



Stage 6 mois – Data Scientist pour l’innovation

Sujet : Prédiction de l’antibiorésistance en métagénomique

Présentation de la société bioMérieux

L’innovation scientifique et technologique est au cœur de la stratégie de bioMérieux. Acteur mondial dans le domaine du diagnostic in vitro depuis 55 ans, bioMérieux conçoit, développe, produit et commercialise des solutions de diagnostic (réactifs, instruments et logiciels) destinées à des applications médicales et industrielles, qui déterminent l’origine d’une maladie ou d’une contamination pour améliorer la santé des patients et assurer la sécurité des consommateurs. Fidèle à sa mission d’améliorer la santé publique dans le monde, bioMérieux est présente dans plus de 150 pays, au travers de 43 filiales, et réunit 10 000 collaborateurs. Son siège social se situe à Marcy l’Etoile, dans la région de Lyon, en France.

Contexte et enjeux

La prédiction de la résistance aux antibiotiques, obtenue directement à partir d’un prélèvement de la flore, sans passer par une étape de culture, est un enjeu clinique très important : cela permet de gagner un temps précieux pour adapter l’antibiothérapie d’un patient. Cela est d’autant plus important dans un contexte où l’antibiorésistance se répand à l’échelle mondiale et constitue aujourd’hui un enjeu majeur de santé publique.

Le stage porte sur la mise au point d’outils pour la prédiction, par des approches d’apprentissage supervisé (Machine Learning), de la résistance aux antibiotiques d’un pathogène présent dans une flore séquencée par approche métagénomique. Des modèles de prédiction appris sur la base de pathogènes isolés par une étape de mise en culture sont déjà disponibles et serviront de base au travail. Il s’agira dans un premier temps d’évaluer si ces modèles peuvent directement s’appliquer à un métagénome. Dans un second temps, le stage visera à concevoir une stratégie d’apprentissage à partir de métagénomomes.

Vous travaillerez plus précisément avec des algorithmes de classification, ainsi qu’avec des outils de bioinformatique dédiés à l’analyse de métagénomomes (ex : Kraken, MetaCherchant) et à la simulation de données de séquençage (« reads »). Vous vous appuieriez également sur les travaux déjà réalisés en interne sur ce sujet. Vous aurez accès au cluster de calcul de bioMérieux. Vous rejoindrez l’équipe Big Data Analytics, composée de Data Scientists, statisticiens et bioinformaticiens, et intégrerez plus spécifiquement une équipe projet à laquelle la problématique de ce stage est rattachée.

Missions

- concevoir et implémenter des stratégies de Machine Learning à partir de données de métagénomique ;
- créer des jeux d’apprentissage et de validation, principalement à l’aide de données simulées;
- évaluer les performances des modèles construits et de proposer des améliorations.

Profil

Etudiant en master Bac+5 ou niveau équivalent (diplôme d’ingénieur), en statistiques, Machine Learning, ou Data Science avec un fort intérêt pour des applications à la biologie, ou bioinformatique avec de bonnes bases en statistiques, ou Machine Learning. Vous aurez à lire et à produire des documents en anglais, et utiliserez les langages R et python ainsi que des outils de bioinformatique basés sur les langages shell, C++, python.

Pour postuler

Plateforme recrutement : sur la page <http://biomerieux-recruitment.com>, avec la référence 53849.

Responsable du stage : magali.dancette@biomerieux.com .

Démarrage possible dès février 2019. Durée 6 mois.

Localisation : Marcy l’Etoile, commune de l’agglomération de Lyon