Ingénieur données de santé

TL ; DR : Si tu veux manger des données pour permettre à la recherche médicale de faire un bond en avant, tu es au bon endroit ☺

Qui sommes-nous ?

Le CHU de Toulouse s’est doté d’une équipe de gestion de la donnée de santé afin d’optimiser la prise en charge des patients. Cette modernisation permet d’envisager une personnalisation de la prise en charge et une accessibilité accrue à la recherche clinique. Par exemple, l’analyse systématisée des images de routine permettrait d’apporter aux experts des indicateurs quantifié pour suivre la maladie. L’équipe Data du CHU de Toulouse offre donc un support technique aux investigateurs souhaitant explorer les possibilités des modèles de machine learning appliqués aux données cliniques.

Dans ce cadre nous avons lancé un programme d’uniformisation de nos données de santé. Cette uniformisation permettra de faciliter le travail des datascientists en leur fournissant des données propres et nettoyées.

C’est quoi le projet ?

Tu intègreras une équipe émergente composée d’informaticiens, de chercheurs, d’ingénieurs et de médecins, riche d’une forte expérience acquise au cours de nombreux projets de recherche ! Tu pourras contribuer à l’amélioration des pratiques hospitalières et au développement de nouvelles approches de traitement des données. Cette mission s’inscrit dans le cadre d’un développement rapide de l’activité « data santé » du site toulousain.

L’enjeu de ce projet est de fournir à l’institution, à la fois une expertise sur les techniques de manipulation de la donnée et de la modélisation, de mettre en place une stratégie institutionnelle mais aussi de mettre la main à la pâte sur certains projets ciblés pour aller chercher des financements ultérieurs. Il faudra donc se familiariser avec différents types de données pour mettre en place des datamarts et des datasets pertinents aux recherches en cours. Les projets sont divers sur tous types de données. Il y a bien évidemment des grosses bases de données pas très propres, mais aussi des données plus spécifiques telles que les données d’imageries, les données génomique ou autres… Il faudra donc apprendre à nettoyer, analyser, entrainer, modéliser les différentes données et ce dans un environnement sécurisé.

*Missions*

* Analyse du besoin métier et des contraintes spécifiques au monde de la santé
* Création de schémas de donnée adaptés
* Développement d’ETL pour aller chercher les données aux bons endroits
* Développement de librairies de traitement de données spécifiques

L'objectif est de fournir des approvisionnements en données qui soient efficaces, stables et maintenables.

Tu travailleras directement avec les médecins pour évaluer et faire évoluer ton travail.

Les travaux seront réalisés directement sur des données patient, dans un environnement de calcul sécurisé adapté.

Qui es-tu ?

**Tu es :**

* Curieux.se
* Autonome
* Master ou Ingénieur en mathématique/informatique
* Ton réseau social favori : Stackoverflow ou Github
* Tu parles python couramment
* Le SQL est ton meilleur ami
* Tu n’as pas peur d’apprendre pleins de nouveaux outils
* Tu as déjà cassé un repo git, et tu ne le referas plus jamais
* Tu adores passer du temps sur wikipédia à apprendre des concepts médicaux bizarre
* Tu sais expliquer le concept d’index à un enfant de 5 ans

Points bonus :

* Tu es un.e dieu/déesse des ETL / ELT
* Les bases NoSQL sont aussi tes potes
* T’es capable de citer les options des fonctions Unix de tête
* Tu as participé à des projets Open-source
* Tu connais l’API Pandas comme ta poche
* Tu préfères vi à emacs (ou l’inverse)
* Tu as déjà entendu les termes ICD10 / SNOMED-CT / FHIR / OMOP