

L'intelligence artificielle dans le secteur public

Spécificités, défis et solutions



Section I:

Défis et opportunités

INTRODUCTION

L'utilisation de la data science, du machine learning et de l'intelligence artificielle (IA) dans le secteur public est en constante augmentation, et les opportunités d'applications et d'extensions nouvelles ne font que s'élargir: les avantages de l'IA s'étendent aujourd'hui à tous les domaines imaginables, qu'il s'agisse de renforcer l'efficacité des procédures administratives ou de prévoir les risques d'occurrence des délits et des activités criminelles, en passant par le pilotage de politiques publiques.



DÉFIS ET SOLUTIONS

Si les opportunités sont nombreuses, les défis ne le sont pas moins - certains d'entre eux étant d'ailleurs identiques à ceux auxquels sont habituellement confrontées les entreprises du secteur privé, pour n'en citer que quelques-uns :

- Freins à l'installation initiale, avec notamment un rythme soutenu d'adoption et d'abandon des technologies, ce qui complique la pertinence du choix d'investissement dans les outils les mieux adaptés
- Questions de réglementation et de confidentialité des données, qui font qu'il est difficile de comprendre comment les équipes peuvent travailler dans le cadre de ces contraintes
- Questions de compétences : passer à l'ère de l'intelligence artificielle ne consiste pas à simplement embaucher des équipes supplémentaires. Il s'agit de mettre en œuvre un changement fondamental tous échelons confondus, afin d'intégrer la donnée à tous les processus associés aux différents cas d'usage. La technologie – et plus spécifiquement les plateformes de data science, de machine learning et d'IA –, peut certes aider à relever ce défi, mais elle n'est en aucun cas une solution miracle.

Outre ces difficultés largement répandues, le secteur public doit relever d'autres défis plus spécifiques :

LENTEUR ADMINISTRATIVE

Nombreuses sont les entreprises et organisations à vouloir innover, mais les cycles d'implémentation peuvent parfois être longs. Compte tenu de la nature très évolutive du paysage en matière d'intelligence artificielle – notamment en termes de rotation des technologies les plus populaires utilisées pour les projets de machine learning (cf. [le post sur Technoslavia](#)), ces entreprises restent systématiquement en queue de peloton au niveau de la courbe d'adoption. Les plateformes de data science, de machine learning et d'intelligence artificielle – à l'instar de Dataiku –, peuvent aider à cet égard, dans la mesure où elles peuvent s'intégrer à différents outils et technologies, permettant ainsi aux entreprises d'assurer une expérience sûre et cohérente en termes de projets data, quelle que soit la technologie sous-jacente.

PÉNURIE DE TALENTS

Embaucher des data scientists n'est pas chose simple. Considéré par certains comme le [job le plus populaire du XXI^e siècle](#), la demande a littéralement explosé. Le défi semble malheureusement encore plus important dans le secteur public, où les budgets sont contraints et où les technologies et les outils IT de pointe – [l'un des moyens clés pour fidéliser](#) les data scientists – sont parfois implémentés plus tard que dans le secteur privé. Cela dit, il ne suffit pas uniquement d'embaucher les meilleurs data scientists, il faut également tirer le meilleur parti possible des personnels existants, qu'ils disposent déjà ou non d'une formation spécifique en data science.



ORGANISATION ET PÉRENNISATION DES SAVOIRS

Une des spécificités principales de tous les services administratifs, c'est qu'ils sont marqués par des rotations de personnel à une fréquence moyenne triennale. Comment valoriser au mieux la diversité des compétences disponibles ? En permettant de pérenniser les travaux, pour capitaliser sur les développements passés, qui doivent être conservés, référencés, pour permettre des usages durables. Nous ne sommes pas dans un usage durable si un service s'arrête parce que la personne qui savait le faire fonctionner est mutée.

Chacun de ces défis est complexe, et ce Livre Blanc n'a nullement vocation à traiter en profondeur chacune des spécificités évoquées. Nous pouvons toutefois recommander de lire plusieurs études articles ou Livres Blancs déjà publiés par Dataiku :

Defining a Successful AI Project: A Framework for Choosing the Right Use Cases, Why Enterprises Need Data Science, Machine Learning, and AI Platforms, Insights on Top Challenges from 100+ Data Professionals, Executing Data Privacy-Compliant Data Projects, Ovum: Profit from AI and Machine Learning, et The 6 Key Challenges to Building a Successful Data Team.

Les autres sections s'intéressent aux principales applications dans le secteur public, avec des cas d'usage et des exemples spécifiques permettant de mettre en lumière l'impact potentiel de l'IA.



Section II:

Quelques exemples d'applications majeures

CRIMINALITÉ ET MAINTIEN DE L'ORDRE

L'analyse avancée peut jouer un rôle majeur pour aider les forces de police à affecter au bon endroit et au bon moment des ressources matérielles, et surtout humaines, particulièrement précieuses.

Des algorithmes qui s'appuient sur de nombreuses années de données criminelles sont par exemple capables de localiser des « points chauds » à surveiller sur la base de signaux faibles qui peuvent échapper à l'œil humain. L'analyse multidimensionnelle permet d'observer des tendances qui échapperaient très probablement à une équipe d'analystes humains, comme certaines heures spécifiques dans la journée, ou certains jours de la semaine, au cours desquels des actes délictueux sont susceptibles de se produire en un lieu donné.

