

Data Scientist : de quoi parle-t-on ?

Le métier de la Data Science (science des données) est un métier interdisciplinaire qui se fait de plus en plus connaître ces dernières années, mais sans en vraiment comprendre le sens. Data Scientist (scientifique des données) est d'ailleurs souvent confondu avec Data Analyst (analyste de données) et même certains professionnels du domaine parviennent difficilement à expliquer leur métier. Auparavant ces professionnels étaient même plutôt appelés des statisticiens !

Mais alors quelle est la différence et en quoi consiste le métier de Data Scientist ? Là où un Data Analyst ira interpréter les données pour en faire des rapports pour répondre à une question factuelle, un Data Scientist ira quant à lui plus en profondeur sur une question de prévision pour de l'aide à la décision. Pour mieux illustrer le propos, voici un exemple de problématique par métier :

- Data Analyst : quelle a été la charge de travail du personnel tout au long des 12 derniers mois et quelles ont été les causes des éventuelles surcharges ?
- Data Scientist : à partir de la charge de travail qu'a eue le personnel les 12 derniers mois et des causes des éventuelles surcharges, quelle sera cette charge pour le mois à venir et quelles pourraient être les causes d'une potentielle surcharge ?

Ici l'analyste va vous donner une interprétation de ce qu'il aura découvert et éventuellement une piste d'amélioration même s'il n'est pas certain que ce qu'il a découvert se reproduise, tandis que le scientifique va vous donner des prévisions de ce qu'il pourrait se produire et pourquoi, pour vous aider à l'anticiper.

Cependant un Data Scientist n'est pas devin ! Pour parvenir à réaliser ses prévisions, il doit d'abord passer par plusieurs étapes :

1. L'analyse de la problématique et des données : oui, c'est une tâche qu'il a en commun avec l'analyste. A travers cette analyse il cherche à comprendre les données, comment les corriger lorsqu'elles présentent des incohérences ou des défauts, puis il cherche à savoir s'il existe des liens entre certaines données et ce qu'il souhaite prédire. Il peut également y ajouter des données externes, comme la météo, qui pourraient lui sembler pertinentes pour ses prévisions.
2. L'analyse du métier qu'il étudie : là ce n'est pas son propre métier qu'il analyse mais celui de son interlocuteur. Afin de dégager des schémas pertinents et ne pas tomber dans le contre-productif, un Data Scientist a besoin de savoir un minimum comment vous travaillez et comment fonctionne votre structure. Par exemple, il pourrait vouloir comprendre quelles sont les possibilités d'un parcours patient lorsqu'il arrive aux urgences. Comprendre le métier et ses outils peut également l'aider à comprendre pourquoi parfois certaines données sont manquantes ou du moins incomplètes.
3. La transformation des données : une fois toutes les informations réunies, il va transformer les données afin qu'elles soient utilisées correctement et produisent le moins d'erreur possible dans l'algorithme de prévision.
4. La création de l'algorithme : l'étape la plus importante pour vous, c'est ici que la magie des mathématiques opère. A l'aide de plusieurs outils possibles, le Data Scientist va utiliser ou créer plusieurs modèles mathématiques et les tester les uns après les autres afin d'obtenir le meilleur résultat possible. Cette étape peut être longue puisque parfois il peut être amené à réadapter encore et encore ses paramètres selon les résultats obtenus lors des tests, ou bien même selon des recherches que lui-même ou d'autres auront menées en parallèle pour

s'adapter aux progrès technologiques, comme en intelligence artificielle par exemple. Cette étape peut même aboutir à compléter les étapes précédentes au besoin.

Une fois tout ce processus complété, le Data Scientist pourra finalement vous présenter ses meilleurs résultats. Ils ne seront pas toujours parfaits mais ils colleront au plus près de la réalité et vous donneront souvent une excellente piste pour faciliter les décisions !

Et l'éthique du Data Scientist dans tout ça ? Elle englobe un ensemble de principes et de valeurs qui guident sa conduite dans la collecte, le traitement, l'analyse et l'utilisation des données. En raison de la nature souvent sensible des données et de l'impact potentiel sur la vie privée, l'éthique revêt une importance cruciale dans le domaine de la Data Science. Voici quelques principes éthiques auxquels un professionnel est généralement tenu de se conformer :

- **Respect de la vie privée** : il doit prendre des mesures pour protéger la vie privée des individus dont les données sont utilisées dans leurs analyses. Cela inclut l'anonymisation des données lorsque cela est possible et le respect des réglementations en vigueur, telles que le Règlement Général sur la Protection des Données (RGPD) en Europe.
- **Transparence** : il doit communiquer de manière transparente sur la manière dont les données sont collectées, traitées et utilisées. Cela implique de documenter les méthodes utilisées, de partager les sources de données, et d'expliquer les choix et les limites du modèle.
- **Consentement éclairé** : Lorsque cela est applicable, obtenir le consentement éclairé des individus dont les données sont utilisées est une pratique éthique. Les personnes devraient être informées de la manière dont leurs données seront utilisées et avoir la possibilité de donner ou de refuser leur consentement.
- **Équité** : le Data Scientist doit s'efforcer de garantir l'équité dans ses analyses. Cela signifie éviter les biais dans les données et les modèles, en s'assurant que les résultats ne discriminent pas certains groupes de personnes.
- **Sécurité des données** : la protection des données contre l'accès non autorisé est une préoccupation majeure et c'est pourquoi il se doit de mettre en place des mesures de sécurité robustes pour prévenir les atteintes à la sécurité et garantir la confidentialité des données.
- **Responsabilité** : le Data Scientist est responsable des conséquences de ses résultats. Il doit prendre en compte les implications éthiques de ses travaux et être prêt à assumer la responsabilité des résultats, en particulier s'ils ont un impact sur les individus ou la société.
- **Utilité sociale** : il est encouragé à utiliser leurs compétences pour le bien de la société. Cela peut impliquer de contribuer à des projets qui ont un impact positif sur la santé, l'éducation, l'environnement, etc...

Il est important de noter que les principes éthiques peuvent varier en fonction du contexte, de l'industrie et des réglementations locales. Les experts en Data Science sont souvent appelés à rester informés sur les meilleures pratiques et à s'adapter aux évolutions dans le domaine de l'éthique des données.