



Stage Machine Learning - 6 mois

Développement d'un outil de validation automatique de données de qualité de l'air

Introduction

La pollution de l'air représente un risque majeur pour la santé. A l'échelle mondiale, elle serait à l'origine d'environ 7 millions de décès prématurés par an (<https://www.who.int/news-room/air-pollution>) dont 48 000 en France. Elle coûte à la France plus de 100 milliards d'euros par an (rapport du Sénat – 2015). Bien que les émissions soient en baisse régulière, chaque année en France, ce sont encore plus de 9 millions de tonnes de polluants qui sont rejetés dans l'air par les activités humaines. En France, plus de 700 stations de mesures de la qualité de l'air quadrillent le territoire.

Ces stations sont équipées d'analyseurs de polluants (gaz et particules fines) et de systèmes d'acquisition de données (DAS). Ces DAS servent à récupérer les données techniques et métrologiques des analyseurs. Les données métrologiques sont validées manuellement par les équipes spécialisées des réseaux de surveillance de la qualité de l'air (AASQA), puis sont rendues publiques. Cette étape de validation nécessite une grande expertise sur la mesure des polluants dans l'air ambiant. Elle prend en compte les données techniques issues des analyseurs, l'évolution simultanée des différents polluants sur plusieurs sites mais également les phénomènes météorologiques et événements particuliers qui ont pu se produire localement ou sur une plus grande échelle (incendie, travaux, épandages ...).

ENVEA est l'une des toutes premières entreprises internationales d'instrumentation, d'analyse et de mesure de l'environnement et des milieux industriels. Depuis sa création en 1978, ENVEA conçoit, fabrique et commercialise des appareils de mesure en continu de l'air ambiant, des émissions et procédés industriels et logiciels de traitement de données.

Créée en 1979, AIRPARIF est l'association agréée par le ministère de l'Environnement chargée de la surveillance de la qualité de l'air sur l'ensemble de l'Île-de-France

Missions

Le premier objectif de ce stage consiste à comprendre et modéliser les mécanismes mis en œuvre par les AASQA pour valider les données issues de différents analyseurs de polluants. Le deuxième objectif consiste à identifier les outils informatiques et statistiques qui devront être utilisés pour le développement d'un outil de validation automatique basé sur un algorithme de machine learning.



Après une phase d'apprentissage, cet outil autonome devra pouvoir juger la pertinence des données issues des analyseurs présents dans une station de surveillance en prenant en compte notamment :

- Des données techniques internes issues des analyseurs (détection de défaillances par exemple).
- La comparaison de l'évolution de plusieurs polluants en inter-station.
- Des liens statistiques entre les différentes stations d'un même réseau.
- Des données externes à la station telles que la température ambiante, l'humidité relative, la vitesse du vent, l'état de la circulation...
- Des données historiques obtenues sur la(les) station(s) disponibles ...

Le stage se décomposera en trois parties :

- Synthèse et modélisation des mécanismes mis en œuvre par les AASQA aux différentes étapes de validation des données, technique et environnementale.
- Etat de l'art des travaux réalisés en Europe (ex. JRC) et aux USA sur la validation automatique des données de la qualité de l'air.
- Identification des outils mathématiques requis pour automatiser la validation de données grâce à des algorithmes de type machine learning et mise en place d'un premier modèle de validation automatique des données. Celui-ci permettra d'évaluer les axes de travail à consolider dans l'optique à moyen-terme d'un outil opérationnel. Le cas échéant, le stage pourrait être poursuivi par une thèse.

Profil

Etudiant en dernière année d'école d'ingénieur ou Master 2 en Informatique ou Mathématiques appliquées, vous disposez de compétences en machine learning et en Programmation (Python).

Des connaissances générales en Physique et Métrologie seraient un plus.

Conditions du stage

Stage de 6 mois basé à Paris (75) dans les locaux d'Airparif, à partir de février/mars 2022.
Déplacements à prévoir dans les bureaux R&D d'ENVEA à Poissy (78) et Bidard (64).

Rémunération

Gratification selon profil et durée du stage

Contact

Lionel GUISEPPIN : lionel.guisseppin@airparif.fr