

Big Data et prise de décision : que faire de toutes ces informations que les organisations produisent ?

23 février 2017

Avec la participation de :

Clotilde Coron, Docteure en Sciences de gestion et chef de projet Big Data & RH dans la Direction Expérience Salarié d'Orange

Claudio Vitari, Maître de conférences à l'IAE de Paris et membre de la Chaire Mutations - Anticipations - Innovations.

Clotilde Coron

Les 20 dernières années ont connu beaucoup d'évolutions technologiques autour d'Internet, des réseaux sociaux, de la téléphonie mobile conduisant à un essor très important du volume de données, à la fois structurées et non structurées. En parallèle, les évolutions technologiques en termes de serveur, de capacités de calcul ont permis d'innover dans les modalités de traitement de ces données et donc de pouvoir incorporer un plus grand volume de données et aussi de pouvoir incorporer des données non structurées.

La définition littérale du Big Data c'est d'avoir un volume important de données structurées ou non structurées. On parle aussi de Data Science. Si l'on dit de la Data Science qu'elle représente tous les traitements statistiques qui nécessitent d'autres outils que des tableurs Excel traditionnels, on peut considérer que le Big Data serait une branche de la Data Science avec quelques spécificités, notamment le gros volume de données qui nécessite des infrastructures spécifiques et le fait d'assimiler des données non structurées. On peut également considérer le Big Data à travers les quatre V :

- volume (gros volume de données),
- vitesse (données dynamiques qui se mettent à jour en temps réel),
- variété (données structurées et non structurées issues de sources variées),
- véracité (données très incomplètes).

Récemment un cinquième V a été ajouté qui donne une idée des bénéfices attendus par les entreprises, celui de la valeur (le Big Data devrait apporter de la valeur économique à l'entreprise).

En RH on a rarement des volumes de données nécessitant l'utilisation de serveurs externes (à part pour certains types de données assez rares). Quand on parle de Big Data RH c'est le fait de se rendre compte que les entreprises ont beaucoup de données sur les salariés et c'est le fait d'utiliser, de mobiliser, d'analyser ces données de façon plus approfondie et novatrice par rapport à ce qui était fait avant dans une optique à la fois d'analyse de compréhension de certains comportements ou phénomènes et dans une optique prédictive de certains comportements ou phénomènes.

Le Big Data en RH est né dans un certain contexte. C'est d'abord l'essor du marketing et ce qu'on appelle le marketing RH dont le principe est d'accorder aux salariés la même attention que l'on accorde aux clients et d'exporter des méthodes que l'on utilisait traditionnellement dans le marketing client à la vie dans l'entreprise, par exemple tout ce qui est personnalisation ou segmentation des services RH. On utilise les méthodes du marketing dans les ressources humaines. Les entreprises ont dû se constituer dans le marketing des bases de données clients qu'elles ne possédaient pas forcément auparavant alors qu'elles disposent déjà de beaucoup de données sur les salariés (structurées et non structurées). Auparavant, on ne pensait pas qu'on pouvait les utiliser. Par exemple le CV d'un ancien candidat qui est devenu salarié est une donnée non structurée, les postes publiés sur le réseau social interne sont des données non structurées.

L'autre élément de contexte est une forme de course à l'innovation. Faire du Big Data RH aujourd'hui est un élément de marque employeur, un élément de marque entreprise. De plus en plus d'enquêtes montrent que le Big Data est devenu une préoccupation majeure des DRH. C'est devenu un élément de communication et cela va de pair avec tout un marché se structurant autour du Big Data RH, avec d'un côté des cabinets de conseil qui proposent de plus en plus de services liés au Big Data RH et d'un autre côté des startups.

De manière prospective, le Big Data en ressources humaines permet d'appareiller des offres de postes et des CV en se fondant sur les mots utilisés par des algorithmes de matching. On peut utiliser la même chose pour la mobilité interne pour laquelle on peut imaginer faire des suggestions de postes aux salariés en utilisant un algorithme de matching. Sur les formations on peut également imaginer faire des recommandations personnalisées, analyser les profils des apprenants et les formats qui leur conviennent le mieux. On peut imaginer de prédire l'absentéisme, le turnover, etc. Ces éléments, dans les prochaines années, vont se développer.

