

OFFRE D'ALTERNANCE

Alternant(e) Bac +5 en data science F/H

Durée : 12 ou 24 mois

Date de publication : 11/04/2024

Lieu : Verneuil-en-Halatte (60) - accessible en transports en commun avec navette de bus privée, à 40 mn au Nord de Paris

Type de contrat : Alternance

Contact : Pour postuler, cliquez [ici](#)

DESCRIPTIF

L'Ineris (Institut national pour l'environnement industriel et des risques), qui compte environ 500 collaborateurs, est un organisme national de référence, sous tutelle du ministère chargé de l'environnement, dont la mission principale est de réaliser des études et des recherches permettant de prévenir les risques que les activités économiques font peser sur la sécurité des personnes et des biens.

Au sein de la direction Milieux et impact sur le vivant (MIV), l'unité d'expertise en toxicologie et écotoxicologie des substances chimiques (ETES), composée d'une douzaine de personnes, a la charge de la veille sur les propriétés de devenir environnemental, de toxicité et d'écotoxicité des substances, ainsi que des évolutions réglementaires liées à ces substances.

Dans le cadre de la mise à disposition de son nouveau [portail sur les substances chimiques](#) et des nouvelles approches méthodologiques faisant intervenir la modélisation, l'équipe recherche un(e) alternant(e) en traitement de l'information et mégadonnées.

Sous la responsabilité d'un ingénieur de l'unité, et avec l'appui des membres de l'unité, vous contribuerez à la veille, la collecte, la préparation, la vérification de la cohérence et la mise à disposition des données dans le portail.

Vous pourrez être amené(e) à travailler sur les tâches suivantes :

- Exploration de données et de textes (Data Mining and Text Mining) :
 - o Utilisation de techniques d'exploration de données pour extraire des informations pertinentes à partir de grands ensembles de données ;
 - o Exploration de textes pour extraire des informations de la littérature scientifique, des rapports et des documents réglementaires ;

- Apprentissage automatique et intelligence artificielle (Machine Learning and Artificial Intelligence) : mise en œuvre d'algorithmes d'apprentissage automatique pour classer, analyser et prédire les données d'(éco)-toxicité et de devenir environnemental. Cela inclut des techniques telles que la classification, la régression, le clustering et le traitement du langage naturel ;
- Création de bases de données, standardisation et harmonisation des données : normaliser les formats de données et harmoniser les données dans différentes bases de données et sources pour garantir la cohérence et la compatibilité pour une analyse et une évaluation plus approfondies ;
- Outils d'extraction de données automatisés : utilisation d'outils et de scripts automatisés pour extraire efficacement les données des bases de données, des rapports et de la littérature, réduisant ainsi les efforts manuels et garantissant l'exhaustivité des données ;
- Contrôle et assurance qualité : mettre en œuvre des mesures de contrôle et d'assurance qualité pour valider l'intégrité des données, identifier les erreurs ou les incohérences et les rectifier avant une analyse plus approfondie ;
- Intégration et visualisation des données : intégrer des données organisées provenant de diverses sources et les visualiser à l'aide de représentations graphiques pour faciliter l'interprétation, identifier des modèles et en tirer des informations significatives.

PROFIL

Etudiant(e), vous préparez un master ou un diplôme d'ingénieur en science des données. Vous vous intéressez aux enjeux liés aux impacts de la contamination chimique sur la santé humaine et l'environnement.

Par ailleurs, vous disposez des aptitudes suivantes :

- Connaissance de l'anglais ;
- Grande rigueur ;
- Sens de l'organisation ;
- Goût pour le travail en équipe ;
- Sens pratique et excellente qualité relationnelle.