

Big Data

Réalisée par Claire Tertrais

26 février 2023

Contents

I. <i>Big Data</i> , que cela signifie-t-il ?	1
II. <i>Big data</i> et agriculture	2
2.1 Les capteurs	2
2.2 Les drones	3
III. <i>Big data</i> et Marketing	3
Bibliographie	4

I. *Big Data*, que cela signifie-t-il ?

Les données massives et l'intelligence artificielle sont des termes de plus en plus entendus et qui concernent de nombreux domaines, sans pour autant être bien compris et parfois difficile à définir. Le mot *big data*, popularisé par John Mashey dans les années 2000, “fait référence aux bases de données trop grandes et complexes pour être étudiées avec les méthodes statistiques traditionnelles – et, par extension, à tous les nouveaux outils d’analyse de ces données” (Thomas Bourany 2018). Concrètement, suite à la progression exponentielle des données et leur hétérogénéité, les *big data* permettent de contourner les limitations des bases de données classiques et offrent à chacun l’accessibilité réel à des bases de données géantes. (BI2B.eu 2019)

Initialement, une liste de 3V permettaient d’analyser cette tendance (Douglas Laney, 2001) mais elle se résume désormais avec 5V : (Thomas Bourany 2018)

- **Volume** : La masse de volume de données créées dans le monde grandit à une allure exponentielle et se mesure en zettabits, soit un milliard de fois plus grand que le téraoctet ;
- **Vitesse** : La vitesse de création, de collecte et de transmission des données se fait de l’ordre de la zeptoseconde. Par exemple en marketing, le choix des publicités se fait sur le modèle d’enchère : les données de l’utilisateur sont analysées par des algorithmes avec une moyenne de 10 millions de prédictions publicitaire par seconde, ce qui permet de choisir la pub en moins de 10 millisecondes ;
- **Variété** : Les *big data* sont utilisées dans des domaines d’applications très divers, sous forme structurées (tableaux) ou non structurées (80% des données, images, textes bruts ...). Les *big data* sont souvent considérées comme une opportunité pour les entreprises et les gouvernements, car ils peuvent être analysés pour extraire des informations précieuses sur les comportements, les tendances et les préférences des consommateurs ;

- **Valeur** : En 2009, l'ensemble des projets liés aux *big data* représentait un profit de 100 millions d'euros. Les revenus n'ont fait qu'augmenter et ont atteint en 2018 une valeur de 42 milliards d'euros sur le marché mondial. Les prédictions évaluent les chiffres à plus de 100 milliards d'euros d'ici 2030 ; (Tristan Gaudiaut 2021)
- **Véracité** : Assez paradoxalement, il y a un manque d'information sur l'origine des données ce qui crée de nombreuses polémiques autour des « infox ». Sont-elles pertinentes ? Sont-elles réelles ? En 2013, un faux tweet publié sur le compte de l'Associated Press indiquant que la Maison Blanche avait fait l'objet de 2 explosions a semé la panique à Wall Street. En moins de quelques secondes, le Dow Jones américain a perdu près de 150 points, ce qui représentait à l'époque une perte de capitalisation boursière équivalente à 150 milliards de dollars ; (Guillaume Maujea 2013)

Comme dit précédemment, les *big data* sont utilisées dans des domaines très divers. Certains d'entre eux vont vous être présentés dans la suite de ce document.

II. *Big data* et agriculture

Il ne fait plus aucun doute que le modèle actuel d'agriculture et de consommation alimentaire n'est pas durable. Mais comment continuer à alimenter plus de 8 milliards d'humains sur la planète Terre tout en diminuant les impacts négatifs engendrés ? Trouver une réponse à cela est bien complexe mais nous pouvons identifier certaines clés.

2.1 Les capteurs

L'agriculture évolue selon un modèle empirique depuis des siècles mais les données collectées n'étaient pas digitales. Les *big data* se différencient en terme de volume et de leur potentiel analytique. L'observation et la connaissance de l'état des facteurs majeurs (sol, humidité...) sont des dimensions importantes en agriculture et font parties intégrantes du processus de décision. Les capteurs apportent une solution avec des mesures objectives et fréquentes (Jean-Pierre Chanet 2016). Les applications potentielles des technologies sans fil en agriculture couvrent un large éventail de scénarios et d'applications.