Quand on met en œuvre opérationnellement un projet Big Data RH on se rend compte que tout ne va pas sans difficulté, ainsi l'expérimentation menée chez Orange pour faire des suggestions personnalisées de formation aux salariés en fonction de leur profil. Ce projet fait suite à une demande des salariés. Une nouvelle plateforme de formations en 2016 promettait d'offrir des suggestions personnalisées de formation sauf qu'Orange s'est rendu compte tardivement que l'éditeur avait survendu cet aspect. On a donc dû faire un projet en parallèle de cette plateforme. On a voulu fonctionner sur un principe de volontariat.

Les salariés devaient être volontaires pour participer à ce pilote. Comme on utilise des données personnelles, on veut avoir l'accord des salariés pour les utiliser. On a envoyé un mail à 10.000 salariés pour les informer, 1.700 salariés ont été volontaires.

L'étape essentielle a été celle de la sélection des données. On s'est confronté assez durement à la réalité. Au début, on avait beaucoup d'ambition sur ces recommandations de formation, on se disait qu'on pouvait le faire en fonction du projet professionnel, des souhaits de mobilité, des compétences sauf qu'on n'a pas de données sur ces éléments. Dans l'entretien individuel on peut indiquer un souhait de projet professionnel sauf que c'est très peu rempli par les salariés. Les salariés ont un espace interne où ils peuvent parler de leurs souhaits de mobilité mais c'est confidentiel donc on ne peut pas utiliser ces données. Le référentiel de compétences met en lien des postes et des compétences mais ce n'est pas une vision individuelle. On peut supposer qu'un salarié sur un poste n'a pas forcément toutes les compétences nécessaires surtout s'il vient d'arriver.

On aurait pu imaginer recommander des formations en fonction de ce différentiel entre les compétences réelles et les compétences nécessaires pour le poste et supposer qu'un salarié n'utilise pas toutes ces compétences dans son métier. Il aurait fallu avoir des données individuelles sur les compétences qu'Orange ne dispose pas. Lors de l'entretien individuel on peut remplir ces compétences mais c'est très peu rempli et très peu actualisé alors que l'on peut supposer que l'on acquiert ou que l'on perd des compétences régulièrement. Sur le réseau social interne, il y a la possibilité d'indiquer des compétences mais cette fonctionnalité est très peu utilisée.

On s'est donc confronté au fait qu'on n'avait pas les données nécessaires pour répondre à nos ambitions. Finalement, on a fait avec ce qu'on avait, c'est-à-dire l'historique de formation, certaines données RH et certaines données issues du réseau social.

On a commencé par faire du filtrage collaboratif, c'est ce qui est le plus utilisé par les moteurs de recommandation. L'idée c'est de dire que si A et B ont à peu près le même historique de formation et si A a suivi des formations en plus, on peut supposer que ces dernières vont intéresser B. C'est une idée très simple mais qui marche plutôt bien. Néanmoins cette démarche a permis de faire des recommandations à seulement 50 % des salariés du pilote.

On a été confronté à une autre difficulté : le catalogue de formation utilisé pour les recommandations change régulièrement. Quand on compare l'historique de formation avec les formations actuelles, il y en a plein qui ne sont plus disponibles. On ne pouvait pas faire de recommandations à un peu plus de la moitié des salariés dans le catalogue actuel. On a donc complété cette première approche par un filtrage thématique. Cette fois on se concentre sur les mots clés des formations suivies dans l'historique et des formations disponibles. Cela a permis de faire des recommandations à 200 ou 300 salariés en plus. Comme il restait encore pas mal de salariés qui n'avaient pas de recommandations, on a modélisé les variables les plus structurantes dans le fait de suivre la même formation qu'une autre personne.

Les variables les plus structurantes identifiées sont le niveau hiérarchique, le domaine métier, le fait d'être manager ou non, le fait d'appartenir au même groupe fermé du réseau social et le fait d'être en lien sur le réseau social. A partir de cette modélisation, on a pu rapprocher les personnes qui n'avaient pas de recommandations des personnes qui avaient eu des suggestions.

On a posté un premier jet de recommandations fin 2016. On a juste demandé des retours sur le réseau social interne pour que les gens puissent s'exprimer un peu. On est en train de poster le second jet. On voudrait ensuite faire un sondage pour évaluer la pertinence globale des recommandations et la pertinence de chaque approche prise séparément.

