

PROPOSITION DE STAGE 2020

Sujet : Fusion de données de consommation d'électricité à différentes échelles

Contexte :

L'équilibre offre/demande est un enjeu majeur pour le Groupe EDF. Pour cela, il est nécessaire de construire des modèles de prévision de consommation d'électricité performants. De même, il est primordial de disposer de données fiables, avec la plus grande finesse possible pour garantir les entrées des modèles. Les sources de données sont nombreuses et souvent très hétérogènes, en termes de granularité, autant temporelle que spatiale.

En effet, les données de consommation peuvent être issues de différentes sources : panel, smart meters, réseau, régional, national. Par conséquent, il est difficile d'obtenir une mesure unique à partir de toutes celles-ci. Des travaux existent sur les prévisions hiérarchiques montrant l'intérêt de considérer ces données à différentes échelles pour la prévision en utilisant des techniques d'agrégation.

Objectif du stage :

L'objectif de ce stage est de comparer les différentes techniques de fusion statistique méthodologiquement et de les appliquer sur les données issues de plusieurs sources afin d'évaluer leur robustesse, leur pertinence et leur universalité.

Les techniques attendues sont : algorithme NIPALS, factorisation de matrices, agrégation multi-niveaux.

Planning envisagé :

- Appropriation du sujet, revue de la littérature sur le sujet métier et les méthodes associées
- Appropriation des différentes données fournies par EDF
- Implémentation des méthodes de fusion statistique, application de celles-ci et évaluation
- Ecriture du rapport de stage

Encadrants et contacts

- Christian Derquenne, chercheur senior (christian.derquenne@edf / 01 78 19 39 13)
- Yannig Goude, ingénieur chercheur, pilote Data Innovation Lab d'EDF (yannig.goude@edf.fr / 01 78 19 39 56)

Profil du stagiaire :

Niveau d'étude : M2 recherche en mathématiques

Domaines de compétences : data science, machine learning, statistique

Informatique : R

Connaissances supplémentaires : autres domaines des mathématiques appliquées, rédaction scientifique

Savoir-être : dynamique, inventif, sérieux, motivé, humour apprécié

Déroulement :

Durée proposée : 6 mois

Date de début souhaitée : février-mars 2020

Localisation : Sur le nouveau site d'EDF R&D à Saclay, un environnement ultra-moderne dédié à la recherche de haut niveau au sein du campus de Paris-Saclay.