Des sinistrés pris en charge plus vite

Grâce à la *data science*, on commence à prédire l'étendue des dégâts… avant même qu'une catastrophe naturelle s'abatte sur une région donnée. De quoi mieux anticiper la prise en charge des futurs sinistrés. Explications…

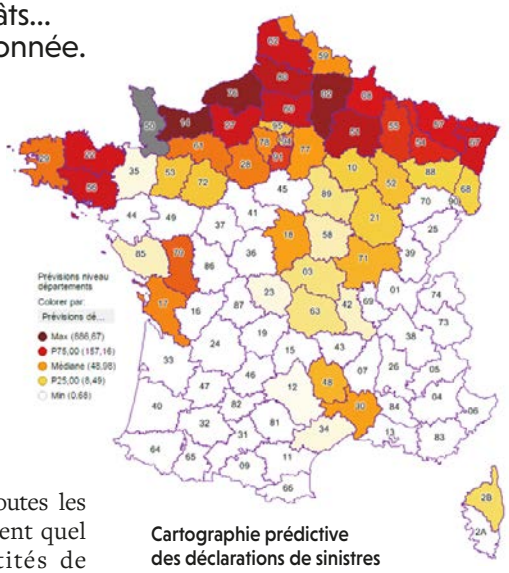
Inondations, tempêtes, cyclones, ouragans… Entre 1988 et 2013, les catastrophes naturelles ont coûté 1,9 milliard d'euros en indemnisations, avec 431 000 sinistrés en moyenne chaque année. Quand des catastrophes de ce type s'abattent sur un territoire, les assureurs activent immédiatement leurs plans de gestion de crise. Principal objectif: traiter au plus vite l'arrivée en masse de dossiers de demande d'indemnisation d'assurés impactés par la catastrophe et qui risquent d'affluer en agences. Mais aujourd'hui les assureurs veulent aller plus loin…

«Data visualisation» des futurs sinistrés

ORLANE MONNET

«*À la MAIF, nous développons par exemple un outil capable de prédire le nombre de déclarations de sinistres sur telle ou telle zone, avant même qu'elle ne soit touchée par une tempête hivernale ou une inondation cévenole*», lance Orlane Monnet du département Études statistiques de cet assureur. Calibré sur la sinistralité des événe-ments climatiques passés les plus significatifs (tempêtes Klaus, Xynthia, Joachim…), ce modèle de prévision utilise deux grands types de données. Tout d'abord, des données sur les communes et les assurés qui risquent d'être touchés: maisons ou appartements, domiciles ou résidences secondaires, propriétaires ou locataires, densité de population, existence ou non d'un plan de prévention des risques, etc. Ensuite, le modèle utilise aussi bien sûr des données météorologiques: force du vent prévue, durée estimée, quantité de précipi-tations à venir, etc.

«*Écrit en langage R*, notre outil se lance et s'utilise* via *une seule et unique interface de* data visualisation** *nommée Tibco Spotfire,* indique Orlane Monnet. *Elle permet de visualiser sur une carte le nombre de déclarations de sinistres prévisibles – plus les zones sont touchées plus elles apparaissent foncées.*» Par simple clic sur des curseurs, elle permet aussi de mesurer la sensibilité du modèle de façon interactive; exemple, si on majore par sécurité toutes les rafales de 10 km/h, on voit directement quel pourcentage augmente les quantités de déclarations de sinistres sur la carte. La MAIF a commencé à utiliser son outil sur des inondations fin 2016 et sur de petites tempêtes début 2017.



Cartographie prédictive des déclarations de sinistres par département

Anticiper les réparations

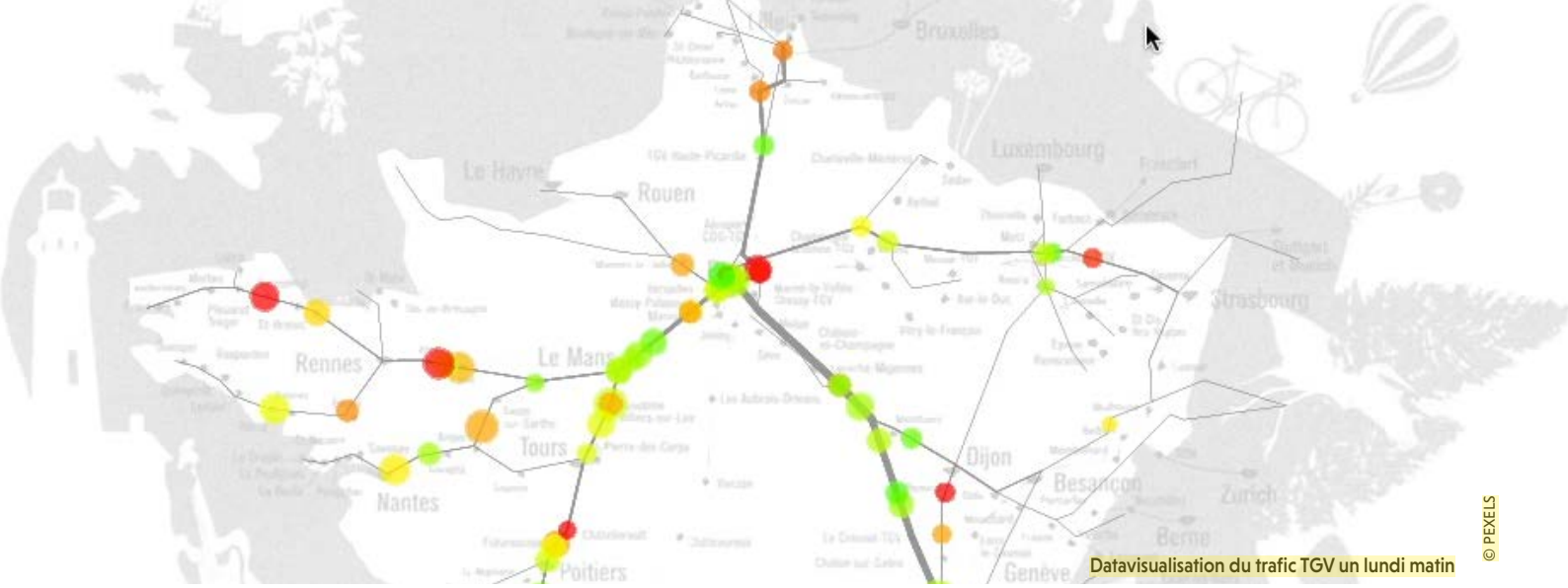
À l'avenir, ce type d'outil pourrait encore être amélioré pour évaluer plus précisément les coûts attendus, et même commencer à mobiliser les entreprises réparatrices (loueurs de motopompes et de groupes électrogènes, réparateurs de carrosseries de voitures…). Mais pour cela, il faudrait collecter des données encore plus précises sur les assurés: équipements, habite à tel étage, arbres sur son terrain, type de toiture, parking aérien ou souterrain, terrain en forte ou faible pente… Enfin, ce type d'outil pourrait aussi servir à alerter par mail ou SMS les assurés concernés et leur prodiguer des conseils de prévention personnalisés.