On se confronte très rapidement à la question de la qualité des données. Le Big Data fonctionne beaucoup par Test and Learn, c'est-à-dire que le premier algorithme est amélioré au fur et à mesure avec les retours des utilisateurs. Le principe est très itératif. On s'est rendu compte que les salariés qui participaient à l'expérimentation du pilote s'attendaient à un produit fini, à des recommandations qui leur correspondaient exactement, sans beaucoup de bienveillance à l'égard de recommandations qui pouvaient ne pas leur correspondre. Ils avaient en plus un très haut niveau d'exigence, d'attente. On peut supposer que c'est dû, en tant que client, à l'habitude d'avoir des algorithmes perfectionnés. On peut également supposer que les salariés d'une grande entreprise comme Orange ne sont pas forcément habitués à cette culture d'itération où l'on améliore peu à peu un produit. Orange est une organisation assez bureaucratique, très en silos, ce qui n'est pas vraiment adapté à ce type de projet qui nécessite une approche transversale.

On a lancé un projet d'expérimentation de présélection de CV via un algorithme, suite à une recherche qui a conduit à élaborer un algorithme qui apparie des offres de postes et de CV. Cet algorithme calcule des proximités entre les CV et les offres de postes. Il attribue un score de pertinence à chaque CV en fonction des mots contenus dans le CV et dans l'offre. Actuellement, cet algorithme semble assez performant. Dans 85 % des cas, il donne à peu près le même classement qu'un chargé de recrutement. Néanmoins, se pose aujourd'hui la question de l'utilisation et de l'usage de cet outil. Les chargés de recrutement sont obligés d'ouvrir tous les CV. Si un candidat se plaint de ne pas avoir été reçu, on ne peut pas lui dire que c'est l'algorithme qui l'a classé en dernier. Les chargés de recrutement continuent donc à ouvrir tous les CV. Ils ne peuvent que constater que l'outil donne quasiment un classement identique. Reste donc la question de l'utilisation de cet outil si juridiquement on doit ouvrir tous les CV.

Au-delà des bénéfices et des promesses liés au Big Data RH, il faut quand même souligner les limites actuelles liées aux données, à leur qualité et à leur organisation :

- les données RH chez Orange sont très fractionnées, morcelées entre différents SI RH,
- la culture de l'entreprise avec une organisation en silos,
- la culture Test and Learn qui n'est pas vraiment partagée par les salariés
- le fait de ne pas utiliser n'importe quelles données.

Il existe un niveau d'attente et un niveau de crainte disproportionnés par rapport à ce que l'on peut faire aujourd'hui en Big Data RH, liés à une méconnaissance du fonctionnement d'un algorithme, du Big Data RH et à une méconnaissance des données dont l'entreprise dispose.

Un enjeu paraît important, celui de la potentielle évolution des métiers RH liés au Big Data. Certains métiers peuvent être confrontés à de l'automatisation. Même si aujourd'hui les chargés de recrutement ouvrent tous les CV, il n'en sera peut-être pas de même dans cinq ou dix ans. Même si les outils Big Data ne se substituent pas du tout aujourd'hui à la prise de décision humaine chez Orange, ce ne sera peut-être pas la même chose dans dix ans. Néanmoins, cela peut amener les RH à se positionner sur des tâches à plus forte valeur ajoutée, à avoir une posture peut-être plus anticipatrice et moins une posture de reporting et une posture où l'on va utiliser les données de façon plus approfondie pour orienter de façon plus efficace les politiques RH.

Claudio Vitari

Les enjeux autour du Big Data sont de plusieurs dimensions. Il faut des solutions Big Data pour arriver à réduire les attaques qui ont lieu à différents endroits de la planète. Les entreprises, les groupes terroristes, les institutions, les espions, etc. utilisent d'énormes quantités de données pour essayer de casser d'autres systèmes d'information pour éventuellement détruire les données. Aujourd'hui, le Big Data peut être vu comme une guerre économique au niveau mondial mais on peut en faire autre chose. Quelles valeurs économiques peut-on lui donner sans forcément forcer les systèmes d'informations des autres entreprises pour obtenir encore plus de Big Data ?

A l'origine, les trois V (volume, véracité et variété) étaient suffisants pour décrire le Big Data. L'objectif était d'améliorer les décisions, être plus efficace, automatiser les processus. Aujourd'hui, certaines entreprises ont dépassé ces niveaux. Elles ont pour objectif d'offrir de nouveaux produits et de nouveaux services grâce au Big Data.

