



**Data science - États-Unis - Données**

# La Kaggle mania gagnera-t-elle les assureurs français ?

■ Alors que les assureurs font de la transformation digitale l'une des priorités de leurs stratégies, quelques pionniers se lancent dans l'aventure des challenges de *data science*, via la plateforme nord-américaine Kaggle.

Le *big data* n'a pas de frontière et c'est tant mieux. Car, aujourd'hui, pour analyser ce flux massif de données que les assureurs s'organisent pour capter, mieux vaut savoir s'entourer des meilleurs *data scientists*. Ces responsables de la gestion et de l'analyse des données, capables d'apporter une nouvelle lumière sur les méthodes d'analyse prédictive et de modélisation de données, deviennent une fonction-clé pour qui entend parfaire sa démarche de transformation digitale. Mais c'est une compétence particulièrement recherchée. Quelques-uns l'ont bien compris et ont développé de véritables communautés de ces génies de la modélisation. Des communautés qui attirent aujourd'hui les assureurs, à l'instar de la plateforme nord-américaine Kaggle. Fondée en 2010, c'est la plus grande communauté de *data scientists* qui entrent en compétition pour résoudre des équations statistiques et des problèmes d'analyse prédic-

tive, de sciences, de mathématiques proposés par des entreprises ou des institutions. Une fois le problème transmis sur la plateforme, la communauté de plus de 480 000 inscrits travaille à trouver la solution. « *L'idée est de créer des algorithmes prédictifs performants et robustes à partir de données observées dans le passé* », explique Michael de Toldi, *chief data officer* de BNP Paribas Cardif.

Parmi les challenges actifs sur la plateforme, les compétiteurs peuvent s'essayer à prédire les résultats d'un tournoi de basket-ball. Récompense : 25 000 \$. Ils peuvent aussi calculer les chances de survie des voyageurs sur le Titanic... Au milieu d'entreprises telles que GE, Facebook, Mercks, Microsoft, Ford ou des institutions comme la Nasa, quelques

## « Accélérer le processus d'indemnisation des sinistres »

« En lançant cette compétition sur Kaggle, nous cherchons à accélérer le processus d'indemnisation des sinistres. Nous travaillons sur la base de nos données historiques pour lesquelles nous connaissons, aujourd'hui, les sinistres qui auraient pu être réglés plus rapidement. À partir de ces informations, nous construisons des algorithmes pour identifier, dès leur déclaration, les sinistres dont le règlement peut être accéléré. Sur Kaggle, ce n'est pas tant les modèles gagnants qui nous intéressent que les méthodologies mises en œuvre. Nous ne savons pas, aujourd'hui, ce que donnera ce challenge, mais il ouvre l'opportunité de soumettre cette problématique à un très grand nombre de *data scientists* à travers le monde. Il s'agit également d'une expérimentation qui pourrait nous permettre de détecter des talents au sein du groupe puisque le challenge est ouvert à l'ensemble de nos collaborateurs. »



**MICHAEL DE TOLDI**,  
*chief data officer*  
de BNP Paribas Cardif



assureurs ont tenté l'aventure. BNP Paribas Cardif vient ainsi d'initier un challenge ouvert jusqu'au 18 avril pour améliorer le traitement de ses processus d'indemnisation des sinistres (voir encadré p. 50). Les trois premiers du classement se partageront 30 000 \$. Les participations peuvent être à titre individuel ou par équipe. La plateforme établit ensuite un classement en fonction de la mesure d'erreur entre la prédiction et la réalité, sur la base des jeux de données fournis par les entreprises qui orchestrent le challenge. Si les résultats de ces concours n'ont pas encore concrètement fait évoluer les offres sur le marché, c'est parce que l'objectif consiste avant tout à apprendre des modèles développés par des spécialistes, les entreprises ne disposant pas toujours de ces ressources en interne. BNP Paribas Cardif ne compte ainsi qu'une vingtaine de data scientists. Ils sont environ 200 au sein du groupe Axa.

### Les Français s'y mettent...

Pourtant, avant BNP, Axa avait déjà lancé un concours fin 2014 sur des données relatives à la conduite de véhicule. « Les assureurs, gérants historiques d'un grand nombre de données, se lancent peu à peu dans de tels concours, mais il est vrai qu'à ce jour, les entreprises françaises ne sont pas encore très actives sur cette plateforme. En effet, la data science est une pratique encore récente et mettre en place un tel challenge nécessite une certaine maturité en terme de gestion de la donnée, ne serait-ce que pour préparer correctement les data que l'on soumet au concours. En effet, une vision claire des besoins en matière de données est impérative. Il faut aussi avoir des processus et des

**ALEXANDRE KEIFLIN**, CONSULTANT SENIOR,  
PRACTICE LEADER BIG DATA ET DIGITAL CHEZ OPTIMIND WINTER

### « L'entreprise qui soumet un challenge sur Kaggle développe une image positive »



#### ■ Pourquoi participer à des challenges sur la plateforme Kaggle ?

La disponibilité des données auxquelles les assureurs ont accès et les capacités informatiques et algorithmiques de traitement sont de plus en plus importantes. Les recours aux méthodes de machine learning et à la data science vont donc se développer dans le secteur de l'assurance. Répondre à ces challenges nous permet de nous étalonner par rapport aux meilleurs data scientists et de continuer à nous améliorer pour répondre au mieux aux besoins de nos clients. Cela contribue à l'enrichissement de l'expertise collective de notre société de conseil en gestion des risques. La plateforme permet,

par ailleurs, de partager les solutions de certains challenges.

#### ■ Et pour les entreprises qui soumettent des concours sur cette plateforme, quel est l'intérêt ?

Double. Pour le prix potentiel proposé, des milliers d'amateurs et de spécialistes travaillent sur la problématique soumise et sont donc susceptibles d'obtenir un résultat bien meilleur que celui des seuls collaborateurs de l'entreprise proposante. Et le groupe qui soumet un challenge sur une plateforme comme Kaggle développe une image positive sur le terrain du digital. Un atout en matière de recrutement ultérieur.

■ PROPOS RECUEILLIS PAR G. B.-F.

mécanismes d'anonymisation éprouvés et robustes», explique Stéphane Royer, responsable des équipes data, risque et processus au sein du Data innovation lab de la compagnie Axa. Si le travail en amont est non négligeable, l'assureur ne compte pourtant pas en rester là et envisage de lancer deux autres challenges en 2016 dont l'un concernera l'assurance vie. « L'aventure Kaggle nous a permis de détecter des talents et de susciter des vocations au sein d'Axa. C'est aussi une vitrine importante quand on fait de la gestion de la donnée un enjeu stratégique de la transformation digitale. Ce challenge a démontré à l'écosystème d'innovation autour de la donnée qu'Axa est un acteur et un interlocuteur qui compte. Cela facilite les collabora-

tions avec les start-up et les sociétés innovantes du domaine », ajoute Stéphane Royer.

Le partage des compétences est, en effet, un des volets majeurs de cette communauté. Certains challenges peuvent aboutir à des recrutements ou, à tout le moins, à une autre forme d'apprentissage. De fait, la plateforme propose à la communauté, des cours, tutoriels, méthodologies, et va jusqu'à mettre à disposition des compétiteurs moins bien équipés, des logiciels pour leur permettre de concourir. Surtout, les participants peuvent partager leurs travaux avec la communauté à l'issue du challenge. Une mine d'or pour les professionnels du secteur.

■ GÉRALDINE BRUGUIÈRE-FONTENILLE