Grâce à ce type d'analyse, les services de police peuvent déployer des équipes sur les zones les plus sensibles. Cela permet non seulement d'augmenter les chances d'appréhender les délinquants, mais surtout, de créer à long terme un effet dissuasif sur la criminalité.

L'analyse prédictive n'est toutefois pas exempte de controverses. Ses critiques estiment que les algorithmes ne feront qu'encourager une présence policière excessive dans des quartiers difficiles. Cela ne signifie pas qu'il faut s'interdire d'utiliser cette technologie. Mais cette perspective met clairement en lumière l'importance des plateformes de data science, de machine learning et d'IA qui introduisent dans ce cas une plus grande transparence (visibilité de la « boîte blanche ») et qui aident à identifier et à remédier à des biais potentiels dans les modèles. Personne ne veut d'une IA aveugle et incompréhensible, l'interprétabilité des modèles et la pédagogie de la méthode sont donc des aspects aussi essentiels que la précision d'un algorithme.



GO FURTHER:

*Get the Guidebook to
Working with Data Under
Privacy Constraints*



CAS D'ESPÈCE EN OREGON

DE L'INFLUENCE DE LA MÉTÉO SUR LE NIVEAU DE CRIMINALITÉ À PORTLAND

Au fil des années, de nombreuses tentatives de prédiction des activités délictueuses et criminelles ont bien sûr été mises en œuvre. Cependant, d'une manière générale, les êtres humains ne sont pas encore très efficaces pour savoir comment intervenir et identifier les endroits où des délits risquent d'être perpétrés. Plus de 200 000 articles spécialisés ont déjà été publiés sur les causes sous-jacentes de la criminalité, et les études déjà réalisées associent généralement ces activités au niveau de pauvreté et d'instruction, et à divers autres paramètres sur lesquels il est somme toute difficile d'agir.

En septembre 2016, le National Institute of Justice (NIJ), l'agence de recherche, de développement et d'évaluation du département américain de la Justice, a lancé l'initiative **Real-Time Crime Forecasting Challenge** (Prédiction en temps réel des actes criminels) destinée à anticiper les points névralgiques de la délinquance dans la ville de Portland (Oregon), et à encourager l'innovation dans ce domaine. L'équipe Dataiku a soumis des propositions à l'occasion de ce challenge et a été sélectionné dans plusieurs catégories, en proposant notamment des capacités de prédiction des vols de véhicules (Auto Theft) beaucoup plus performantes que celles des autres concurrents. Ce cas spécifique est illustré par une présentation générale du projet, avec la méthodologie adoptée par l'équipe pour le mettre en œuvre.

Explication succincte de ce projet et de son utilisation. Pour une explication technique détaillée, [consulter l'article Data from the Trenches](#).

LA MÉTHODOLOGIE DATAIKU

L'équipe Dataiku a d'abord observé que l'une des caractéristiques clés des données criminelles est la présence associée d'une dimension géographique et d'une dimension temporelle (les événements criminels surviennent dans différents lieux, à différentes dates). Ces événements peuvent également être modifiés par de nombreuses autres variables – conditions météorologiques, infrastructures urbaines, données démographiques, événements publics, politique gouvernementale, etc.

L'équipe a entamé son projet avec les données publiées par le NIJ, et les a enrichies avec différents jeux de données publiques, notamment les rapports de police, les informations de recensement, les données de l'application Foursquare, les informations détenues par la presse et les médias, et les données des services météorologiques.



UN DÉFI MAJEUR : ANTICIPER LES ACTIVITÉS CRIMINELLES

Les mesures préventives sont récemment devenues un sujet brûlant : est-il possible de prévoir, et donc de prévenir, les actions criminelles avant qu'elles ne surviennent vraiment ? L'objectif de projets tels celui initié par le NIJ ne consiste nullement à s'attaquer aux délinquants potentiels, mais plutôt à mettre à la disposition des décideurs politiques et des instances policières des techniques opérationnelles leur permettant d'anticiper ce type d'événements.

De fait, **les études déjà réalisées sur les politiques de prévention** suggèrent que le renforcement aléatoire de la présence policière ne s'avère pas aussi efficace pour réduire le taux de criminalité qu'un ciblage plus précis des points ou des secteurs névralgiques, où la probabilité de futurs incidents à caractère criminel demeure élevée. De surcroît, **les éléments tangibles collectés** semblent montrer que les interventions menées dans ces secteurs névralgiques n'entraînent pas de déplacement des activités criminelles vers des zones limitrophes, mais induisent effectivement une **diminution des taux de criminalité et d'incarcération** dans l'ensemble du secteur géographique considéré.

Pouvoir comprendre les caractéristiques qui favorisent les activités criminelles dans un secteur donné constitue un outil intéressant pour soutenir les efforts déployés par les forces de police, avec différentes politiques permettant de modifier les facteurs en cause.

PRÉVOIR QUOI ?