J'étudie cet aspect depuis plusieurs années. Ma thèse, terminée en 2006, portait sur les systèmes d'information en RH. Le terme de Big Data n'existait pas mais on essayait de voir comment recommander des experts au sein des entreprises à partir des données RH. On regardait les CV, les missions, les compétences, les formations qui étaient théoriquement recensées dans le SI RH afin de créer une sorte de service pages jaunes interentreprises. En cas de besoin d'un expert en comptabilité, en tapant les mots-clés sur un moteur de recherches interne, on espérait avoir un retour de trois experts en comptabilité que l'on pouvait contacter pour résoudre un problème ponctuel.

J'ai terminé en 2006 et les questions étaient déjà présentes. Beaucoup de données très hétérogènes qui évoluent, des compétences qui changent, le titre des postes également. Le système ne fonctionnait donc pas à l'époque. Les difficultés sont toujours du même genre. C'est une guerre économique, un marché. Comme il est important de pouvoir vendre, le terme de Big Data permet d'ouvrir et de créer de nouveaux marchés pour des solutions qui parfois existaient auparavant.

Les données telles que les CV, les postes, les compétences, les formations sont souvent des données déclarées officiellement et saisies par les ressources humaines. Ce sont donc des données générées par l'homme. Cette tendance est de moins en moins prégnante dans le Big Data parce que la plupart des données sont désormais créées par des machines (téléphones, capteurs, caméras, etc.). Elles sont de plus en plus importantes parce qu'on met de plus en plus de capteurs dans la domotique, les vêtements, les téléphones, les voitures. Ces données dépassent maintenant celles générées directement par l'homme.

Si traditionnellement les recherches s'appuyaient sur la méthode scientifique, aujourd'hui on laisse les ordinateurs trouver les corrélations, les relations entre des données diverses et variées. Les gros calculateurs traitent, analysent, comparent et classent les données. Il n'y a donc plus d'hypothèses de départ, plus de théorie. On attend que l'ordinateur nous révèle la réalité qui nous est cachée. Néanmoins, il faut faire attention de ne pas laisser une complète autonomie à la machine.

Si le Big Data sert à améliorer nos décisions, à comprendre la réalité des choses et à automatiser certaines tâches, aujourd'hui il sert de plus en plus à créer de la valeur, de nouveaux produits et de nouveaux services. Les phrases du genre « Si c'est gratuit, c'est vous le produit » deviennent de plus en plus vraies. Aucun mécène n'offre les services en ligne gratuits de manière philanthropique. Facebook c'est 2 milliards d'utilisateurs et 22 milliards de CA. Grosso modo un utilisateur est vendu 15 dollars par an. Nos données sont vendues à des agences publicitaires, des entreprises. En moyenne, un européen vaut 600 dollars par an sur le marché mondial du Big Data.

Tout ceci est-il légal ? Pour la plupart des entreprises, c'est légal, autorisé par des licences. L'utilisateur souscrit un contrat de licence. Facebook indique que son utilisation est gratuite, c'est vous le produit et on ne va pas vous faire payer. Dès que vous finalisez votre inscription, vous acceptez les conditions générales d'utilisation. Ce faisant, vous donnez votre consentement à l'utilisation de vos données. On pense que nos données sont anonymes mais l'anonymat est perdu par le croisement des données. Des recherches l'ont démontré à de nombreuses reprises.

Qui a le pouvoir de manipuler ce Big Data a le pouvoir de nous connaître même si on ne le souhaite pas. Les Google Glass ont fait scandale à l'époque avec leur application de reconnaissance faciale. Aujourd'hui, on passe aux lentilles connectées. Ainsi notre interlocuteur n'est pas dérangé par nos lunettes un peu encombrantes. On arrive à mettre toute la technologie nécessaire de reconnaissance faciale sur des lentilles connectées.

Nous avons interrogé 200 grandes et moyennes entreprises françaises au sujet de leur opinion sur le Big Data et de leur solution d'analyse du Big Data. On a posé des questions autour de la valeur transactionnelle (le Big Data baisse-t-il les coûts de transaction ?), la valeur stratégique (le Big Data facilite-t-il, améliore-t-il l'atteinte des objectifs stratégiques ?), la valeur transformationnelle (le Big Data permet-il de réfléchir à de nouveaux business models, à développer de nouvelles compétences ?) et la valeur informationnelle (le Big Data donne-t-il de meilleures informations pour une prise de décision ?). On a posé la question aux entreprises qui n'avaient pas de Big Data et aux entreprises qui l'avaient.

