

Scientifique de données/Data Scientist

Le Centre d'intégration et d'analyse des données médicales (CITADEL) du Centre hospitalier universitaire de l'Université de Montréal (CHUM) s'inscrit dans les initiatives facilitant un CHUM apprenant, enseignant et communiquant. Le projet est né du besoin de structurer les sources de données utiles (cliniques, administratives, financières, rapportées par les patients) afin de les exploiter dans un but de recherche et d'amélioration continue des interventions en santé.

Les principaux objectifs poursuivis par CITADEL :

- Intégrer, vers un bassin de données institutionnel, les informations nécessaires pour améliorer les soins aux patients et augmenter la performance du système de soins.
- Fournir un accès facile, sécurisé, approprié et dans un délai raisonnable aux données clinico-administratives pour promouvoir et faciliter la recherche et l'innovation.
- Fournir un service d'analyse des données médicales.

Vos responsabilités incluront:

- Transformer les données disponibles dans une forme permettant l'analyse statistique des données (par exemple, une observation par ligne par patient ou événement, trajectoires de soins, des formats facilitant la création de tableaux de bord avec des statistiques sommaires).
- Recommander des plans d'analyse des données qui répondront de façon optimale aux objectifs d'utilisation des clients.
 - Être familier avec les stratégies d'apprentissage automatique et d'analyses biostatistiques (voir expertises spécifiques), et savoir dans quelle situation les utiliser.
- En collaboration avec l'équipe de CITADEL, contribuer à développer, valider et gérer le bassin de données institutionnel.

Nous recherchons les expertises spécifiques suivantes:

- Maîtrise ou Doctorat dans un domaine approprié (biostatistique, épidémiologie, génie, mathématiques, ou autres domaines pertinents aux sciences des données, etc.).
- Connaissance de techniques d'apprentissage automatique couramment utilisées, y compris: les modèles de régression, les forêts d'arbres décisionnels (random forest), k-nearest neighbors, gradient boosting, analyse de séries temporelles, les réseaux neuronaux récurrents, convolutionnels pour l'analyse des données d'imagerie médicale, représentations vectorielles de mots tels word2vec et autres stratégies d'analyse de texte libre.
- Connaissance de techniques d'analyse couramment utilisées, y compris: la modélisation prédictive, la visualisation et les statistiques descriptives, les modèles de survie, l'analyse de séries chronologiques, l'analyse des données en grappes, l'inférence bayésienne, l'analyse causale, le calcul de taille d'échantillon, analyse coût-efficacité.
- Compétences démontrées avec les langages de script et autres langages de programmation:
 - Requis : Expérience avec R ou Python.
 - Utile : Bash, C++, Linux, autres.
- Expérience de travail dans le domaine du « Big Data », entrepôts de données, bassin de données ou autres domaines jugés équivalents.
- Compétences démontrées en matière de requêtes et de gestion de bases de données.
 - Requis : SQL de base (ex. PostgreSQL, MySQL, Oracle).
 - Utile : MS Access, NoSQL (ex. MongoDB), données DICOM.

- Expérience dans la mise sur pied et la maintenance d'interface utilisateur pour faciliter l'accès aux données et aux analyses aux usagers non codeurs (ex. Shiny, Tableau).
- Expérience avec les algorithmes et bibliothèques d'apprentissage automatique:
 - Pandas, SciPy, Scikit-learn, Keras, TensorFlow, PyTorch, etc.
- Expérience dans la visualisation et la présentation des données.
 - ggplot2, matplotlib, etc.
- Expérience de travail avec systèmes de collaboration et gestion des versions (git ou autres).
- Expérience et connaissance en matière de cybersécurité un atout majeur.
- Expérience et connaissance dans les méthodes d'analyse distribuées un atout majeur.

Compétences personnelles:

- Habilité marquée à équilibrer l'autonomie et le travail en équipe dans un contexte multidisciplinaire.
- Curiosité intellectuelle et volonté d'apprendre de nouvelles méthodes et technologies.
- Facilité à communiquer avec les clients.