Les données publiées par le NIJ ont formé la base du projet. Elles constituent un dataset de tous les appels passés au central de la police, ayant conduit à des interventions des forces de l'ordre pour la criminalité urbaine, les cambriolages et les vols de véhicules, entre les mois de mars 2012 et mars 2017. Telles sont les données utilisées par l'équipe Dataiku comme éléments de base pour élaborer une prévisibilité des événements criminels.

Ces données sont produites sous la forme d'un point géographique, d'une heure et d'une date, avec les spécificités de l'appel passé au central. Lorsqu'on tente d'identifier les endroits où une activité criminelle risque de survenir, il est important de prendre en compte les caractéristiques géographiques (en quoi un secteur ou un quartier est-il différent par rapport à un autre ?) et les caractéristiques temporelles (en quoi une heure particulière constitue-t-elle une opportunité par rapport à une autre ?)

QUELLES DONNÉES UTILISER POUR PRÉVOIR LES ACTIVITÉS CRIMINELLES ?



OpenStreetMap:

Les premières données fournies par l'équipe consistaient en points d'intérêt avec OpenStreetMap. Outre les meilleurs itinéraires et les curiosités géographiques, OSM note les points présentant un intérêt particulier. Les spécificités d'un endroit sont enrichies par des informations et par des catégories générées par l'utilisateur – services publics, hôtels, magasins, restaurants, services postaux, etc. L'équipe a recueilli ces points et marqueurs géographiques.



Foursquare:

Plus que de simples points d'intérêt locaux, l'équipe Dataiku voulait connaître le comportement réel des citoyens dans les différents quartiers de Portland. Pour mieux comprendre comment les gens passent leur temps dans chaque quartier, l'équipe s'est intéressée aux données de l'API Foursquare et en a extrait les réservations et les entrées dans les restaurants et autres lieux publics.



Postes de police:

L'équipe s'est ensuite intéressée à la carte des postes de police à Portland, sous forme d'une mosaïque de la ville montrant la répartition des secteurs par poste de police.



Données de recensement U.S.:

Aux États-Unis, les services de recensement collectent plus de 20 000 informations sur chaque quartier (entre 600 et 3 000 habitants par quartier) : nombre d'habitants, sexe, revenus moyens, nombre de véhicules automobiles par foyer, etc. Afin d'effectuer un tri parmi ces quelque 20 000 informations, l'équipe a uniquement extrait les caractéristiques en corrélation avec notre cible (appels au central de police – le 911 aux États-Unis) pour les années concernées (2014, 2015, 2016). Ces données sont d'ordre géographique (quadrillage) et historique, dans la mesure où l'on peut observer les variations au cours de l'année



Données météorologiques :

Afin d'explorer l'aspect temporel des activités criminelles, l'équipe a extrait des données de l'API météo de la NOAA, l'Agence américaine d'observation océanique et atmosphérique. Ce qui lui a permis de disposer de la quasi-totalité des informations météo nécessaires (températures, précipitations, vents dominants, etc).



Données météorologiques :

L'objectif était ici de « nettoyer » le bruit des données associées au jour de la semaine ou aux événements (Constate-t-on plus ou moins de délits lors des fêtes de fin d'année ?)

EN SAVOIR PLUS ?

Si vous souhaitez de plus amples détails sur la manière dont l'équipe a utilisé ces données pour remporter le concours, [consultez l'article technique](#). Dataiku a également conduit un projet permettant de cartographier et de prévoir les taux de criminalité dans le Grand Londres – [voir les détails ici](#). Pour chacun de ces projets, les équipes ont utilisé la plateforme Dataiku DSS pour l'ensemble du processus, de l'intégration initiale des données, jusqu'à leur nettoyage et à leur modélisation.

Optimisation des services de santé

La crise sanitaire du COVID-19 a montré une forte implication des personnels de santé, mais aussi souligné l'importance d'une donnée de qualité, agrégée, partagée, actualisée en temps réel, afin de permettre un pilotage efficace des politiques sanitaires. Quand les données sont hétérogènes et en silos, la capacité à les exploiter de manière unifiée et cohérente à travers une plateforme de data science devient d'autant plus cruciale.

L'utilisation de l'intelligence artificielle dans le domaine de la santé peut apporter des bénéfices directs et indirects aux patients :

- les bénéfices directs sont ceux qui aident les professionnels de santé à améliorer la prise en charge des patients, et ces patients à mieux s'impliquer dans leur démarche de soins : gestion de la capacité d'accueil en temps réel, pilotage plus fin des ressources humaines nécessaires, ...
- les bénéfices indirects sont ceux qui ont attirés à l'amélioration des systèmes de santé, et qui vont permettre de travailler plus vite, plus efficacement, de limiter les dépenses ou d'optimiser les revenus : automatiser des processus de reporting, détecter des corrélations entre facteurs de contamination, ...

Transparence et détection d'anomalies

Un bon exemple est fourni par l'Union européenne, qui a récemment lancé le projet Digiwhist dont la vocation est de recueillir et de rendre aisément disponibles les données relatives aux marchés publics des 28 États membres. Ce système répertorie plus de 20 millions de marchés publics et classe automatiquement chaque contrat selon un certain nombre de métriques qui établissent une corrélation avec de possibles actions de corruption, comme le nombre de soumissionnaires, la longueur de la sollicitation et le volume d'informations publiées en relation avec les appels d'offres. Ce système aidera l'UE à identifier les marchés pouvant avoir été obtenus par des méthodes inappropriées, assurant ainsi une plus grande transparence dans la procédure d'attribution des marchés publics.