Au final, les entreprises ayant le Big Data trouvent qu'il génère de la valeur transactionnelle, stratégique, transformationnelle et informationnelle. Celles qui n'ont pas de Big Data considèrent que cela pourrait détruire de la valeur transactionnelle (le Big Data représente un coût mais on n'arrive pas à produire quelque chose pour abaisser les coûts de production ou les ventes). Elles imaginent que cela pourrait être bénéfique pour les trois autres types de valeur.

Comment les entreprises s'approprient-elles le Big Data ? Quels sont les modèles d'affaires qui peuvent en naître ? Nous avons fait une analyse sur 176 initiatives internationales en se focalisant sur les temps réels. Comment les entreprises arrivent-elles à offrir des solutions à partir de données en temps réel ? On a identifié cinq catégories de modèles d'affaires, des services qui peuvent naître à partir du Big Data :

- L'efficacité. L'entreprise TaKaDu identifie les fuites dans les circuits de distribution d'eau potable afin d'intervenir plus facilement et de réduire les pertes.
- Des services d'analyses de données. L'entreprise Act Value utilise des solutions pour comprendre les parcours des touristes dans les villes pour gérer le trafic, les musées, etc.
- Des services de génération de données. L'entreprise TomTom vend des données sur des questions de trafic, d'embouteillage.
- L'agrégation. L'entreprise Netvibes vend des données structurées venant de différents sites.
- Les nouveaux services. L'entreprise TripIt récupère les données des réservations d'hôtel, billets de train, d'avion, rendez-vous avec des fournisseurs, clients, etc. pour offrir en temps réel le planning de votre journée, l'agenda de votre voyage avec des alertes sur les rendez-vous, les éventuels retards d'avion, etc.

Ces services ne peuvent pas s'imaginer sans la disponibilité du Big Data en temps réel, en volume varié pour combiner toutes ces informations.

Nous réfléchissons à la question de la perception des gens et des entreprises autour des Big Data, autour du fait que nous n'avons plus la confidentialité des données, que nous perdons notre anonymat. Certains disent qu'on est espionné, d'autres que c'est tant mieux, qu'ils n'ont rien à cacher. Nous avons commencé à travailler sur la question des modèles d'affaires que les entreprises peuvent mettre en place à partir de leurs données. Quelles sont les compétences, les capacités organisationnelles nécessaires pour arriver à créer de la valeur ? Quels sont les processus organisationnels qui permettent de transformer en potentiel dans une réalité concrète un service, un produit ?

Comment le Big Data s'inscrit-il dans l'écologie en termes de relations entre partenaires mais également en termes d'impacts sur l'environnement ? Google à lui seul consomme l'équivalent de trois réacteurs nucléaires pour des services dits immatériels.

Florent Noël : Vous avez pointé le côté très anxiogène du Big Data, le côté effrayant. Que ce soit dans le champ de la consommation ou dans le champ de la prise de décision RH, le fait d'avoir des robots ne nous permet-il pas de garder des choses confidentielles ?

Si j'ai un projet professionnel et que j'aimerais bien avoir des formations pour pouvoir le suivre, il est peut-être plus facile de le déclarer à un robot qu'à son manager ou à son RRH.

Sur Amazon, on peut souhaiter préférer être conseillé en termes de lecture par un robot plutôt que d'aller voir son libraire.

Netflix a un objectif de maximisation de la consommation de formation et je ne suis pas sûr qu'Orange ait le même objectif. Comment fait-on pour articuler les besoins du salarié avec les besoins de l'entreprise ?

Clotilde Coron

Le but du projet n'est pas de maximiser la consommation de formation. On n'a pas pris des formations qui nécessitent une approbation managériale, d'où le choix des formations e-learning. L'objectif d'Orange est plutôt d'offrir un nouveau service aux salariés à la fois dans une optique d'amélioration du climat social et dans une optique d'amélioration de l'image employeur.

Sur la question de lier les besoins de l'entreprise à ceux des salariés, on est parti du plan de formation, de la politique de l'entreprise en matière de formation. On a inclus cette donnée dans les formations recommandées. Mais, au final, les formations recommandées ne sont pas souvent liées aux métiers et ne sont pas dans le plan de formation d'Orange.

Les salariés sont inquiétés par le fait que l'entreprise ait accès aux données. Un robot est de toute façon paramétré par l'être humain. Si le robot part d'une information, cela veut dire que cette information existe quelque part. C'est ce qui peut angoisser certains salariés, plus que l'idée qu'un être humain peut avoir l'information.

