

MACHINE LEARNING :

ATTENTION AUX SOLUTIONS CLÉ EN MAIN

AVIS D'EXPERT



Bouzid AIT AMIR, Team Lead Data Science | Keyrus



Khalil EL MAHRSI, Tech Lead Machine Learning & Intelligence Artificielle | Keyrus

Automatiser la sélection & le paramétrage d'un modèle de Machine Learning grâce à l'AutoML¹ : l'idée est séduisante. Mais dans la pratique, la préparation des données reste une étape clé et la plus chronophage du processus de Data Science qui reste hors du périmètre de l'AutoML.

Le Machine Learning est, aujourd'hui, de plus en plus considéré comme la clé du succès pour un projet de Data Science.

De quoi s'agit-il ? Concrètement, le Machine Learning est une branche de l'Intelligence Artificielle dans laquelle les algorithmes fournissent au système informatique une nouvelle capacité : accomplir, par apprentissage, des tâches pour lesquelles il n'a pas été explicitement programmé. Exemple ? Vinci Energies a développé en 2017 un outil destiné à assurer une maintenance prédictive des structures métalliques (pylônes électriques, éoliennes...). En lui présentant une série de photos de structures portant des traces de corrosion plus ou moins sévères, l'équipe projet a permis au système d'apprendre à reconnaître une corrosion légère, moyenne ou importante et à recommander une intervention. Autre exemple : l'éditeur de logiciels Adobe développerait en ce moment, avec l'université du Maryland, un algorithme censé détecter les images retouchées. Ce sont également des algorithmes de Machine Learning qui sont implantés dans les véhicules autonomes pour apprendre à reconnaître les obstacles et à réagir en conséquence.

Le développement actuel de ces méthodes s'appuie sur la progression continue des capacités de calcul mais également sur la croissance exponentielle de son carburant privilégié : la data. On estime que 90 % des données informatiques, aujourd'hui disponibles dans le monde, ont été créées au cours des quelques dernières années. D'où un faisceau de facteurs favorables au développement rapide de projets de Data Science s'appuyant sur du Machine Learning.

UN PRINCIPE PROCHE DU CLOUD

Pour répondre à la demande croissante d'entreprises de différentes tailles, ne disposant pas toutes de compétences

en Data Science ou des moyens matériels en interne (serveurs de calculs, GPUs...), de grands noms comme Amazon, Google, IBM ou Microsoft proposent des plateformes de Machine Learning accessibles en ligne, selon un principe calqué sur celui du Cloud Computing et du SaaS. On parle de MLaaS (Machine Learning as a Service) et le concept a trouvé rapidement son public. Selon une étude publiée en septembre 2017 par Infoholic Research, le marché du MLaaS devrait connaître une croissance annuelle mondiale de 49 % jusqu'à 2023.

Aujourd'hui, le MLaaS trouve un prolongement dans l'AutoML. Là où le MLaaS propose plusieurs services tels que des modèles de Machine Learning sur étagère ou des plateformes de développement & de déploiement de modèles, l'AutoML automatise le processus de sélection & de validation d'un modèle de Machine Learning, jusqu'à en faire un Workflow (ou flux de travaux) clé en main, voulu accessible et exploitable par des non spécialistes. Parmi les acteurs les plus emblématiques de l'AutoML, Google figure en première place avec son service Cloud AutoML. Dans la même veine, une série de start-ups proposent d'automatiser tout ou partie d'un projet de Data Science grâce à l'AutoML.

PRÉPARATION DE DONNÉES & CHOIX D'UN MODÈLE

Pour autant, une réalité s'impose : un projet de Data Science va bien au-delà de la seule évaluation d'un modèle de Machine Learning adapté à la tâche à accomplir. Et la place de l'AutoML dans un projet doit être évaluée avec réalisme. Pour y voir clair, il est important de bien comprendre les différentes étapes d'un projet de Data Science.

La première porte sur l'expression du besoin, c'est-à-dire cerner la problématique à traiter et déterminer comment on se propose d'y répondre. Une fois ce besoin clairement

¹ L'autoML consiste à automatiser les différentes étapes de construction d'un modèle de Machine Learning (comparaison de différents algorithmes, optimisation des hyperparamètres, sélection du meilleur modèle, etc.).

défini, la seconde étape consiste à identifier le périmètre de données sur lesquelles le projet va s'appuyer. Il faut également qualifier les informations et s'assurer qu'elles sont réellement exploitables & pérennes (des données obtenues ponctuellement ne peuvent alimenter un outil destiné à être utilisé sur la distance).