On peut retrouver :

- Les réseaux pour la collecte de données scalaires (température, humidité) : les capteurs permettent l'acquisition de données élémentaires pour l'aide à la décision ou le pilotage d'actionneurs (irrigation, pilotage de bâtiments, météo) ;
- Les réseaux pour la collecte de données multimédia (images, vidéo) : en agriculture ces réseaux peuvent être mis en œuvre pour la détection de maladie ou de ravageur et la surveillance troupeaux ;
- Les réseaux avec des nœuds mobiles ou avec une connectivité discontinue : ces réseaux regroupent les applications utilisant un véhicule (drone, tracteur) pour la collecte de données ou les WSN nécessitant d'éteindre les nœuds pour augmenter la durée de vie du réseau ;
- Les systèmes basés sur les puces RFID (Radio Frequency IDentification) : au-delà de leur utilisation pour l'identification et la traçabilité, les technologies RFID permettent aussi de faire de la collecte de données ;
- Les applications basées sur la connexion à un smartphone : la généralisation des smartphones a permis le déploiement de nombreux objets connectés pouvant offrir des réponses aux besoins de l'agriculture.

Des chercheurs de chez Arvalis, Cap 2020, IRSTEA et Acolliance ont travaillé sur un projet nommé CROCUS, qui avait pour objectif de développer des Réseaux de Capteurs Sans Fil. Ces RCSF sont une des technologies les plus prometteuses car ils sont complémentaires des solutions existantes (satellites, capteurs embarqués, stations météorologiques), qui présentent des résolutions spatiales ou temporelles trop faibles. Les sondes des capteurs permettent un suivi quotidien, puis via une passerelle de communication, les données sont récoltées dans un serveur. Des algorithmes transforment les données brutes en variables agronomiques telles l'indice foliaire, la teneur en chlorophylle, l'état hydrique, la température du sol etc... qui sont ainsi mises à disposition des utilisateurs. Un diagnostic fiable de l'état et du fonctionnement de la culture est alors construit à partir de l'exploitation d'une quantité énorme de données, les *big data*. (EcophytoPIC 2011)

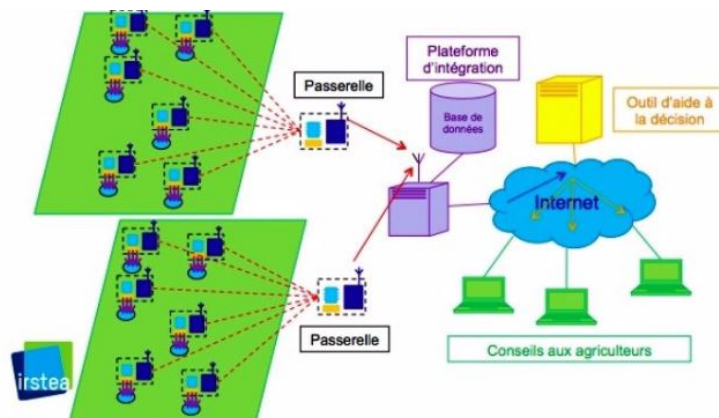


Figure 1: Circulations des données

Aux côtés de ces réseaux de capteurs sans fil, d'autres technologies prometteuses sont étudiées. Pour obtenir des images, il faut se tourner vers les drones.

2.2 Les drones

Une équipe de chercheurs autrichiens travaille depuis environ 10 ans sur l'utilisation des drones en agriculture, notamment pour anticiper et limiter le développement des adventices sur les cultures. Le drone prend en photo les champs puis transmet les images à un serveur en temps réels à l'aide de la 5G. A partir de ces nombreuses données, une application sur le serveur crée une carte à haute résolution qui est ensuite utilisée pour identifier l'adventice. Dans le cas de ces chercheurs, ils se sont focalisés sur le rumex. Cette plante, de nature très robuste et invasive, ne présente aucun intérêt fourrager. L'identification de l'adventice se fait grâce à l'apprentissage automatique : une fois détectée et localisée, les données sont retransmises sur le terrain pour que le rumex soit traité directement. Un robot agricole, tiré par un petit tracteur, reçoit les coordonnées de localisation. Il peut alors éliminer "écologiquement" et efficacement la plante par pulvérisation ciblée. Cette technique de "spot spraying" permet de traiter les plantes individuellement et réduit jusqu'à 90% l'utilisation de pesticides. (Georg Roman Schneider, Josef Scharinger, and Claudia Probst 2023)

III. *Big data* et Marketing

Quel est le produit le plus adapté pour ce marché spécifique ? Comment promouvoir un tel produit dans ce marché ? A travers quels canaux de communication ? A quel moment et à quel prix ? Quels genre de promotion ? Toutes ces questions sont les éléments clés d'une bonne stratégie marketing. C'est donc, sans surprise, que ce domaine est pionnier dans l'utilisation des *big data*.