Claudio Vitari

Une fois qu'on a donné l'information au robot on ne la contrôle plus. Étant donné que le coût du stockage de l'information est aujourd'hui quasiment nul, on ne jette jamais les données. Un jour, on aura peut-être des algorithmes qui pourront traquer des cryptages. Une fois que la donnée est quelque part, tôt ou tard elle sera utilisée.

Je préfère un robot impartial et sans moralité. Quand je montre une vidéo à des étudiants sur YouTube, je préfère que les suggestions sur le bandeau de droite n'apparaissent pas. On interagit avec un robot et les données sont déjà utilisées ou le seront. On peut être associé à une personne ayant un comportement différent mais avec un profil identique.

Deux personnes ayant les mêmes compétences souhaitent peut-être avoir les mêmes formations, deux personnes ayant le même profil d'âge et de sexe veulent voir les mêmes films mais si les goûts sont différents cela peut être embêtant. Finalement, on est classé par catégorie, étiqueté.

Ce sont des techniques de marketing qui existent depuis des années. Que ce soit un robot ou une équipe marketing ne change rien.

Claudio Vitari

Cela ne change rien.

Florent Noël : Le fait que le robot analyse les données peut-il éviter les discriminations ?

Clotilde Coron

Tout dépend des paramètres. L'utilisation de la variable « genre » dans un algorithme pour recommander des formations ou pour recruter présente plusieurs possibilités. Il est possible de ne pas la mettre et l'algorithme se sert des données du passé pour prédire le futur mais il y aura un biais genré. On a aussi la possibilité d'inclure la variable « genre » de différentes façons. Soit pour contrôler ce biais genré, soit pour l'accentuer en disant que les femmes et les hommes ne vont pas accéder aux mêmes postes ou aux mêmes formations.

J'ai l'exemple d'un cabinet de conseil qui a conçu un modèle sur ses salariés dans le cadre de son recrutement. Ce modèle explique la réussite des salariés par différentes variables (diplômes, vie associative, etc.). Selon le cabinet, avoir ce modèle permet de prendre des décisions plus rationnelles, avec moins de discrimination. Ils se sont rendu compte que ce n'était pas le fait d'avoir un diplôme de telle école qui jouait mais d'avoir par exemple un double diplôme, ou que le fait d'avoir fait du secteur associatif prédisait mieux la performance que le diplôme. Cela enlève des biais liés au diplôme. En revanche pour le genre, comme ils souhaitent recruter plus de femmes, ils mettent un coefficient plus positif.

Il y a toujours des humains derrière les robots. Ce sont les humains qui les paramètrent. Les robots peuvent permettre d'éviter des discriminations si l'humain y a pensé en amont. Sur le projet de recommandation de formation, les juristes nous ont dit qu'on ne pouvait pas utiliser la variable « genre ». Mais actuellement on ne sait pas s'il y a un biais qui fait que les femmes et les hommes ne suivent pas les mêmes formations. Si on n'a pas la variable « genre », on ne peut pas savoir s'il y a un biais ou non, donc on ne peut pas le contrôler.

Florent Noël : Les données ne sont jamais jetées. Est-ce que cela ne va pas devenir un véritable enjeu ? Peut-on arriver à un moment où l'on ne pourra plus les stocker ?

Claudio Vitari

Notre planète est physiquement limitée. L'écologie nous l'enseigne. On pourrait avoir la planète remplie de disques durs et arriver à saturation. Des études ont démontré que l'on peut stocker des données dans notre ADN. Dans un brin d'ADN on peut mettre des milliards de gigabits de données. On serait donc capable de stocker toutes nos données dans notre corps. Mais aujourd'hui la question des limites de données n'est pas atteinte.

Florent Noël : J'ai fait partie des gens qui se sont intéressés à l'intelligence artificielle à la fin des années 80. Je me souviens avoir travaillé avec des entreprises sur ces sujets, notamment EDF. Il y a eu pendant une dizaine d'années une équipe qui a travaillé pour l'assistance au management via l'intelligence artificielle avec cette idée que comme les gens manageaient mal, grâce aux systèmes experts ils allaient enfin bien manager. Je me suis laissé aller pour comprendre les limites de ces outillages à développer un système expert d'orientation des carrières. Il y avait toujours une espèce d'incompréhension des apports réels de la technique. On attend souvent que la technique fasse des choses qu'on n'est pas capable de faire soi-même.