Les véhicules autonomes offrent l'opportunité de réduire le coût des transports publics, mais aussi d'accroître leur fiabilité. Un certain nombre de constructeurs expérimentent déjà des bus autonomes. Cette technologie permettra de disposer de transports en commun plus fluides, plus sûrs, avec la possibilité de mettre simultanément en service de multiples véhicules pour répondre à une demande accrue. Grâce à cette capacité de montée en puissance, les villes seront en mesure de mettre en œuvre des lignes de bus offrant les mêmes capacités de transport que le rail, traditionnellement considéré comme la meilleure option pour le transport en commun à forte densité. Ce nouveau mode de transport sera toutefois beaucoup moins coûteux, dans la mesure où, contrairement au rail, il ne nécessite pas de lourds investissements en voies ferrées, tunnels et autres infrastructures majeures.



REFERENCES ADDITIONNELLES

Consultez la présentation faite par Luca Maria Aiello, Senior Research Scientist @ Nokia Bell Labs

The Spirit of the City: Capturing Network-Generated Data for Better Cities (en anglais)

Le secteur des transports induit un autre sujet d'importance : la maintenance prédictive basée sur l'IA, qui s'applique aussi bien aux réseaux de transport public qu'aux flottes de véhicules en tous genres, des avions aux matériels roulants. Les algorithmes de maintenance prédictive offrent aujourd'hui la perspective de coûts de maintenance réduits, avec pour corollaire une durée de vie accrue des équipements.

A l'instar des propriétaires de véhicules particuliers, le secteur public assure actuellement la maintenance des équipements à intervalles réguliers, sur la base de l'âge du véhicule et du kilométrage enregistré. Cette approche est parfaitement raisonnable, jusqu'à ce que l'on trouve une meilleure façon de procéder, avec des applications IA capables de surveiller en temps réel l'état d'un bus, d'un train, d'un camion, ou de toute autre machine coûteuse, avec la possibilité de prévoir les problèmes d'entretien potentiels, permettant ainsi aux professionnels de maintenance d'anticiper de manière adéquate.

EN SAVOIR PLUS :

Un Livre Blanc pour vous guider pas à pas dans la mise en œuvre de l'AI pour la maintenance de vos équipements (en anglais)



Défense

Les applications de la data science irriguent également les secteurs régaliens, qui ont des contraintes plus particulières s'agissant par exemple de la sécurité et de la confidentialité des données. Les plateformes d'analyse et de traitement de données comme Dataiku DSS peuvent ainsi fonctionner en offline sur des réseaux sécurisés, ce qui permet d'être en adéquation avec les politiques de sécurité extrêmement poussées imposées à certains ministères, et particulièrement le ministère des armées. Pour ce dernier, les cas d'usages sont multiples et peuvent s'apparenter à ceux que l'on constate dans le secteur privé (maintenance prédictive, optimisation des flux logistiques), avec une adaptation aux besoins de la défense (mise en condition opérationnelle, protection des données).

D'autres thèmes d'application sont en revanche plus spécifiques. Le **renseignement** par exemple, peut bénéficier en premier lieu des atouts de l'IA. Il constitue, avec la dissuasion, un des principaux enjeux de souveraineté de part la sensibilité des données que les experts ont à traiter. Cela impose un respect sans faille du besoin d'en connaître, transposé à la gestion de ces données. L'IA permet ainsi d'identifier des comportements suspects, de réduire les données à étudier, d'anticiper des tendances sur des médias de communications.

Il en est de même pour la **cybersécurité**. Ce domaine doit faire face encore à de nombreux enjeux lorsque l'on parle d'IA, parmi lesquels l'"interprétabilité" des modèles. En effet, l'IA s'avère adaptée à certaines thématiques, comme la réponse à incidents, l'analyse post-mortem, pour identifier les traces d'une attaque. Elle permettrait aux éditeurs d'anti-virus d'aller plus loin que la simple identification par signature. En revanche, des efforts doivent être maintenus concernant l'exploitation (encore complexe) des données car générant trop de faux positifs. Il s'avère donc nécessaire que les projets IA dans ce domaine soient interprétables, mesurables, configurables, en mode "boîte blanche". Ils doivent aussi favoriser la collaboration entre analystes cyber et experts IA.

La philosophie de Dataiku est de privilégier l'autonomie des clients. Cette approche correspond parfaitement aux attentes aux acteurs du secteur de la défense, qui doivent légitimement garder la main sur ses projets sensibles.

Pour un exemple d'usage parmi d'autres, cf. le blogpost sur la détection de radicalité en ligne (Twitter): <https://blog.dataiku.com/guest-blog-predicting-connections-to-isis-based-on-tweet-content>



Transports et ville intelligente

Les villes génèrent leurs propres écosystèmes de données : réseaux de transport, réseaux sanitaires, données relatives à la sécurité, informations collectées par les infrastructures urbaines, ... Comme dans tout écosystème de données, il existe un enjeu à savoir tirer parti de cette masse d'informations pour appuyer le pilotage des politiques publiques. D'ailleurs, selon le baromètre 2018 du syndicat Syntec numérique sur la transition numérique des territoires, 92 % des collectivités de plus de 5 000 habitants ont déjà lancé une initiative de ville intelligente.