Ayant pour objectif d'exploiter les *big data*, des chercheurs ont développé des nouvelles techniques pour capturer, traiter, analyser et visualiser de grandes quantités de données en un temps limité. Ces techniques

impliquent diverses disciplines : mathématiques, statistiques, méthodes d’optimisation, traitement d’un signal, apprentissage automatique, visualisation et analyse des réseaux sociaux.(Paulo Cortez and Sérgio Moro 2018)

Ces données peuvent permettre d’anticiper le comportement d’un consommateur ou l’évolution d’une tendance. Prenons l’exemple de Target (Hill 2012), une grande chaîne de magasin américaine. L’enseigne assigne un “guest ID number” à chaque client, relié à la carte de crédit utilisée, à leur nom ou à leur adresse mail. Cela crée une base de donnée commune entre tous les magasins qui garde en mémoire tous les achats.

Dans les années 2010, Target développe une nouvelle méthode d’analyse pour identifier les femmes en début de grossesse. Une analyste a remarqué que les abonnées du programme Target Baby achetaient une quantité plus importante que la normale de crème hydratante pour le corps en début du 2nd trimestre, et particulièrement sans parfum. Un autre analyste a noté que dans les 20 premières semaines, les femmes enceintes consommaient davantage de compléments alimentaires tels que le calcium, le magnésium ou le zinc. De plus, il a été capable de mettre en place, à partir de 25 produits consommés dans un temps rapproché, un score de “pregnancy prediction” à chaque client. Il est également possible d’estimer le terme d’une grossesse en fonction des achats de sorte à ce que Target envoie des coupons de réduction spécifiques à l’avancée de la grossesse.

En 2012, un scandale a éclaté lorsque l’enseigne a envoyé des réductions sur les habits de nouveaux nés à une adolescente. Le père, scandalisé par le fait que Target envoie ce genre de pub spécifiquement à sa fille, est allé se plaindre au magasin. Il s’est avéré que sa fille était bien enceinte mais qu’elle n’en avait pas informé sa famille.

Target a rapidement compris que les gens étaient effrayés que l’entreprise soit au courant de leur grossesse à l’avance. Une nouvelle stratégie marketing a alors été adoptée : au lieu d’envoyer directement des coupons sur des articles de grossesses, ce sont désormais des publicités d’apparence quelconque et non personnalisées que les clients reçoivent. En réalité, Target positionne des offres pour les couches, poussettes etc... entre d’autres articles dont ils savent pertinemment que cela n’intéressera pas la femme enceinte (tondeuse, verre de vin...). Ainsi, tant que la personne ne se sent pas espionnée, elle utilisera les coupons.

L’analyse des données sur les consommateurs et leurs habitudes s’avèrent bien plus courante et utile que ce que le grand public s’imagine. Il existe des milliers d’exemples comme celui de Target où une publicité qui nous semble anodine se révèle plus que ciblée.

Bibliographie

Cette synthèse a été réalisée à partir des éléments ci-dessous :

BI2B.eu. 2019. “Big Data.” *BI2B*. <https://www.bi2b.eu/nos-domaines/big-data/>.

EcophytoPIC. 2011. “Projet CROCUS.” <https://ecophytopic.fr/recherche-innovation/piloter/projet-crocus>.

Georg Roman Schneider, Josef Scharinger, and Claudia Probst. 2023. “Weed Detection in Grassland and Field Areas Employing RGB Imagery with a Deep Learning Algorithm Using Rumex Obtusifolius Plants as a Case Study.” *Engineering Proceedings* 27 (1): 87. <https://doi.org/10.3390/ecsa-9-13950>.

Guillaume Maujea. 2013. “Comment Un Faux Tweet a Fait Plonger Wall Street.” *Les Echos*. <https://www.lesechos.fr/2013/04/comment-un-faux-tweet-a-fait-plonger-wall-street-338456>.

Hill, Kashmir. 2012. “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.” *Forbes*. <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>.

Jean-Pierre Chanet. 2016. “Réseaux de Capteurs Sans Fil Et Approches Contextuelles Au Service de l’agriculture.” HDR Ecole doctorale Sciences pour l’ingénieur de Clermont-Ferrand.

Paulo Cortez, and Sérgio Moro. 2018. “Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis.” *European Research on Management and Business Economics* 24 (1): 1–7. <https://doi.org/10.1016/j.iedeen.2017.06.002>.

Thomas Bourany. 2018. “Les 5V Du Big Data.” *Regards Croisés Sur l’économie* 23 (2): 27–31. <https://doi.org/10.3917/rce.023.0027>.

Tristan Gaudiaut. 2021. “Infographie: Le Big Bang Du Big Data.” *Statista Infographies*. <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde>.