Ce mythe est toujours présent, réactualisé à chaque nouvelle vague technologique. Que sont devenus les systèmes experts aujourd'hui en RH ? Ils sont devenus l'actualisation des réglementaires de paye. Par rapport à la grande vogue de l'époque, on allait révolutionner la manière de gérer les ressources humaines sur les mêmes thèmes : l'offre de formation, l'adéquation homme/poste, la gestion de la mobilité, constituer une base de connaissances, de compétences requises. On est toujours dans des systèmes humains et dans la difficulté de produire l'information que l'on voudrait voir traiter par l'automate. Dans les applications RH, il n'y a aucune raison de se faire peur parce qu'elles ne produiront jamais ce qu'elles prétendent produire.

Clotilde Coron

Pour sortir de ce mythe, il suffirait d'ouvrir la boîte noire, de former les salariés et les RH sur le fonctionnement d'un algorithme et sur les statistiques pour qu'ils aient moins d'attente et moins de crainte.

Florent Noël : Pouvez-vous nous parler des problématiques d'anonymisation dans les Big Data ? A partir du moment où les données sont traitées par des robots, comment arrive-t-on à faire en sorte qu'il n'y ait pas de biais décisionnel du fait d'éventuel manquement d'anonymisation ?

Clotilde Coron

Quand c'est un service individualisé (on veut faire des recommandations de formation aux salariés) on ne peut pas anonymiser puisqu'on est obligé de retrouver le salarié après. Dans ce cas, on utilise seulement le matricule du salarié et on recoupe à la fin avec son nom. Du coup, c'est plutôt de la pseudonymisation.

Dans le cas d'analytics RH, le fait d'utiliser les données pour faire des études approfondies, c'est anonymisé.

Claudio Vitari

Les discriminations sont inscrites dans les algorithmes, les paramétrages, parfois de manière involontaire. Elles sont enregistrées dans les outils d'analyse par des paramétrages, des choix de couleurs, tout un tas d'éléments qu'on ne contrôle pas. Les couleurs d'un graphique peuvent avoir des valeurs symboliques pour certains, différentes pour d'autres. On n'y pense pas dans le choix mais cela peut avoir ensuite un impact pour vendre un produit ou pour recruter un candidat.

Même si ce n'est qu'un robot qui traite les données, la question des paramétrages est très importante. Quand Google a annoncé qu'il souhaitait paramétrer les résultats de la recherche pour filtrer les sites de qualité des sites qui n'étaient pas de qualité, on filtre des informations considérées comme fake news. Mais qui décide de ce qui est pertinent, de ce qui est véritable ? Est-ce un robot ? Non, c'est l'ingénieur informatique qui paramètre le robot.

Une autre question intéressante est celle de la neutralité du Net. Aujourd'hui, on parle d'améliorer le service mais qui va décider quelles informations doivent circuler ? Il y aura des robots mais qui va les paramétrer ? Les ingénieurs, dans leur bonne foi, cherchent la meilleure solution mais ils ne sont pas forcément formés aux questions d'éthique dans leur cursus. Quels sont les bagages nécessaires à la mission de réalisation d'un algorithme pour améliorer les résultats d'un moteur de recherche ? On ne les a pas. Le biais reste implicite.

Florent Noël : Quelles solutions d'affichage visuel du Big Data doit-on privilégier ?

Clotilde Coron

Sur tout ce qui graphique, il y a toujours des biais d'affichage. C'est à la personne qui fait le graphique de choisir celui qui va limiter les biais.

Claudio Vitari

Plein de boîtes vendent des solutions. Il en existe également des libres et gratuites.

Florent Noël : Les données des collaborateurs à l'extérieur de l'organisation ont-elles été utilisées ? D'un point de vue juridique, y a-t-il eu des discussions à ce propos ?

Quand on parle de robots qui vont traiter et proposer des choses, on pense tout de suite à la personne qui avait ce rôle et qui va disparaître. On se dit qu'on est dans une transformation numérique qui va quand même tuer un certain nombre d'emplois. Sur votre projet, est-ce que la conduite du changement a été abordée ?