Les applications principales de l'exploitation des données au niveau des collectivités locales concernent la maintenance prédictive des équipements publics, le pilotage en temps réel de données de consommation énergétique, le suivi de la qualité de l'air, ... Il peut s'agir aussi de gérer l'éclairage, le chauffage, la consommation et la récupération d'eau. On voit également émerger des systèmes intelligents de gestion de l'éclairage public avec des capteurs de mouvements.

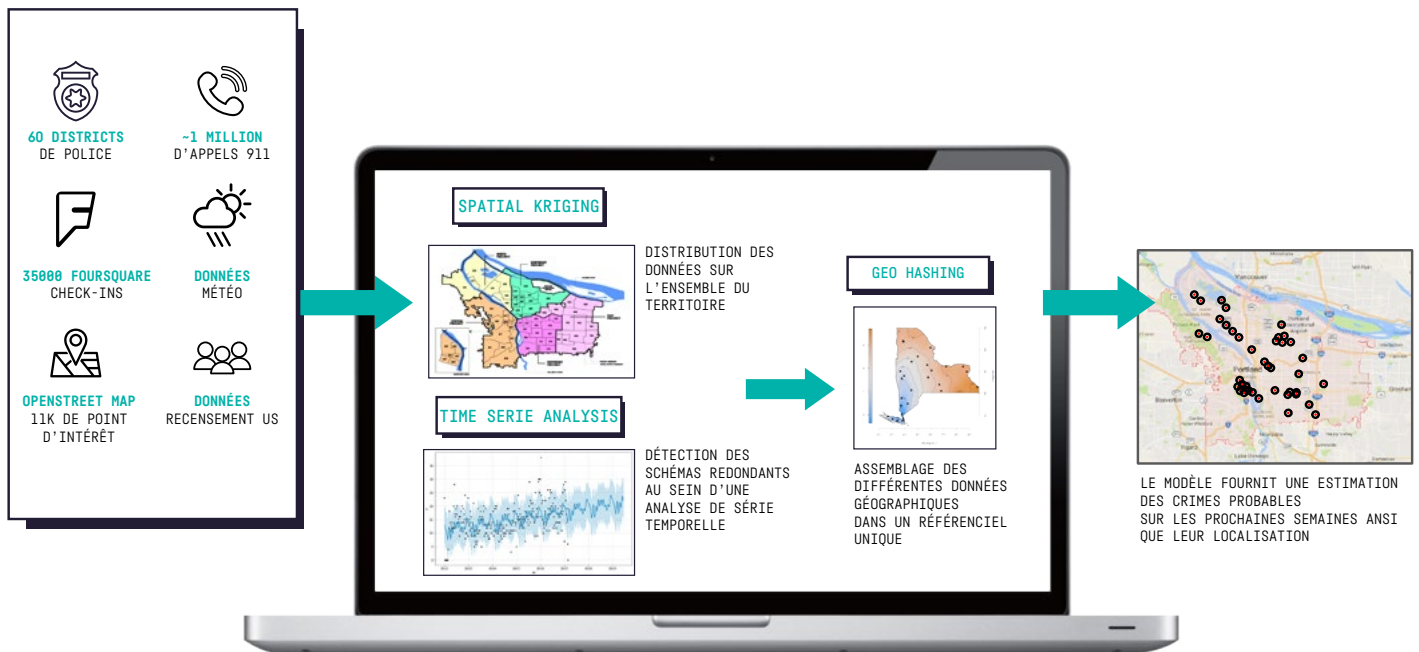
Le domaine des transports est intimement lié à ces questions de gestion des données des collectivités, notamment si on veut s'adresser à la saturation croissante des réseaux dans les principales métropoles, ou réduire le nombre du million d'automobilistes, d'utilisateurs des transports en commun, de cyclistes et de piétons qui meurent chaque année sur les routes dans le monde entier.

On perçoit que les bénéfices sont importants, mais beaucoup de maires notent des problèmes d'alignement des systèmes d'informations aux exigences et aux enjeux de la ville numérique (seules 35% des collectivités estiment que leur SI est adapté). De ce fait, beaucoup de projets sont au stade de l'expérimentation et peinent à entrer dans une phase de production.

L'utilisation de plateformes de Machine Learning - IA peut permettre de passer de ce stade expérimental à une phase d'industrialisation des projets.



