

## Stage : Machine Learning Engineer (H/F) Développement d'un design pattern data science

### Qui sommes-nous ?

Nous pourrions commencer en nous présentant comme l'un des leaders sur le marché international du conseil en data, avec un chiffre d'affaires défiant ceux des G.A.F.A. Mais non. Nous préférons mettre en avant notre cadre de travail, nos réussites et surtout nos consultants. Allez venez, c'est par là...

Quantmetry c'est une centaine de personnes qui travaille de concert pour accompagner nos clients dans leurs réflexions, P.O.C., industrialisation autour de la Data, l'Intelligence Artificielle, le Machine Learning et la Big Data. Nous sommes ce qu'on peut appeler communément un cabinet de conseil pure player en Data.

N'étant pas sectorisés, nous intervenons sur différents sujets (voici une petite liste non-exhaustive) :

- La maintenance prédictive,
- La segmentation clients, le scoring,
- La détection de fraude, de corruption, de blanchiment d'argent,
- Le développement des architectures de plateformes Data,
- L'analyse de textes et d'images dans le cadre de la lutte contre le cancer du sein.

Dans le cadre de notre incessante croissance, nous recherchons des personnes motivées pour nous accompagner et ainsi travailler sur des projets de recherche.

## Préalable

Quantmetry propose ci-dessous le volet R&D d'une offre de stage. Tout stagiaire entrant chez Quantmetry, outre le travail de R&D qui lui est proposé et pour lequel il sera encadré, aura aussi pour objectif de participer à certaines missions de conseil chez des clients variés, lui permettant d'aborder le monde du consulting.

## Contexte

Un **design pattern** est une solution pratique à un problème de conception logiciel. Il existe de nombreux design pattern classiques répondant à différentes problématiques logicielles (web, microservices ...).

Actuellement nous n'avons pas de design pattern **spécifiquement adaptés** pour les projets data science. Pourtant certaines problématiques subsistent :

- Comment garantir une évaluation des performances fiable ? (notamment garantir l'absence de leakage dans le processus d'évaluation des performances, c'est à dire l'utilisation d'informations qui ne seraient pas disponibles en production) ?
- Comment être robuste au changement de l'infrastructure ou des sources de données en cours de développement ?
- Comment l'architecture choisie facilite le versionnement des modèles et des données ?

L'objectif de ce stage est donc de construire **cette méthodologie de développement** en interaction avec nos data scientists, nos data engineer et nos architectes. Cette solution sera par la suite intégrée à nos bonnes pratiques de développement.

Plusieurs étapes seront nécessaires au cours du stage :

- Une veille scientifique sur les design pattern data science existants
- Une veille interne pour recenser les problèmes techniques rencontrés par les data scientist et les data engineer en mission
- La rédaction d'un guide méthodologique formalisant ces différents design pattern "data science", leurs conditions d'applicabilité ainsi que des exemples d'implémentations.
- Le test de ces design pattern, soit dans le cadre d'une mission soit dans la modification du template de code actuel.