

Entreprise : comment maîtriser l'usage du Big Data

Les technologies Big Data se multiplient, les algorithmes prédictifs sont de plus en plus puissants et la responsabilité entourant l'exploitation de la donnée pèse lourd sur les épaules des entreprises. Quelle démarche adopter afin de ne pas franchir des limites incertaines tout en maintenant son processus d'innovation ?

Le quotidien Big Data en entreprise - La formule 1 de la BI

Les Big Data (comprendre ici les technologies de traitement des gros volumes de données) sont avant tout une boîte à outils destinée à accompagner une entreprise dans sa prise de décision, à affiner la perception qu'elle a de ses clients, de leurs choix et comportements d'achat et à disposer d'informations essentielles pour mieux marketer et positionner ses produits et services. Les solutions Big Data répondent au besoin d'une plus grande interactivité et l'accès à une meilleure compréhension de la donnée disponible. En pratique, l'expérience montre que les demandes Big Data portent principalement d'une part sur les données internes de l'entreprise, et d'autre part, sur les difficultés de traitement de la donnée. Ces difficultés, depuis longtemps identifiées par les métiers en charge de la BI, les directions marketing, les chefs de produits, relèvent de problématiques de chargement, de compilation et de restitution des données ou sont liées à la recherche d'une granularité plus fine des reportings.

Les solutions Big Data ont permis de répondre à ce besoin en offrant une plus grande réactivité dans le traitement et l'intégration de la donnée, ainsi qu'une perception affinée des process et du marché. Elles ne règlent pour autant pas la question des données personnelles ou confidentielles ou nominatives, dont l'exploitation, par l'effet des algorithmes, peut aboutir à des résultats surprenants. L'ingénierie algorithmique repose sur des choix humains, dont les conséquences seront la résultante des Big Data de demain.

La définition de son besoin Big Data - D'une contrainte technique à la préservation de la vie privée

Les modèles d'apprentissage algorithmique, (ou Machine Learning) reposent sur la soumission d'un contexte et d'un certain nombre de probabilités en fonction desquels la machine répondra avec plus ou moins de pertinence à une question spécifique, en fonction des données fournies. Le caractère prédictif des technologies Big Data est guidé par la recherche précise d'une réponse à une question donnée. Une autre méthode, voire un autre discours, consiste à s'appuyer sur des algorithmes auto-apprenants, susceptibles d'intervenir quand l'entreprise ne sait pas précisément ce qu'elle cherche. L'objectif ici relève de l'expérimentation, voire la validation d'hypothèses, qui, par l'effet des co-occurrences découvertes à cette occasion, permettront de dégager de nouvelles pistes de réflexion.

Cette démarche est principalement adoptée par les entreprises disposant d'équipes digitales ou innovation assez mûres sur l'usage des technologies Big Data pour projeter dans ces corrélations

de nouveaux use cases. Le principal écueil de la méthode repose sur l'absence de maîtrise de la donnée utilisée et du but à atteindre. Dans cette recherche de corrélation, la machine fournira un résultat qui peut ne correspondre ni à la stratégie de l'entreprise, ni à ses valeurs, ni aux règles en vigueur au regard de la protection des données à caractère confidentiel.

Depuis juillet 2015, les entreprises sont fortement incitées par la CNIL à identifier les risques engendrés par leurs traitements. La démarche d'étude d'impact débute par une étude du contexte, visant à décrire les traitements et leurs enjeux. Dans le cas d'une recherche générale de corrélations, l'impact sera évidemment très variable selon que l'entreprise intervient par exemple dans la métallurgie, cherchant à améliorer ses processus de production ou intervient dans le domaine bancaire, souhaitant disposer de modèles élaborés pour appuyer, par exemple, ses décisions d'octroi de prêt.

Dans les deux cas les données internes sont utilisées. Dans les deux cas, la recherche repose sur une demande métier d'amélioration des processus. Mais dans un seul de ces cas, a priori, l'impact décisionnel peut avoir des conséquences à large spectre, dont la maîtrise n'est pas garantie.

L'approche par la définition du contexte et des enjeux via la formulation de la question et la définition du modèle de réponse, offrent en revanche le cadre exigé pour la réalisation d'une étude d'impact sur la vie privée (le cas échéant), puisque le contexte est contrôlé dès l'origine par l'entreprise. C'est en outre une démarche adaptée aux entreprises débutant dans le Big Data. Sécurisant, le processus s'appuie en général sur les compétences d'un intégrateur qui assurera le transfert de compétence. Egalement plus pragmatique, au fort potentiel de croissance, elle permet de trouver rapidement de nouveaux axes de travail en réponse aux besoins métiers identifiés de longue date.

Quel avenir donner aux Big Data ? Une responsabilité entre les mains des entreprises

Pour être exploitables, les résultats doivent avant tout s'appuyer sur une donnée de qualité, qu'il aura fallu correctement qualifier. A ce titre, les technologies Big Data ne changent pas la façon dont les entreprises abordent leurs différentes problématiques mais le panel de solutions disponibles est autrement plus important. Les enjeux se sont déjà déplacés, tant les progrès technologiques ont dépassé le cadre actuel de l'usage et de la protection des données personnelles. Des algorithmes à visée prédictive très puissants existent déjà. Il s'agit plutôt de s'interroger sur ce que feront les entreprises des données dont elles disposent. Ayant en leur possession des informations très diverses, nominatives et confidentielles ou non, techniquement aujourd'hui, rien n'empêche leur exploitation et seule la réflexion en amont sera gage de pratiques éthiques et rationnelles.

Un des garde-fous potentiels de demain sera peut-être le contrôle par l'individu lui-même de ses données. En attendant de lui en donner les moyens, les entreprises sont finalement les seules en charge de cette protection. Concrètement, elles n'en seront capables qu'en avançant par étape, en commençant par s'interroger sur la qualité des données dont elles disposent mais également sur leur propriété. La donnée est chose précieuse. C'est aux entreprises qu'il appartient de démontrer qu'elles comprennent l'usage qu'elles peuvent en faire et surtout, leur capacité à en fixer les limites.

Elles ne maîtriseront la véritable signification de l'exploitation des données numériques qu'en naviguant suffisamment près des frontières. Il y a nécessairement une gestion du paradoxe à effectuer. Progresser pour ne pas se laisser distancer par ses concurrents tout en maintenant sous surveillance les limites à ne pas franchir, c'est un exercice de haute voltige. C'est pourquoi la banalisation des technologies Big Data ne doit pas intervenir trop tôt, n'en déplaisent aux pourvoyeurs de solutions tout-en-un.