Troisième étape : le traitement et l'enrichissement des données. On parle de Feature Engineering, étape souvent réalisée itérativement entre le data scientist et les métiers. Il s'agit de transformer la donnée pour la rendre exploitable. Il est rare qu'un projet de Data Science opère à l'aide de données brutes. C'est seulement après que l'on passe à l'étape charnière de la sélection d'un modèle de Machine Learning. Le but ? Trouver le juste équilibre entre la performance pure du modèle et la facilité avec laquelle il pourra être industrialisé.

Ce dernier point est essentiel. Le produit fini est en effet un outil clé en main, livré sous forme d'API ou directement intégré à son système d'information et pleinement exploitable par ses collaborateurs. C'est ainsi que travaille l'équipe de 30 data scientists dont dispose **Keyrus** en Île-de-France.

LA PARTIE CACHÉE DE L'ICEBERG

On l'aura compris, l'AutoML ne peut intervenir que sur la partie aval du projet pour automatiser le test des modèles, l'évaluation de leur performance au regard de la tâche envisagée et la qualification de leur output. Or, ce n'est ni la partie la plus importante du projet, ni l'opération la plus délicate. L'AutoML laisse hors de son périmètre les étapes de Data Preparation, Data Management, Data Qualification et Feature Engineering, parties les plus déterminantes d'un projet et qui représentent à elles seules 60 à 70 % de l'effort total. C'est pourquoi un projet de Data Science ne peut se limiter aujourd'hui à la collecte de données brutes à soumettre à une API d'AutoML. La préparation de données nécessite une intervention humaine s'appuyant sur de véritables compétences. Réussir un projet de Data Science exige donc l'intervention d'un conseiller possédant une vision à 360 degrés du projet et capable d'actionner les bons leviers (AutoML compris) au bon moment.

Dans l'état actuel des technologies, se fier à 100 % à un moteur AutoML est certainement la meilleure façon de passer à côté de son projet.

K.E.M.

B.A.A.

À PROPOS DES AUTEURS

Khalil EL MAHRSI

Khalil est Docteur en Informatique et spécialiste en Machine Learning & Deep Learning. Avec plus de 7 ans d'expérience, il a contribué à la réalisation de bout en bout de plusieurs projets de Data Science au déploiement des solutions adéquates pour le stockage des données, en passant par la mise en place d'algorithmes d'apprentissage automatique, jusqu'au développement de visualisations de données. Tech Lead Machine Learning & Intelligence Artificielle chez **Keyrus** depuis un an, il accompagne ses clients sur leurs projets en IA et apprentissage automatique.

Bouid AIT AMIR

Après une formation en économie mathématiques à la Toulouse School of Economics puis en informatique appliquée à Télécom Bretagne, Bouid a intégré le département d'allocation stratégique d'une banque de gestion d'actifs puis une ESN spécialisée en optimisation et maintenance prédictive dans l'aéronautique et le ferroviaire.

Chez **Keyrus** depuis 4 ans en tant que Team Lead Data Science, Bouid y a créé et développé une practice Data Science. Aujourd'hui, il conseille et accompagne ses clients sur la réalisation de projets de Data Science, d'Intelligence Artificielle ainsi que des choix technologiques qui en découlent.

À PROPOS DE KEYRUS

Keyrus, créateur de valeur à l'ère de la Data et du Digital

Acteur international du conseil et des technologies, spécialiste de la Data et du Digital, **Keyrus** a pour mission d'aider les entreprises à tirer profit du paradigme de la Donnée et du Numérique pour accroître leur performance, faciliter et accélérer leur transformation et générer de nouveaux leviers de croissance, et de compétitivité.

Plaçant l'innovation au cœur de sa stratégie, **Keyrus** développe une proposition de valeur unique sur le marché autour d'une offre novatrice qui s'appuie sur la combinaison de trois expertises majeures et convergentes :

- **Data Intelligence**

Data Science – Intelligence Artificielle – Big Data & Cloud Analytics – Business Intelligence – EIM – CPM/EPM

- **Digital Experience**

Innovation & Stratégie Digitale – Marketing Digital – DMP & CRM – Commerce Digital – Performance Digitale – User Experience

- **Conseil en Management & Transformation**

Stratégie & Innovation – Transformation Digitale – Pilotage de la Performance – Accompagnement des Projets

Présent dans une quinzaine de pays et sur 4 continents, le Groupe **Keyrus** emploie plus de 3 000 collaborateurs.

Keyrus est coté sur le compartiment C de l'Eurolist d'Euronext Paris (Compartiment C/Small caps – Code ISIN: FR0004029411 – Reuters : KEYR.PA – Bloomberg : KEY:FP)

Plus d'informations sur : www.keyrus.fr