PREDICTION DE LA CRIMINALITE À PORTLAND



PREDICTION D'ÉVÉNEMENTS SUR LA BASE D'ANALYSE DE SÉRIE TEMPORELLE

Plus d'information :

<https://innovationhub-act.org/news/innovation-challenge-winners>

PLANIFICATION DES OPÉRATIONS

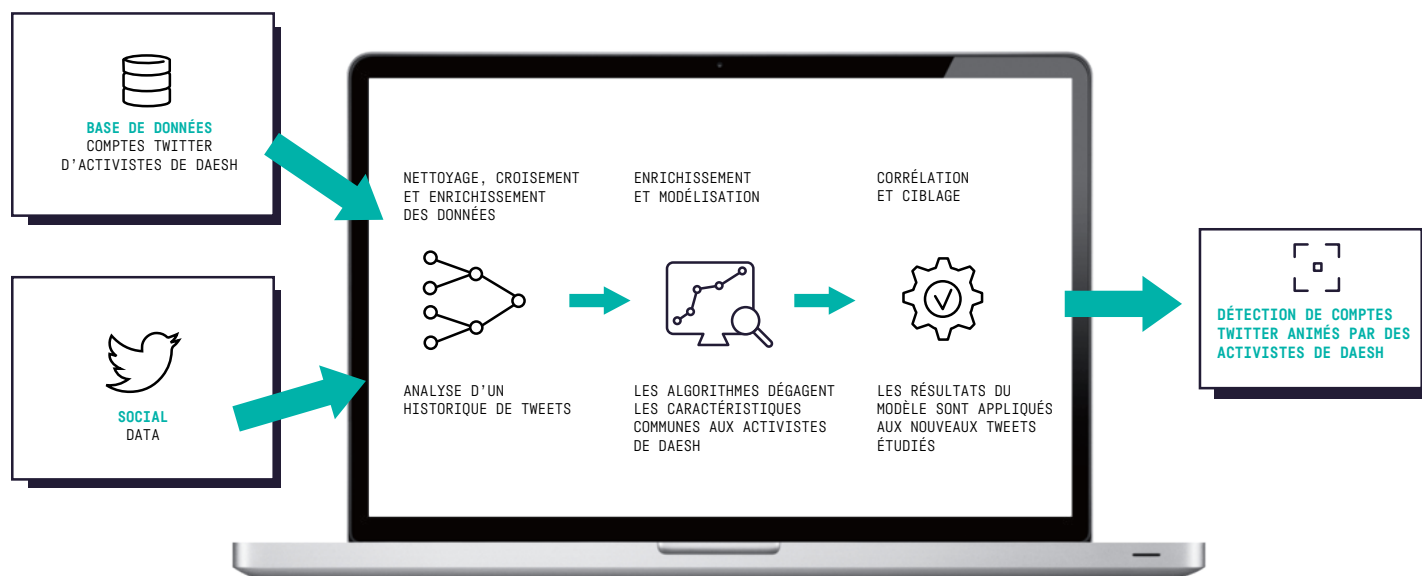


PREDICTION D'ÉVÉNEMENTS SUR LA BASE D'ANALYSE DE SÉRIE TEMPORELLE

Plus d'information :

<https://innovationhub-act.org/news/innovation-challenge-winner>

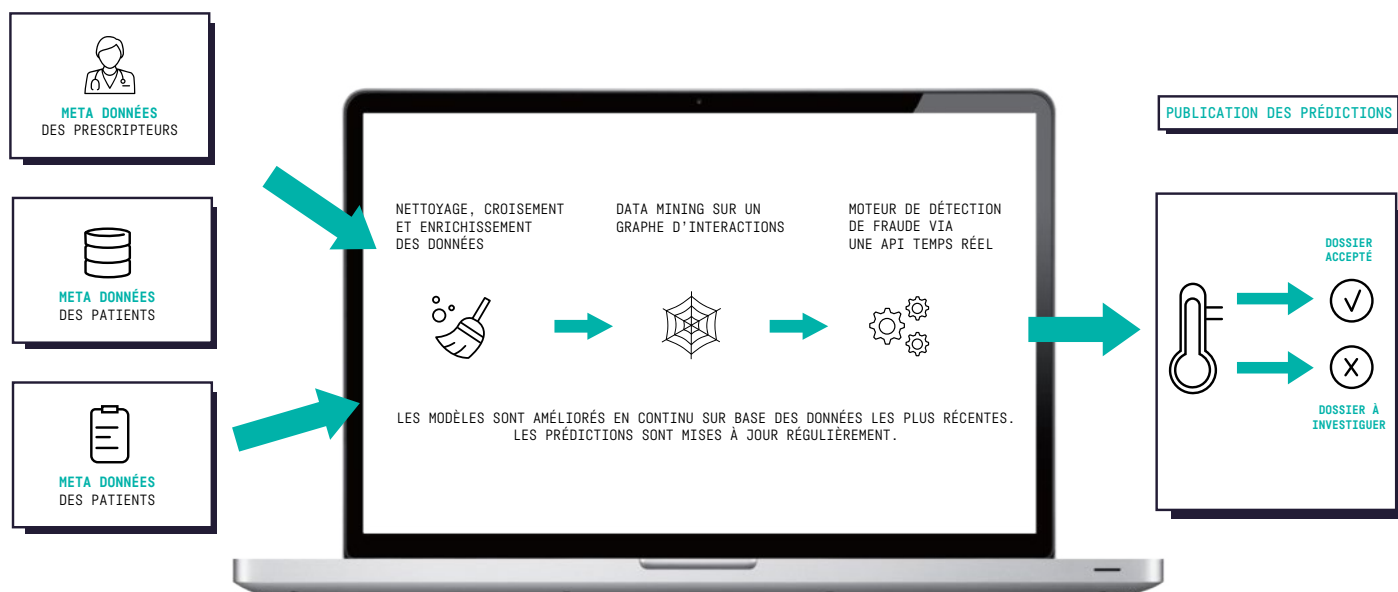
IDENTIFICATION DE SOUTIENS ACTIFS DE DAESH À PARTIR D'UNE VEILLE TWITTER



Plus d'information :