Statut:

- Poste régulier à temps complet, 35 heures par semaine.
- Salaire et avantages sociaux selon les normes du CRCHUM.

Pour poser votre candidature:

Les personnes intéressées doivent faire parvenir une lettre de présentation ainsi que leur curriculum vitae à:

Mme. Christine Bellefeuille, CRHA
Conseillère en ressources humaines sénior
CRCHUM
900 Rue Saint-Denis, R05.448
Montréal, Québec H2X 0A9

Christine.bellefeuille.chum@ssss.gouv.qc.ca

**Toutes les candidatures sont confidentielles*

**English version follows*

Data Scientist

The Center for the Integration and Analysis of Medical Data (CITADEL) at the University of Montreal Health Centre (CHUM) is an integral part of the initiative to further CHUM's presence as a learning, teaching, and communicating health institution. To support this goal, CITADEL was launched to structure CHUM's administrative health data sources (clinical, administrative, financial, and patient reported data) for research and continuous health system improvement initiatives.

The primary goals of CITADEL include:

- Integrating into an institutional data lake the necessary data sources for research and improving patient care and health system performance
- Providing secure, easy-to-use access to clinical-administrative data within a reasonable amount of time to help facilitate research and innovation
- Providing a service for analyzing medical data

Your responsibilities will include:

- Transform available data in a form facilitating data analysis (e.g., one row observation per patient or event, care trajectories, and formats that facilitate the creation of dashboards with summary statistics)
- Recommend and execute analysis plans based on the optimal use of data to meet clients' research objectives;
 - Be familiar with machine learning and biostatistical analysis strategies (see specific expertise), and know in which situations to use them.
- In collaboration with the CITADEL team, contribute to the development, validation and management of an institutional data lake.

We seek candidates with the following expertise:

- Masters or Ph.D. in an appropriate field (biostatistics, epidemiology, statistics, engineering, mathematics, bioinformatics, or other data-science related fields)
- Knowledge of commonly used machine learning techniques, including: regression models, random forest, k-nearest neighbors, gradient boosting, time series analysis, recurrent neural networks, convolutional neural networks for medical imaging data analysis, vector representations of words such as word2vec and other free text analysis strategies.
- Knowledge of commonly used analytical techniques, including: predictive modelling, visualization and descriptive statistics, survival models, time-series analysis, clustered data analysis, Bayesian inference, causal analysis, sample size calculation, cost-effectiveness analysis.
- Demonstrated skills with scripting and other programming languages:
 - Required: Experience with R or Python.
 - Useful: Bash, C++, Linux, others.
- Experience working in the field of "Big Data", data warehouses, data lakes or other areas deemed equivalent.
- Demonstrated skills in querying and database management.
 - Required: Basic SQL (e. g. PostgreSQL, MySQL, Oracle).
 - Useful: MS Access, NoSQL (e. g. MongoDB), DICOM data.
- Experience in setting up and maintaining user interfaces to facilitate access to data and analysis for non-programmers (e. g. Shiny, Table).

- Experience with machine learning algorithms and libraries:
 - Pandas, SciPy, Scikit-learn, Keras, TensorFlow, PyTorch, etc.
- Experience in visualizing and presenting data.
 - ggplot2, matplotlib, etc.
- Experience working with collaboration systems and version management (git or other).
- Experience and knowledge of cybersecurity is a major asset.
- Experience and knowledge of distributed analysis methods a major asset.

Other Relevant Skills:

- Superior ability to balance autonomy and teamwork in a multidisciplinary context
- Intellectual curiosity and willingness to learn new methods and technologies
- Good client management skills

Statut:

- Regular, full-time position (35 hours per week)
- Salary and benefits according to CRCHUM standards

To apply for this position:

Interested candidates should submit their resume and a cover letter to:

Ms. Christine Bellefeuille, CHRP
Human Resources Advisor
CRCHUM
900 Rue Saint-Denis, R05.448
Montreal, Quebec H2X 0A9

Christine.bellefeuille.chum@ssss.gouv.qc.ca

**All applications will be confidential.*