

Machine Learning Engineer (ingénieur en apprentissage automatique)

Structure d'accueil

L'activité s'exerce au sein de l'UMR1252 SESSTIM, Sciences Economiques et Sociales de la Santé & Traitement de l'Information Médicale, à la Faculté des Sciences Médicales et Paramédicales de la Timone, à Marseille. L'UMR SESSTIM s'efforce de produire une recherche d'excellence, pluridisciplinaire et interdisciplinaire, en sciences humaines et sociale et en santé publique pouvant conduire à des changements dans les différents champs de la médecine prédictive, personnalisée, préemptive et participative. Les chercheurs du SESSTIM développent, ou sont associés à des travaux de recherche tentant d'apporter des réponses pour faire face à des enjeux actuels pour la société et ses populations et contribuent à des développements méthodologiques et à des avancées méthodologiques. En termes de pathologies d'intérêts, nos travaux de recherche portent principalement sur la maladie cancéreuse et les maladies infectieuses et transmissibles. Nos questionnements portent sur des facteurs individuels, populationnels ou contextuels. Nos travaux ciblent des populations de France, ou plus largement du Nord, mais aussi du Sud, que ce soit au niveau du bassin méditerranéen ou de l'Afrique subsaharienne, principalement.

Missions principales

Le/La candidat-e travaillera au sein de l'équipe multidisciplinaire « Méthodes Quantitatives et Traitement de l'Information Médicale ([QuanTIM](#)) », composée de chercheur-se-s en épidémiologie et santé publique, statisticien-ne-s, biostatisticien-ne-s, informaticien-ne-s et data scientists. Il/Elle sera plus spécifiquement affecté-e à un projet d'applications de techniques d'Intelligence Artificielle aux données de la [Plateforme de Données de Cancérologie](#), qui est une extraction du Système National des Données de Santé ([SNDS](#)). Ce projet financé par l'Institut National du Cancer implique des membres de l'équipe « Cancer, Biomédecine & Société ([CanBIOS](#)) », ainsi qu'un médecin du Département d'oncologie chirurgicale de l'Institut Paoli Calmettes. L'objectif du travail sera de **prédire la survenue de complications post-opératoires après une tumorectomie ou une mastectomie dans le cadre du cancer du sein.**

Activités

L'INCa a mis en place l'entrepôt de données intitulé « **Plateforme de données en cancérologie** ». Cette base médico-administrative est accompagnée d'algorithmes déjà développés et a donné lieu à des publications innovantes en raison de la quasi-exhaustivité des données. Selon la provenance des données (DCIR ou PMSI), il peut y avoir peu de tables mais très volumineuses ou beaucoup de tables mais peu volumineuses. Le/La candidat-e devra acquérir les connaissances de cette base. A cette fin, des formations sont prévues, notamment pour découvrir l'architecture de ces données. Le/La candidat-e bénéficiera de la présence d'une documentation interne à la plateforme pour comprendre les tables supplémentaires créées par l'INCa. Il/Elle travaillera à distance sur la plateforme sécurisée dans une « bulle ». En tant qu'ingénieur en apprentissage automatique, il/elle sera en charge du preprocessing des données comprenant : 1) la mise en place de requêtes pour **l'extraction des variables** en collaboration avec les chercheur-se-s impliqué-e-s dans le projet ; 2) l'examen des données pour déceler d'éventuelles incohérences ou anomalies ; et 3) la vérification de la complétude des données et de leur précision. Le/La candidat-e devra gérer le **développement d'un modèle de deep learning** afin de prédire la survenue de complications post-opératoires après une tumorectomie ou une mastectomie dans le cadre du cancer du sein. Il/Elle attachera une importance tout particulière à **l'interprétation du modèle** (technique SHAP ou autre) avec un focus sur les traitements (ou associations de traitements) associés aux survenues de ces complications. Pour cela, il/elle travaillera en étroite collaboration avec notre équipe multidisciplinaire. Le/La candidat-e contribuera à la rédaction d'un

article sur ce travail en vue d'une publication dans une revue internationale avec comité de lecture.

Compétences

Savoir

- Solides connaissances dans les techniques de machine learning et de deep learning ;
- Expertise dans l'interprétation des modèles (e.g., technique SHAP) ;
- Bonnes connaissances opérationnelles en pgSQL ou SQL ;
- Gestion de grandes bases de données ;
- Maîtrise de Python et/ou R, des compétences en SAS seraient un atout.

Savoir-faire

- Excellent sens de l'organisation ;
- Rigueur ;
- Aptitude à travailler dans un environnement dynamique et à respecter les délais ;
- Capacité à collaborer avec des équipes multidisciplinaires ;
- Être force de proposition.

Compétences linguistiques

- Anglais : niveau scientifique

Niveau de diplôme et expérience

- Bac +5 ou plus en data science, ou biostatistiques.
- Minimum 1 an d'expérience dans un poste similaire.

Environnement et formations

Environnement professionnel - Lieu d'exercice

L'activité s'exerce au sein de l'UMR1252 SESSTIM, Sciences Economiques et Sociales de la Santé & Traitement de l'Information Médicale, à la **Faculté des Sciences Médicales et Paramédicales de la Timone, à Marseille**.

Contrat

Début : Dès que possible, selon les délais administratifs de recrutement

Durée : 12 mois

Rémunération : Niveau ingénieur.e d'étude avec expérience ou ingénieur.e de recherche ; grille de salaire d'Aix-Marseille Université.

Candidature :

SVP envoyer à roch.giorgi@univ-amu.fr et nathalie.graffeo@univ-amu.fr le dossier constitué

- D'une lettre de motivation expliquant comment le/la candidat.e pense qu'il/elle pourra contribuer à ce projet ;
- D'un curriculum vitae.

Date de fin de candidature : 31/01/2024.

Les entretiens seront réalisés par visio ou à Marseille.