

Data Scientist

Vous interviendrez sur nos projets clients afin d'analyser, consolider et exploiter efficacement les données mises à disposition, dans le cadre d'une problématique métier et dans un contexte opérationnel.



Artelys est une entreprise en forte croissance spécialisée en **optimisation**, **aide à la décision** et **modélisation**. Grâce à une expertise de haut niveau dans les techniques quantitatives, ses ingénieurs et consultant(e)s conçoivent et mettent en œuvre des solutions dans des secteurs d'activité diversifiés : **énergie**, logistique & **transport**, télécommunications, finance et défense. Artelys est une équipe dynamique de 50 personnes principalement basée à Paris.

A travers son activité soutenue de R&D et ses projets clients orientés données, Artelys exploite toute la filerie de la data-science, de la fouille approfondie des données à l'implémentation efficace de modèles, en passant par le développement de solutions innovantes de visualisation et d'exploration.

Descriptif du poste

Au sein d'une équipe dynamique de haut niveau, vous interviendrez conjointement sur des projets de R&D et clients, centrés sur l'**analyse** et la **consolidation** de données (de grande et de petite taille), la conception et l'implémentation de méthodes de **modélisation** performantes, ainsi que la **restitution visuelle** de ces traitements.

En particulier, vous serez amené(e) à :

- Explorer des jeux de données bruts (potentiellement très volumineux et de sources disparates), implémenter des techniques de visualisation pour en intuitier le sens, appliquer des méthodes de consolidation afin de leur donner une structure robuste servant de base à la modélisation.
- Implémenter et tester un large spectre des solutions de modélisation (des statistiques classiques aux techniques modernes du **machine learning**), afin d'en exploiter la meilleure combinaison.
- Développer des solutions de visualisation et d'exploitation des résultats dans un contexte opérationnel.

Quelques exemples de projets chez Artelys :

Projet de R&D d'envergure de l'Agence de l'Environnement de la maîtrise de l'énergie (ADEME), **NEXT** vise à développer un logiciel innovant de référence pour le dimensionnement des réseaux électriques flexibles (**smart-grid**), apportant ainsi une brique essentielle à la **transition énergétique et écologique**.

Au cœur du projet NEXT se trouvent la compréhension et la modélisation du processus de consommation électrique à une échelle temporelle et spatiale très fine. Des techniques modernes de **clustering** permettent notamment de mettre en évidence des caractéristiques communes de consommation. D'autre part, ce projet est l'occasion de mettre en oeuvre et de tester de nombreuses techniques de **modélisation prédictive**, des classiques modèles GAM aux techniques modernes du machine learning, et en particulier les Generative Adversarial Networks et les Variational AutoEncoders.

Partenaires : INRIA, Gaz Electricité Grenoble, L2EP

<http://www.ademe.fr/next>

Pour le compte d'un acteur majeur du secteur pétrolier, nous développons une application web de restitution et d'exploration des résultats d'un traitement statistique complexe. Basé sur le framework Python Dash (Flask, Plotly.js, React.js), l'application met en oeuvre, dans un contexte de très grande dimension, une technique de régression parcimonieuse (linéaire ou non) permettant de sélectionner les variables les plus impactantes et restitue les résultats sous la forme dashboard de graphes interactifs.

Appelé à la demande, l'exécution de l'algorithme est transparente pour l'utilisateur qui souhaite analyser les résultats sans prérequis de compréhension des techniques mathématiques implémentées.

Profil recherché

Titulaire d'un master et/ou d'un doctorat avec une spécialisation dans le domaine des data-sciences et des statistiques, vous possédez une double compétence en mathématiques et informatique, notamment scientifique (R, Matlab, Python).

Vous êtes le candidat idéal si :

- Vous êtes curieux et vous abordez la fouille d'une nouvelle source de données comme un véritable jeu.
- Vous maîtrisez et avez implémenté des méthodes de clustering, de classification et de regression.
- Vous considérez que la compréhension du contexte métier associé à une problématique data est indispensable à sa résolution.
- Animé par un pragmatisme prononcé lors de vos choix de modélisation, vous savez que le meilleur modèle n'est pas forcément le plus compliqué.
- Vous êtes conscient que l'univers du « Small Data » présentent des enjeux et des défis à la hauteur de ceux du « Big Data ».
- Vous êtes autonome, rigoureux et animé d'un fort esprit d'équipe.

Sans être un prérequis obligatoire, la connaissance des environnements et outils de Big Data (Hadoop, Hive, Spark, MapReduce, Bigtable, NoSQL) est un plus.

En outre, une bonne maîtrise de l'anglais sera fortement appréciée.

En **forte croissance**, Artelys est à la recherche de collaborateurs proactifs et créatifs, soucieux de la valeur ajoutée de leurs travaux et motivés pour prendre part au développement de nos activités.