Clotilde Coron

Concernant les données externes, aujourd'hui on ne le fait pas du tout pour des raisons juridiques. La seule chose que l'on pourrait faire c'est de demander aux salariés si l'on peut utiliser ces données mais il est relativement difficile d'appareiller des données d'Orange à des données de LinkedIn par exemple. Certaines entreprises créent une nouvelle plateforme où les salariés vont reproduire leur profil LinkedIn. Je n'ai jamais entendu parler d'entreprises qui utiliseraient les données LinkedIn en France.

Ce projet est un nouveau service qui n'est pas rendu aujourd'hui par des personnes humaines donc il n'y a pas de problématique de poste. Sur d'autres sujets, par exemple la présélection de CV par un algorithme, il y a une problématique. Il y a effectivement des craintes de la part des salariés et des syndicats. Orange a signé un accord sur la transformation numérique fléchissant les métiers qui peuvent être impactés par le digital, les nouveaux métiers et ceux qui risquent de disparaître et en réfléchissant sur les moyens d'accompagner les salariés concernés.

***Florent Noël** : Comment peut-on mieux anticiper l'évaluation de la création de valeur d'un projet ? L'objectif initial est-il le même que celui qui a été atteint ? Avez-vous transformé au fur et à mesure cet objectif ?*

Clotilde Coron

On a transformé en fonction de la réalité des données. Selon les résultats obtenus avec le sondage, on transforme le projet. On demande aux salariés s'ils trouvent cette fonctionnalité pertinente, ce qu'ils proposeraient d'autre. On essaie d'aller dans une culture où l'on va faire des choses mais on essaie d'avoir des retours de salariés améliorés.

***Florent Noël** : Existe-t-il des publications sur tout ce qui est recherche de business model, notamment dans le domaine de l'industrie ? Pensez-vous que l'on peut mettre de la créativité dans des grands groupes français ?*

Claudio Vitari

On peut mettre de la créativité. La visualisation est un instrument permettant de montrer des choses qu'on ne verrait pas autrement. Il existe des pistes de recherche autour de la datavisualisation. La visualisation peut concerner un objet particulier, un ensemble d'objets, un réseau social, un individu. Elle peut aider fortement la créativité.

***Florent Noël** : On parle maintenant de l'usine 4.0 dans l'industrie avec des capteurs dans des chaînes de production. Par exemple, Rio Tinto fait de l'aluminium avec des machines qui travaillent à 800 degrés. L'homme ne peut pas intervenir. Lorsque l'on s'aperçoit qu'il y a un problème en fin de chaîne, on a perdu le coût de fabrication et la matière première. Avec des capteurs qui détectent les pannes, on parle de maintenance prédictive, on arrive à diminuer ces problèmes et à ne pas fabriquer des choses inutiles et coûteuses.*

A titre de remarque, il faut bien distinguer causalité et corrélation. A partir de quelle quantité de données peut-on considérer que c'est du Big Data ? Peut-on créer des modèles qui vont permettre de prendre des décisions à partir de peu de données ?

Clotilde Coron

Avec le Big Data on passe d'un paradigme de la causalité à un paradigme de la corrélation. Ce qui comptait avant c'était de comprendre pourquoi le client allait faire telle ou telle chose. Aujourd'hui, si une courbe montre que les clients blonds achètent plus de céréales de telle marque, est-ce vraiment important de comprendre pourquoi si on arrive à vendre plus de céréales de telle marque ? Pour les entreprises, la réponse n'est pas toujours évidente. En revanche, en ressources humaines la question se pose de manière très différente. Les gens ont besoin de comprendre pourquoi on met le genre, le sexe dans un algorithme de formation alors qu'a priori il n'y a pas de lien.

En ressources humaines on a rarement un volume énorme de données, à part les données sur les mails. Sur 1.700 salariés ce n'est pas du Big Data au sens où l'on n'a pas de serveur de calcul supplémentaire. J'appelle cela du Big Data au sens où l'on va créer un nouveau service, on va personnaliser la relation avec le salarié. L'usage du terme Big Data est un peu faux mais on peut tirer un modèle de prise de décision à partir de 1.700 personnes. Tous ces modèles sont fondés sur des modèles statistiques ou économétriques. En économétrie on peut très bien avoir un modèle explicatif à 90 % même avec 200 personnes.

Claudio Vitari

La difficulté était de combiner la variété des données. Dans les trois V volume, vitesse et variété, c'est la variété qui complexifiait l'exercice plutôt que le volume. Une fois qu'on a un modèle pour 1.700 salariés, on peut le répliquer pour 10.000, 100.000, etc.