<https://blog.dataiku.com/guest-blog-predicting-connections-to-isis-based-on-tweet-content>

PRÉDICTION DE FRAUDES DANS LA SANTÉ



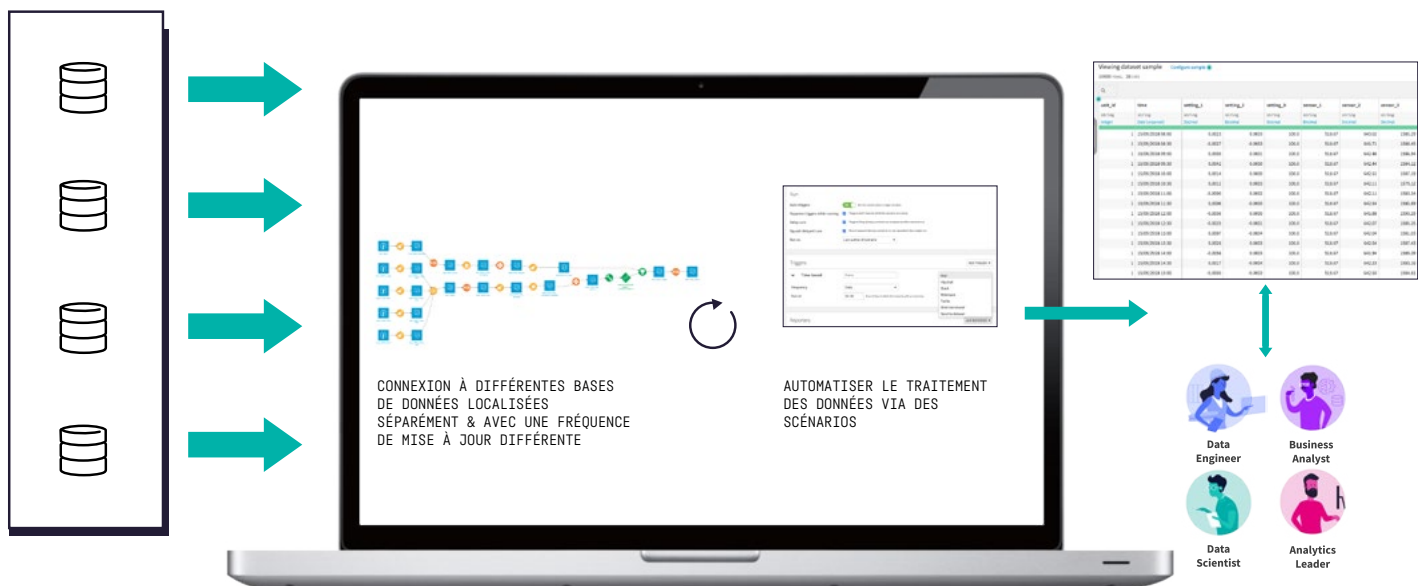
LES 10% DES CAS LES PLUS SUSPECTS SONT DÉTECTÉS 3 FOIS MIEUX QU'ACTUELLEMENT

Statistique

Les domaines de la statistique sont tout particulièrement concernés par les enjeux d'intelligence artificielle et de transparence de la donnée : comment faire en sorte de fournir une information fiable, lisible, exploitable par des publics plus larges qui fondent certaines de leurs décisions sur des indicateurs précis ?

Dans le secteur de la statistique, l'enjeu se situe autant au niveau des cas d'usage (prédiction de l'évolution d'un indicateur économique sur la base d'analyse de séries statistiques temporelles, par exemple) qu'au niveau de la méthodologie de travail : aux nombreux statisticiens qui ont des connaissances en R, Dataiku DSS fournit une infrastructure permettant de mutualiser et de gérer collaborativement du code, afin de démultiplier la productivité des équipes.

AUTOMATISER L'AGRÉGATION DE DONNÉES POUR FACILITER L'ACCÈS À L'ANALYTICS EN LIBRE-SERVICE



METTRE À DISPOSITION L'ANALYTICS EN LIBRE-SERVICE
ET À GRANDE ÉCHELLE POUR 2 000 UTILISATEURS

Section III: Conclusion

L'IA AU SERVICE DE L'HUMAIN

En intégrant la data science, le machine learning et, in fine, l'IA dans le secteur public, l'objectif est de multiplier les aptitudes humaines, et non de se substituer à elles. Pour que les efforts mis en œuvre dans ce domaine s'avèrent profitables, nous devons répéter ce message, qui doit être clairement compris et assimilé à tous les niveaux, et dans les deux sens des structures hiérarchiques, puis mis en pratique au quotidien, par tous ceux qui œuvrent dans le secteur public.

Elaborer une stratégie IA véritablement inclusive – impliquant chacun au sein de l'organisation, et non pas divisée en silos, avec les équipes data d'un côté et les data scientists de l'autre –, mais qui soit centrée sur l'humain en valorisant l'apport de chacun et en évitant de s'en remettre aveuglément à des machines pour prendre une décision – telle est la clé pour mettre en œuvre une intelligence artificielle véritablement utile et évolutive.

