



**Figure 1**. SCM Data Volume and Velocity vs. Variety

*Logiciel :*

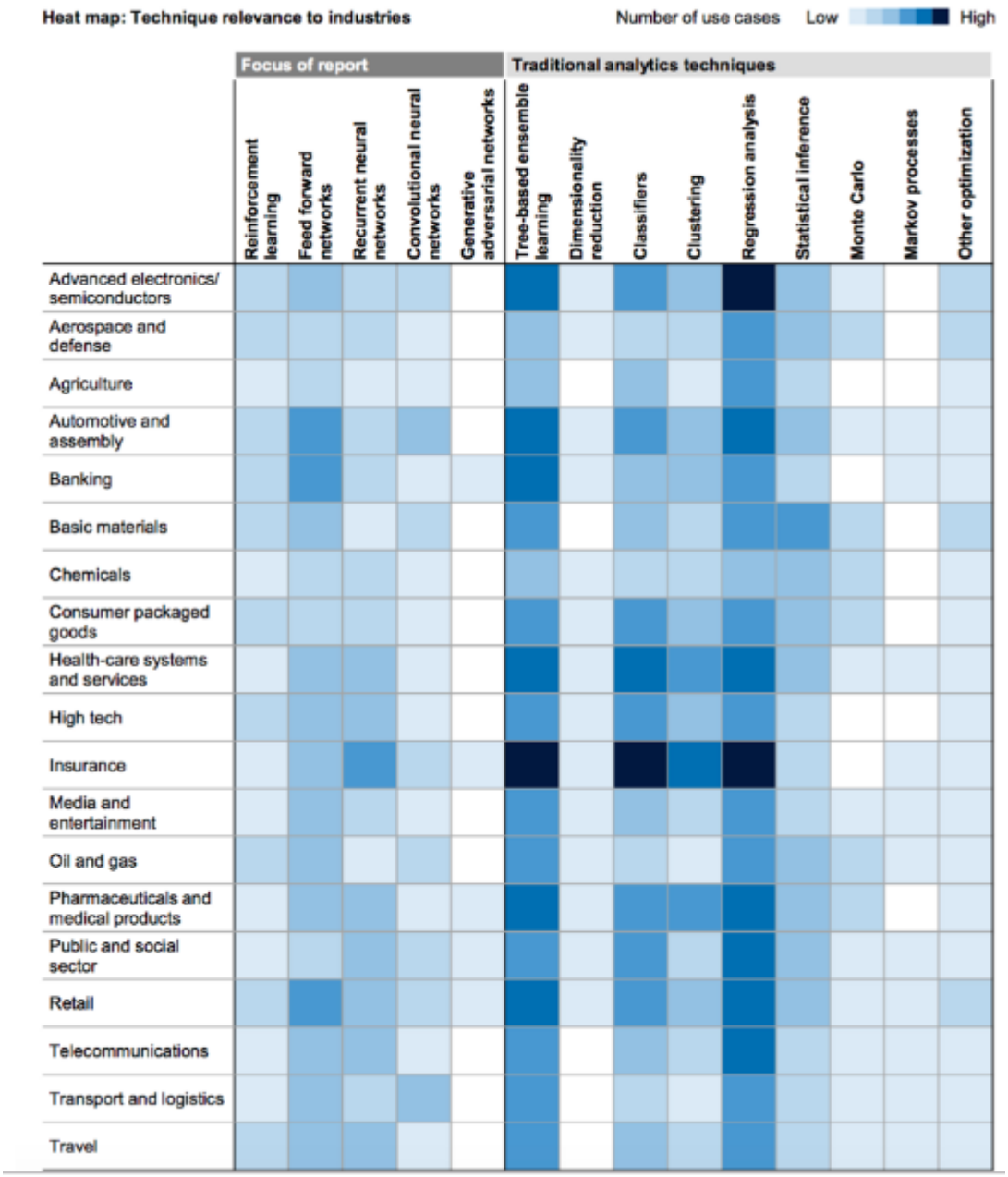
- Les langages : Comme son nom l'indique, ce sont les différents moyens de communication qu'utilisent les ingénieurs pour transmettre des ordres aux machines et d'apporter un sens à cette donnée. Les plus connus en data science sont : R (orienté statistique) et Python (orienté développement web).
- Les frameworks: Ce sont les cadres mis en place par certaines sociétés notamment les GAFA ensuite mis à disposition des développeurs en open source. Ils permettent l'utilisation de certaines librairies d'algorithmes ou de structures pré faites afin de faciliter le travail des développeurs. Les plus utilisés sont Tensor Flow de Google,Theano, Keras..
- Hardware: Ce sont les outils utilisés soit pour récoler de la donnée (capteurs Iot ou encore Lidar, Radar, Caméra etc.… dans une voiture autonome) soit pour permettre de la traiter et exécuter les calculs pour l'apprentissage (GPU de 'Nvidia de Qualcomm ou d'Intel…)

# 2. Les technologies et les cas d’utilisations

Parmi toutes les catégories de machine learning utilisées tels que le deep learning, le reinforcement learning ou encore le transfer learning, il est souvent difficile lorsqu’on regarde les utilisations du côté application de faire la différence entre tous les modèles et surtout de savoir quels sont ceux les plus utilisés et pour quels cas d’utilisations concrètement.

Le rapport de McKinsey sur l’utilisation du machine learning en entreprise, nous permet de mieux appréhender l’utilisation de ces modèles et surtout de comprendre dans quelles verticales ils sont utilisés. On peut ensuite déduire les cas d’utilisations récurrents pour chaque secteur.

On remarque par exemple que c’est dans l’assurance que le machine learning est actuellement le plus utilisé et qu’en terme de modèle c’est la régression linéaire qui est encore le modèle le plus exploité.




# 3. Les orchestrateurs

- Les outils d'acquisition de données : Ce sont des plateformes ou outils qui permettent de récolter de la donnée rendue disponible sous forme d'API ou de base de données. Parmi ces outils, on peut notamment citer PhantomBuster, une startup française à fort potentiel qui crée une infrastructure permettant à des personnes non techniques, d'extraire ces données en quelques clics.
- Les infrastructures: Ce sont les créateurs d'environnements dans lesquels évolue l'apprentissage : Google ML, Facebook fb Learner, Michaelangelo d'Uber, Azure ML de Microsoft, Cognitive Scale, Skymind etc... Ces derniers fournissent des plateformes pour développeurs sur l'intégralité de la chaîne sans qu'ils n'aient à connecter plusieurs infrastructures de la base de données aux frameworks évoqués plus haut.

- Les plateformes clés en main de Datascience: Ce sont les outils qui permettent la création de modèles de machine learning sur mesures et exportables. La plus performante et connue étant Datarobot.


# 4. Les produits

Les produits sont délivrés sous formes de plateformes Saas ou d'API/SDK et ont pour chacun d'entre eux une approche verticale (avec une spécialisation sur l'industrie) ou bien horizontale (avec une spécialisation sur la technologie utilisée). D'un point de vue purement technique, les deux approches sont valables en terme de données. L'un analyse tout type de données mais ciblées sur un marché précis, le deuxième analyse un type de donnée sur tout type de marchés. La plupart du temps, les plateformes offrent aussi via une API leur service soit pour accélérer la distribution de leur service actuel soit pour diversifier leur business model.


---

Article écrit par Eytan Messika