« L'usage d'algorithmes de machine learning nous aidera à encore mieux assurer les biens de nos sociétaires »

© Darri MAIF

STÉPHANE RENOUX
responsable du programme Data & intelligence artificielle à la MAIF

* Langage informatique dédié aux statistiques et à la science des données
** En mettant en images les données, la « dataviz » est un moyen d'identifier des tendances et corrélations



Datavisualisation du trafic TGV un lundi matin

© PEXELS

Améliorer l'information temps réel

SNCF met au point des algorithmes de *big data* capables de prédire en temps réel l'affluence à bord d'un train, et l'heure prévisionnelle de ses futurs passages en gare en cas de situation perturbée. Explications avec Maguelonne Chandesris qui dirige l'équipe Innovation & Recherche « Statistique, Économétrie et Datamining » de SNCF.

MAGUELONNE CHANDESRIS

Des futurs du passé... au futur du présent

Quel est l'objectif principal de ce projet ?

M. C. Baptisé RELUANCE (qui contient « retards et affluence en avance »), ce projet vise deux grands objectifs : prédire en temps réel le niveau de confort à bord (par exemple si on trouvera une place assise) et à quelle heure est prévue l'arrivée d'un train dans telle ou telle gare en cas de situation perturbée... Si le second point intéresse l'ensemble des 5 millions de voyageurs qui empruntent nos trains chaque jour, le premier concerne ceux qui prennent des trains régionaux, sans réservation. Pour y parvenir, nous avons développé des algorithmes de *machine learning* déjà testés sur plus de 3000 trains en circulation. Chaque jour, ils s'alimentent de centaines de milliers de données, à partir desquelles ils fournissent des millions de prédictions, au fur et à mesure de la remontée en temps réel des informations !

Sur quels types de données se basent-ils ?

M. C. Dans la pratique, nos algorithmes utilisent deux grands types de données. Tout d'abord celles qui concernent la localisation précise des trains en temps réel, collectée par des capteurs placés au sol et à bord des trains. Mais aussi des données sur le nombre de passagers qui montent et descendent des trains à chaque gare, obtenues par des capteurs installés à bord, au niveau des portes. Toutes ces données sont acheminées en temps réel, *via* des interfaces de programmation (API), vers nos algorithmes de calcul, hébergés sur nos serveurs situés près de la gare de Lyon.

Comment fonctionnent ces algorithmes ?

M. C. Leur principe général est facilement compréhensible. Très concrètement, ils vont chercher dans l'historique de ce type de données les situations les plus semblables à la situation actuelle. Une fois celles-ci identifiées, ils s'appuient sur ce qui est advenu après ces situations semblables passées (les « futurs du passé »)... pour prédire ce qui va se passer maintenant (le « futur du présent »). Point important : ces algorithmes dits « d'apprentissage non-paramétrique » réalisent leur phase d'apprentissage en continu, sans aucun reparamétrage à effectuer, facilitant la maintenance industrielle. Au final, ils améliorent sensiblement les prédictions de l'heure d'arrivée des trains à destination, et une marge d'erreur sur l'affluence inférieure à vingt voyageurs... pour des trains pouvant transporter plus de neuf cents personnes.

Quelles sont les prochaines étapes ?

M. C. Forts de ces résultats, nous avons déjà démarré les travaux d'industrialisation. Il nous faut notamment implémenter ces algorithmes de manière robuste dans « l'immeuble numérique » du groupe SNCF, afin que ces prédictions puissent alimenter tous nos systèmes d'analyse et de diffusion : affichages en gare, applications clients, centres opérationnels de supervision...
Ces informations en temps réel seront en effet aussi très utiles aux agents qui régulent la circulation des trains et les flux de passagers en gare. Au final, tous nos trains devraient bénéficier de ces prédictions algorithmiques à l'horizon 2020.

© SNCF