POUR ALLER PLUS LOIN

Il existe de bonnes pratiques que le secteur public peut adopter pour améliorer son approche de l'IA et assurer sa pérennité :

- 1. Adoption du machine learning.** Les modèles déterministes basés sur des règles statiques (par ex. des enchaînements logiques de type : si $X =$, alors $Y =$) sont rassurants pour l'esprit en raison de leur simplicité, mais ils sont très inférieurs à des modèles plus sophistiqués capables de prendre en compte de nombreuses variables afin d'émettre des recommandations. La modélisation probabiliste et les graphiques de dépendance partielle offrent des recommandations pondérées et démontrent le degré d'interférence des variables spécifiques pour contribuer à la recommandation émise, offrant ainsi transparence dans des modèles mathématiques sinon largement indéchiffrables.elusive mathematical models.

EN SAVOIR PLUS :

Les bases du machine learning : guide illustré pour les non techniciens (en anglais)



2. Adopter l'open source. La confidentialité revêt une importance fondamentale lorsqu'il s'agit des données et des systèmes du secteur public. L'open source offre gratuitement une technologie IA véritablement exceptionnelle : la communauté de développement collaboratif et les multiples apports se traduisent par de meilleures performances logicielles et s'avèrent plus fiables que les seules alternatives privées. Les outils les plus innovants pour élaborer et appliquer les technologies de machine learning sont en open source. Mais quand il s'agit de rendre opérationnels les projets, les organisations ont besoin d'une autre couche propriétaire pour assurer au mieux l'intégration avec leurs applications spécifiques.

3. Investir dans des outils qui génèrent de la valeur..

En matière de données, les bons outils démystifient l'IA et permettent à chacun de comprendre comment les systèmes basés sur l'IA s'adaptent à leurs recommandations et pourquoi. En outre, un environnement central avec des capacités d'audit offre l'expérience collaborative requise pour faciliter l'analyse des données et la prise de décision, en limitant l'accès des utilisateurs aux données les plus sensibles. Ce qui permet un engagement adéquat, sans mettre en péril les exigences de conformité.

EN SAVOIR PLUS :

Pourquoi les entreprises ont besoin des plateformes de Data Science, Machine Learning et IA (anglais)





Your Path to Enterprise AI

Dataiku est un leader mondial dans les plateformes d'Analytique et de Machine Learning, et soutient les efforts d'agilité data des organisations en offrant la possibilité de disposer d'une IA élastique et responsable accessible à tous à l'échelle de l'entreprise. Des centaines d'entreprises utilisent aujourd'hui Dataiku afin de soutenir leurs opérations et d'assurer leur pérennité dans un monde en perpétuelle mutation.

300+
CLIENTS

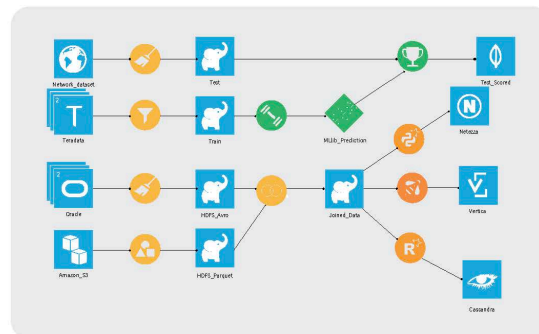
30,000+
UTILISATEURS ACTIFS

*scientifiques de données, analystes, ingénieurs, & plus

1. Nettoyage & data wrangling

Name	Sex	Age
Harvard Mag	Female	Integer
Braynd, Mr. Owen Harris	male	22
Moran, Mr. James	male	38
Haskelton,		36
Hayes, Mr.		26
Allen, Mr. Y		35
McCarthy,		
McCarthy,		
Hayes, Mr. Y		29

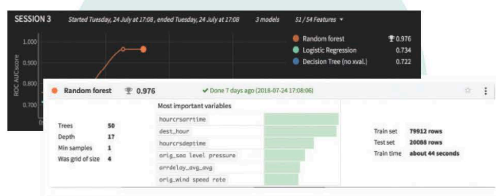
Remove rows containing NA
Keep only rows containing NA
Split column on NA
Replace NA by ...
Reset rows equal to Moran, Mr. James
Keep only rows equal to Moran, Mr. James
Clear cells equal to Moran, Mr. James
Filter on Moran, Mr. James
Filter on NA
Toggle row highlight
Show complete value



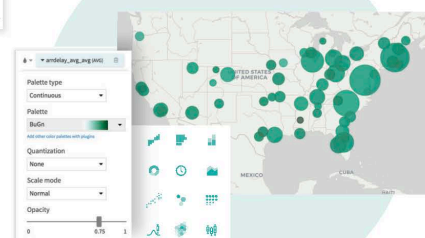
5. Contrôle & ajustement



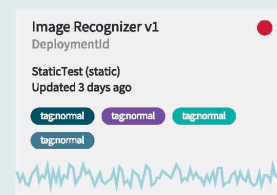
2. Élaboration + application du machine learning



3. Mining & visualisation



4. Déploiement en production





WHITE PAPER

www.dataiku.com