

OFFRE POST-DOCTORAT Janvier 2022 (Projet SIMULBUS) LS2N

L'équipe DUKe (Data User Knowledge) du LS2N, UMR CNRS 6004, recherche un.e postdoc dans le cadre du projet SIMULBUS sur le sujet suivant:

Fouille de données variées et hétérogènes pour l'apprentissage de modèles pour la prédiction du comportement de courses de bus

Contexte :

Ce contrat s'intègre dans le cadre du projet R&D Collaborative Pays de la Loire SIMULBUS. Ce projet implique les équipes DUKe et PSI du LS2N et les entreprises A1 Statistiques et TELLAE.

Le développement des technologies Big Data et la prolifération des données des réseaux de transports publics ont favorisé le développement de solutions pour l'analyse des temps de parcours et de la performance des réseaux. Les données de base sont celles fournies par les Systèmes d'Aide à l'Exploitation (SAE) des réseaux de transports publics. Les données plus précises spatialement ne sont pas exploitées. Le croisement des sources de données est de plus un élément absent de ces technologies qui s'appliquent à restituer une source de données sans faire le lien avec le contexte de ligne (voirie, contexte urbain, météo, etc). Enfin, ces technologies sont également déconnectées des approches de simulation et fournissent simplement des analyses sans pouvoir évaluer de nouveaux scénarios de gestion.

L'objectif du projet est de croiser des données variées en termes de format, structure ou représentation en termes de nature (données temporelles, déterminants géographiques, voirie, météo, etc.), mais également des données avec une hétérogénéité sémantique afin de créer de modèles riches expliquant avec fiabilité la ponctualité et la régularité d'une course.

Mission :

Le ou la candidat(e) retenu(e) devra participer activement à la proposition et à l'implémentation d'une approche combinant la fouille de motifs séquentiels et les modèles de mélanges de régressions, intégrant des données variées, pour modéliser une course d'un bus et prédire son comportement. La validation des modèles obtenus sera réalisée en partie à l'aide de la simulation dans un processus itératif et incrémental. Les modèles développés dans ce travail seront testés sur des données de deux réseaux de transports publics : SEMITAN exploitant du réseau de l'agglomération nantaise et RTM exploitant du réseau de la métropole de Marseille.

Le ou la candidat(e) retenu(e) participera aussi à la vie du projet :

- organisation des réunions,
- interactions avec les différents partenaires,
- participation à la rédaction des livrables du projet (rapport, publications scientifiques, etc.).

Encadrement

Au LS2N, DUKe, le.a post-doctorant.a sera encadré.e par Mounira Harzallah et François Queyroi, en collaboration avec Bruno Rivière de A1 Statistiques et Vincent Leblond de TELLAE.

LS2N <https://www.ls2n.fr/>
A1 Statistiques <https://a1statistiques.fr/>
Tellae <https://www.tellae.fr/>

Calendrier: Début 01/01/2022 (durée 12 mois).

Compétences

Doctorat en Informatique, avec des compétences en fouille de données/machine learning. Une expérience en clustering à base de modèles et aspects temporels serait un vrai plus.

Candidature, par mail à mounira.harzallah@univ-nantes.fr et francois.queyroi@univ-nantes.fr. Merci de joindre :

- CV détaillé
- Lettre de motivation
- Noms d'une ou deux personnes référentes