

Prévoir la panne d'un train : à la SNCF, ce sera bientôt possible

Prévoir les pannes des trains avant que celles-ci ne perturbent le trafic ferroviaire. Pour la SNCF, l'enjeu de la maintenance prédictive est évidemment majeur. Pas étonnant donc de voir la compagnie nationale s'y intéresser de près, d'autant que, dans le même temps, les trains eux-mêmes émettent de plus en plus de données, renseignant sur l'état de fonctionnement de leurs systèmes. « Pour optimiser la maintenance, nous cherchons à utiliser au mieux les données des trains, d'autant que les nouvelles fonctions qu'ils proposent sont elles-mêmes sources de complexité, donc de pannes », raconte **Philippe de Laharpe**, chef de projet télédiagnostic à la direction du matériel de la SNCF. Une volonté qui a conduit la compagnie à mettre le doigt dans l'engrenage du Big Data.

« On a commencé par une analyse de 'Small Data', avec des analystes examinant les données envoyées par les rames, reprend le chef de projet. Rapidement, face à la croissance des volumes, on a développé des systèmes automatisés embarquant le savoir-faire des équipes pour mettre en place une maintenance corrective connectée. L'étape suivante consiste à passer à la maintenance proactive, capable de détecter une panne quand elle se produit mais n'a pas encore impacté l'exploitation ou encore d'analyser des signaux faibles, pas forcément prévus lors de la conception des rames, mais qu'on va valoriser via des analyses de plus en plus complexes. »

Un réceptacle Hadoop

C'est pour cette étape que la SNCF a décidé de miser sur les technologies de Machine Learning. Expérimentées dans le cadre d'un pilote portant sur une génération récente de trains Bombardier circulant sur le réseau Transilien, ces technologies vont très vite passer en phase de pré-production « afin de concrétiser les premiers résultats sur des données de production ». Le projet implique la **direction du matériel**, qui apporte des techniciens de télédiagnostic, la **direction du digital** de la compagnie nationale ainsi que le prestataire **Quantmetry**. La plate-forme technique – un datalake Hadoop couplé au framework d'analyse de données Spark – est amenée par VSCT, la DSI de Voyages-SNCF qui intervient comme fournisseur de services pour la SNCF.

Menées en 2015, les phases d'expérimentation ont permis de mettre en évidence quelques difficultés liées à l'emploi du Machine Learning. « Quand on construit un modèle avec une boîte noire algorithmique, la difficulté consiste à surveiller les dérives de ce modèle, analyse **Héloïse Nonne**, data scientist senior chez Quantmetry. Sans oublier les difficultés de communication avec les techniciens qui ont besoin de preuves tangibles pour intervenir. Par ailleurs, si l'objectif est de fiabiliser le matériel, encore faut-il savoir expliquer la panne, ce que ne propose pas le Machine Learning. »

S'intégrer à la chaîne de maintenance de la SNCF

Le modèle doit aussi être amendé avec les données réelles de circulation des rames car, même à l'arrêt, un matériel continue à émettre des données qui ont toutes les chances de fausser le modèle si on les interprète comme des informations publiées pendant un trajet. Sans oublier une dérive

bien connue des algorithmes d'intelligence artificielle : la sur-performance, autrement dit la propension des algorithmes à surinterpréter les événements, surtout dans des cas où, comme ici, on fait face à un taux de panne assez bas. « *Les équipes métier ont donc retravaillé les modèles pour les améliorer* », note Héloïse Nonne, pour qui la principale difficulté du projet résidait toutefois dans la maîtrise de la qualité des données et dans la gestion des sources hétérogènes d'informations. Un classique des projets Big Data.

De son côté, Philippe de Laharpe apporte une vision davantage tournée vers les métiers des défis qui attendent encore le projet. « *Une des difficultés réside dans le fait que les matériels qui font l'objet du pilote connaissent assez peu de pannes. D'où la question : comment faire baisser ce taux déjà faible et à quel prix y parviendra-t-on ? Il faut aussi prendre en compte l'impact organisationnel, note le chef de projet. Traiter une panne probable signifie qu'il faut faire évoluer la façon de travailler, et changer les processus de maintenance* ». Car, faire des prédictions de panne ne suffit pas ; ces dernières doivent s'intégrer dans un « *système de management* », comme dit Philippe de Laharpe, aux côtés d'autres sources d'alerte afin, en fonction de la sévérité de l'incident, de planifier la réparation dans le centre de maintenance idoine de la SNCF.

Un test sur 176 rames Bombardier

Malgré ces obstacles, Philippe de Laharpe se veut confiant. « *Je suis à peu près convaincu du retour sur investissement de la solution* », explique-t-il dans un entretien avec *Silicon.fr*. Et de citer un exemple mis en évidence lors du pilote, où les algorithmes prédisaient deux fois plus de pannes que les analystes métier. « *Cela nous encourage à aller plus loin avec le Machine Learning, d'autant que les spécialistes des systèmes techniques que nous surveillons sont rares. Ce sont donc des compétences qui coûtent cher* », remarque le chef de projet.



Début avril, la SNCF lancera donc un test grandeur nature, exploitant les données de 176 rames Bombardier du réseau francilien. Des trains qui possèdent chacun 7 ou 8 ordinateurs de bord, et émettent quelque **70 000 données par mois et par rame**. Ces informations sont transmises toutes les 30 minutes aux systèmes d'information de la compagnie ferroviaire. « *On va laisser les pannes apparaître comme avant et mesurer pendant 6 mois les améliorations qu'aurait pu amener l'application prédictive* », explique Philippe de Laharpe. Des résultats qui permettront à l'entreprise de décider si le projet doit oui ou non être mis en production et généralisé. Car, au-delà même du matériel roulant, la maintenance prédictive a aussi un intérêt pour l'entretien des voies par exemple, comme l'a d'ailleurs montré le tragique accident de Brétigny-sur-Orge en juillet 2013. « *Progressivement, tous les silos de l'entreprise ont vocation à rejoindre le datalake* », note Philippe de Laharpe.

A lire aussi :

[BluData : comment Auchan bâtit son bras armé Big Data](#)

[Big Data : le Machine Learning protège les Livebox Orange de la foudre](#)

[Big Data : Mappy accélère son cluster Hadoop... sans acheter de serveurs](#)