

DATA SCIENTIST (EXPERT EN DONNÉES)

Contrat : CDD de 1 an

Salaire : Selon le barème

**Lieu de travail : Université de Limoges, laboratoire VieSanté, Institut
d'appartenance Omega-Health**

L'unité VieSanté est une unité de recherche émergente dédiée à l'étude de la fragilité médico-psycho-socio environnementale de la personne âgée et l'apport des outils numériques pour optimiser le parcours de santé des personnes âgées. L'originalité de la démarche scientifique de l'unité VieSanté réside dans une approche centrée sur les personnes âgées dans le cadre de la « vie réelle », c'est-à-dire sur leur lieu de vie et dans les réseaux de soins gériatriques en prenant en compte la problématique des déserts médicaux ou assimilés (domicile ou en institution).

Elle développe plusieurs thématiques pertinentes et originales comme celle centrée sur les parcours dans le champ de la prise en charge des personnes et celle centrée sur les médicaments inappropriés chez le sujet âgé vers la fragilité pharmacologique. L'unité VieSanté réunit des complémentarités permettant, dans une approche multidisciplinaire, le développement d'une recherche originale sur les déterminants et facteurs de décompensation de la fragilité ou sur une intervention innovante dans le parcours de santé des personnes âgées.

Les collaborations nouées intègrent dans cette analyse d'optimisation de parcours, les gérontotechnologies et l'intelligence artificielle, en s'appuyant sur des acteurs de terrain en proximité du lieu de vie de la population cible.

ACTIVITÉS PRINCIPALE :

- **Collecte et préparation des données** : exploiter des données des « sociétaires » de l'Assurance Retraite (données publiques et données de santé) : nettoyage des valeurs manquantes, normalisation des variables
- **Traitement des aspects éthiques et réglementaires** : anonymisation et respect du RGPD
- **Analyse exploratoire** : identifier les corrélations entre variables (ex. : âge et risque cardiovasculaire), détection des biais (ex. : sur-représentation d'un groupe), visualisation des distributions pour comprendre les patterns
- **Sélection des variables explicatives** : variables les plus pertinentes pour construire le modèle seront sélectionnées à l'aide de méthodes de sélection automatique et statistique, telles que la sélection pas à pas (forward, backward ou stepwise), la régularisation (LASSO, Ridge, Elastic Net) ou encore des méthodes basées sur l'importance des variables (par exemple via les arbres de décision ou les forêts aléatoires). La sélection tiendra compte de la multicollinéarité entre variables en éliminant celles fortement corrélées afin d'assurer la robustesse et l'interprétabilité du modèle.
- **Choix du modèle et des algorithmes** : le choix du modèle se fera en fonction des caractéristiques et de la configuration des données collectées (nature des variables, taille de l'échantillon, présence de valeurs manquantes, etc.). Différentes familles de modèles (régression, arbres de décision, réseaux de neurones, modèles ensemblistes, etc.) seront évaluées selon leurs performances et leur capacité d'interprétation. L'objectif final sera de mettre en place un modèle reposant sur les principes de l'intelligence artificielle explicative (XAI) permettant de comprendre et justifier les décisions ou prédictions du modèle à l'aide d'outils tels que SHAP, LIME ou les importances de variables.
- **Entraînement du modèle** : l'entraînement du modèle sera réalisé à partir de la majorité des données disponibles afin d'assurer un apprentissage suffisant des schémas et des relations présentes dans les données. La répartition des données se fera selon une logique adaptée au volume d'observations et au nombre de participants, par exemple en jeu d'entraînement et de test (train/test) ou en jeu d'entraînement, de validation et de test (train/validation/test).
- **Évaluation du modèle (prototype)** : l'évaluation du modèle sera d'abord réalisée sur l'ensemble de test, afin de mesurer sa performance sur des données indépendantes de l'entraînement. La robustesse et la *généralisabilité* du modèle pourront être vérifiées et consolidées à l'aide d'un nouvel ensemble de données provenant d'autres participants, mais comportant les mêmes variables d'entrée. Cette étape permettra de confirmer la capacité du modèle à reproduire ses performances sur des données réelles et inédites.
- **Implémentation et suivi** : l'implémentation du modèle débutera dès la collecte des premières données. En effet, selon le nombre de participants et la taille de l'échantillon, certains types de modèles pourront s'avérer plus adaptés que d'autres, notamment en termes de complexité et de capacité de généralisation.

ACTIVITÉS SECONDAIRES :

- Participer au rayonnement du Laboratoire
- Jouer un rôle de conseil auprès du directeur du Laboratoire VieSanté

COMPÉTENCES REQUISES

- Le data scientist doit avoir un goût prononcé pour le travail d'équipe.
- Expertise en gérontologie indispensable

Compétences techniques

- Mathématiques & statistiques : Probabilités, statistiques inférentielles, modèles statistiques (régressions, séries temporelles), optimisation, algèbre linéaire,
- Machine Learning & IA : apprentissage supervisé (régression, SVM, random forest...), apprentissage non supervisé (clustering, PCA, anomalies...), deep Learning (CNN, RNN, transformers), validation des modèles (cross-validation, métriques, surapprentissage),
- Programmation : Python (indispensable) : NumPy, Pandas, Scikit-learn, TensorFlow/PyTorch, R, automatisation & scripting,
- Gestion et manipulation des données : SQL, manipulation de grands volumes de données (Pandas, PySpark), connexion à des bases,

Compétences analytiques

- Analyse exploratoire des données (EDA) : comprendre la structure des données, détecter anomalies, biais, valeurs manquantes, identifier les insights clés
- Visualisation & dataviz : Matplotlib, Seaborn, Plotly, Power BI ou Tableau

Savoir être :

- Curiosité scientifique
- Rigueur analytique
- Travail en équipe (data engineers, professionnels de santé, agents publics...)
- Gestion de projet & autonomie
- Responsabilité, discrétion et confidentialité, capacité d'organisation et gestion des priorités
- Capacités à travailler en autonomie et de manière transversale
- Sens du dialogue et de l'écoute, capacité à rendre compte
- Disponibilité, capacités d'adaptation

LIAISON HIERARCHIQUE

- Le Directeur du laboratoire VieSanté : Professeur Achille TCHALLA

INFORMATIONS SUPPLÉMENTAIRES / SPÉCIFICITÉS / CONTRAINTES LIÉES AU POSTE

- Participation aux instances de gouvernance du projet (horaires atypiques).
- Déplacements à prévoir (en région)

Merci d'adresser votre candidature au Pr Tchalla : achille.tchalla@unilim.fr

et son secrétariat : laura.reis@chu-limoges.fr