

Janvier 2022

OFFRE DE STAGE

<u>INTITULÉ DU POSTE :</u> Développement des modèles d'estimation de la fréquentation TER	<u>ENTITÉ :</u> DIRECTION MARKETING, INNOVATION & DEVELOPPEMENT Département DEVELOPPEMENT DE LA MOBILITE
--	---

<u>DATE DE DÉBUT SOUHAITÉE :</u> AVRIL2022
<u>DURÉE DE STAGE :</u> 6 MOIS
<u>LOCALISATION :</u> PARIS - AVEC QUELQUES DÉPLACEMENTS PONCTUELS À PRÉVOIR
<u>NOM ET PRÉNOM DU TUTEUR :</u> RUIXUE SANG

CONTEXTE ET DESCRIPTION DES MISSIONS

La fréquentation des trains TER n'est pas connue précisément. Actuellement elle est estimée sur un périmètre réduit des circulations TER en utilisant des modèles empiriques basés sur les comptages, les données de ventes et des règles « métiers ».

Les enjeux de la connaissance de la fréquentation des trains TER sont nombreux et importants (reporter aux autorités organisatrices, fiabiliser la répartition des recettes, améliorer le plan de transport, mieux connaître le réseau, dimensionner les trains, optimiser le cadencement, améliorer l'information Voyageur, la lutte anti-fraude...)

Dans le but de connaître le plus précisément possible la fréquentation, la Direction TER, en collaboration avec la Factory Data & IA d'ITNOVEM, a mis en place un projet de construction de modèle(s) de Machine Learning permettant d'estimer le nombre de montées et de descentes à chaque arrêt et pour chaque train (LightGBM). Les premiers résultats sont encourageants mais la trop faible volumétrie des données de comptages limite l'apprentissage et donc la qualité de la prévision à une maille fine.

Dans l'attente d'accroître la volumétrie des données d'entrée (travaux en cours au sein de TER), deux missions devront être réalisées au cours du stage que nous proposons :

- ! Tester, développer et évaluer les modèles de ML construits et proposer des évolutions / optimisations pour améliorer la performance des modèles prédictifs
- ! Optimiser les estimations existantes basées sur des règles métiers en intégrant notamment les spécificités régionales (transfrontaliers, distributeurs multiples, connections aux réseaux urbains...)

COMPÉTENCES REQUISES

COMPÉTENCES REQUISES

TRÈS BONNE CONNAISSANCE EN DATA SCIENCE / STATISTIQUES, MATHÉMATIQUES

BONNE MAÎTRISE DU PYTHON

CONNAISSANCE DE L'ENVIRONNEMENT BIG DATA ET EXPÉRIENCE DANS L'UTILISATION DE PYSARK SERA UN PLUS

INNOVATION / R & D ET FORCE DE PROPOSITION

RIGUEUR, AUTONOMIE ET COMMUNICATION

PROFIL RECHERCHÉ :

STAGE DE FIN D'ÉTUDES D'ÉCOLES D'INGÉNIEUR OU MASTER 2 EN DATA SCIENCE, STATISTIQUE, MATHÉMATIQUES APPLIQUÉES, MACHINE LEARNING

CONTACT

RUJUE.SANG@SNCF.FR