



## **DATA ENGINEER / ANALYSTE (F/H) – CDD 1 an – Lyon**

Depuis 2005, HEVA est leader en France dans le traitement et l'analyse des données de santé en vie réelle, et plus particulièrement celles issues des bases du Système National des Données de Santé (SNDS) et des bases en OPEN DATA. Nous élaborons et produisons pour nos clients (industries pharmaceutiques, fabricants de dispositifs médicaux, institutionnels) des études pharmaco-épidémiologiques, médico-économiques et de parcours de soins à partir de solides méthodes statistiques et de solutions innovantes s'appuyant sur l'Intelligence Artificielle (Machine Learning, Deep Learning, ...). Afin de rendre nos analyses et nos données facilement compréhensibles pour les non-experts, nous avons également développé un savoir-faire spécifique de Datavisualisation autour du développement d'outils digitaux (Sites web, Applis iOS).

Dans le cadre de nos développements R&D, en lien avec notre partenaire académique Mines Saint-Etienne, et suite à la mise à disposition des données du Système National des Données de Santé (SNDS) aux acteurs privés, nous cherchons à renforcer notre pôle Data.

Mines Saint-Etienne (UMR CNRS 6158 LIMOS) apporte son expertise au travers du département ingénierie des systèmes de soins et des services de santé (I4S) du Centre Ingénierie et Santé (CIS). L'équipe I4S possède une expérience de recherche en génie industriel et informatique appliqués à la santé de plus de 10 ans. Les travaux de l'équipe se répartissent en trois axes : (i) l'analyse, de la modélisation et de la simulation des systèmes (évaluation de performances des nouvelles technologies appliquées en santé notamment), (ii) l'optimisation et l'aide à la décision (tournées d'ambulances, organisation de soins à domiciles), et (iii) l'analyse de données de santé (application de techniques de data mining, process mining et machine learning aux bases de données de santé médico administratives telle SNIIRAM, PMSI).

Pour notre équipe Data-management, nous recherchons un :

### **DATA ENGINEER / ANALYSTE**

#### **Le contexte :**

La mission d'HEVA est de participer à la compréhension et à l'amélioration du système de santé français. Notamment par la réalisation de véritables études épidémiologiques à partir des données de l'assurance maladie. HEVA constitue ainsi des cohortes de centaines de milliers de patients, et est capable de les suivre sur une dizaine d'année et de repérer les événements de leur prise en charge. Les enchaînements particuliers de la prise en charge médicale sont ainsi mis en évidence et analysés.



Pour cela, une équipe projet pluridisciplinaire (pharmaciens, épidémiologistes, data-managers, data scientists et biostatisticiens) travaille sur la plus grosse base de données médicale du monde, le Système National des Données de Santé (SNDS). Même si nous ne travaillons que sur des extractions spécifiques (à chaque fois sur une thématique précise), ces données sont volumineuses (plusieurs To) et complexes (complétude fluctuante, règles de gestions extrêmement variables et évolutives en fonctions de la réglementation) avec plusieurs centaines d'entités à manipuler.

Aujourd'hui, une première partie de ce travail est réalisé sous SAS (extraction des données brutes, jointures, identification des variables d'intérêt, ...), une seconde sous Python (analyses data science). Nous avons pour ambition de **faire évoluer notre organisation technique pour nous tourner vers des technologies pleinement orientées big-data**. Nous voulons améliorer notre plateforme de traitement et d'analyse de nos données. Cette plateforme aura pour avantage d'unifier notre process de production, de faire gagner en performance le data-management et facilitera les analyses de data-science. Cela facilitera la création d'outils de data visualisation pour permettre la compréhension des données et contribuer aux travaux sur l'explicabilité des modèles de data science.

**Le poste :** nous sommes donc à la recherche d'un **data engineer / analyste** pour développer ces analyses et ce système de base de données. Les grandes étapes de cette mission seront à mener autour de 2 axes :

#### Axe principal (analyse de données) :

- Prise de connaissance des aspects spécifiques des **données de santé** (type de données et environnement technique),
- Contribuer aux réflexions sur des outils de **data visualisation** permettant de tester des scénarios (analyses de sensibilités) d'extractions de cohortes sur la plateforme
- Etude des solutions techniques existantes pour une infrastructure d'**exploration de données** et de reporting (CubR, Redash, SuperSet, Metabase, ...)
- Implémentation d'outils de **data science** (clustering, classification) pour contribuer à l'étape de data review, en travail collaboratif avec l'équipe Data Science : développer un outil de **détection et de diagnostic d'erreurs** (anomalies de données)
- Contribution à l'élaboration de solutions pour **l'explicabilité** des modèles de machine learning pour les data review (et diagnostics) sur les études épidémio.

#### Axe secondaire (système de base de données) :

- Etude des solutions techniques existantes pour une infrastructure de **stockage et de requêtage de données** pour la réalisation d'étude épidémiologiques (Hadoop/Spark, format Parquet, ...),
- Construction d'une Preuve de Concept (POC) et test des performances en situation réelle (cas d'étude sur une pathologie donnée) au sein du pipeline de production,
- Formations des utilisateurs.



### **Votre profil :**

- Jeune diplômé d'une Grande Ecole d'Ingénieur française, filière informatique / mathématiques appliquées / statistiques / data science,
- Être curieux, indépendant dans le travail et bien organisé,
- Être pragmatique et capable de délivrer un résultat,
- Bases solides en analyses de données et analyses statistiques (une première expérience en Python & SQL est appréciable),
- Notions de machine learning,
- Bases en administration système (Linux, système de base de données, ...),
- Envie de travailler sur des projets impliquant des données volumineuses,
- Aimer travailler sur le développement d'outils qui servent directement à une équipe de production d'études épidémiologiques,
- Autonome et rigoureux, vous avez le sens de l'organisation et avez développé votre esprit analytique et le sens de la synthèse,
- Volonté de travailler en équipe et d'apporter vos compétences techniques à des projets ambitieux.

### **Vos conditions de travail :**

Au sein de notre siège basé à Lyon, vous êtes accueilli dans une ambiance conviviale et bienveillante. Chaque membre de l'équipe participe à une atmosphère de travail privilégiant l'entraide, favorisant la créativité et générant ainsi l'amélioration continue des compétences et de la performance au service de la satisfaction de nos clients. Vous aurez également accès au laboratoire situé dans le Centre Ingénierie et Santé de Mines Saint-Etienne (déplacements ponctuels à prévoir), situé sur le Campus Santé Innovation, à proximité du CHU de Saint-Etienne. Un double encadrement vous donnera accès d'une part à l'excellence académique et scientifiques de l'équipe de Mines Saint-Etienne, et à la grande expertise dans l'analyse des données de Santé de l'équipe d'HEVA.

Dans ce cadre du plan de relance R&D de l'État, vous serez employé par le laboratoire et votre rémunération est de : ~32 000€ brut annuel (montant fixé dans le cadre du Plan de Relance de l'emploi accordé par l'État, via le Centre de Recherche de Mines Saint-Etienne). Vous aurez accès et pourrez participer aux activités de communications scientifiques (colloque, congrès nationaux/internationaux, revue scientifique, ...).

Offre d'emploi : CDD (1 an), statut cadre.

Date limite de candidature : **25/07/2021**

Entretiens entre le **25/07/2021** et le **31/07/2021**

Début souhaité : **01/10/2021**

<https://hevaweb.com/fr>

[mines-stetienne.fr](https://mines-stetienne.fr)

Merci d'envoyer votre dossier de candidature (CV + lettre de motivation) à

[mprodel@hevaweb.com](mailto:mprodel@hevaweb.com)