

OFFRE DE STAGE

Consolidation d'une base d'apprentissage et de modèles de machine learning pour la prédiction d'effondrements de cavités souterraines

Nos réf. : Ineris - - ID 2816879

Date de publication : 04/11/2024

Lieu : Nancy – Campus Artem – accessible en transport en commun, à 15 mn du centre

Type de contrat : stage

Contact : nathalie.conil@ineris.fr - Tél. : 03 54 40 66 17 pour plus d'informations

Ineris, Institut National de l'Environnement Industriel et des Risques. Etablissement public à caractère industriel et commercial sous la tutelle du Ministère de la Transition l'Ecologique, d'environ 510 employés et implanté à 50 km au Nord de Paris. Site internet : www.ineris.fr.

Le stage proposé s'effectuera au sein de l'unité RMC2, Risques après-mines, cavités et carrières. L'unité compte plus de 15 permanents la plupart basés à Verneuil en Halatte où se trouve le siège de l'Ineris. L'équipe travaille sur des problématiques en lien avec l'après-mines et les cavités souterraines anthropiques abandonnées et notamment sur l'impact potentiel du changement climatique sur l'augmentation des instabilités des ouvrages souterrains.

Depuis 2022 un inventaire comprenant des effondrements de grande ampleur au droit de carrières souterraines abandonnées a été initié, il comprend 550 cas et 50 paramètres en lien avec la localisation de la carrière, l'environnement géologique de la carrière, la géométrie de la carrière, les dates d'exploitation. Lors d'un précédent stage de master 2, un premier modèle de machine learning pour lier la taille de l'effondrement aux paramètres embasés a été testé sur un faible nombre de ces cas (143) et sur une liste restreinte de paramètres en raison du faible taux de remplissage de la base d'apprentissage. Néanmoins, les premiers résultats sont encourageants et nous poussent à compléter les informations manquantes pour améliorer les résultats du modèle.

Ce stage s'articulera en priorité autour des sujets suivants.

Collecte de données dans des documents

Une bonne partie de ces informations manquantes se trouvent dans des documents de formats différents qu'il serait fastidieux d'analyser manuellement pour y rechercher ces informations. Aussi, le stagiaire développera des techniques de « fouilles » de texte soient pour extraire ces informations avec leur contexte, en réponse à une requête formulée par l'utilisateur en langage naturel. Différentes solutions (LLM sur étagère ou modèles réentraînés sur des corpus de texte spécifiques) pourront être expérimentées et comparées les unes aux autres.

Prise en compte de nouvelles caractéristiques environnementales

Par ailleurs, nous souhaitons compléter ou incorporer dans le modèle des grandeurs explicatives issues de bases de données environnementales (positions des nappes

phréatiques, données de terrain, conditions hydroclimatiques, cartographiques...) gérées par d'autres producteurs de données (MétéoFrance, Géorisques, Géoportail etc...). Le travail de stage consistera à identifier et automatiser la collecte des données utiles pour compléter la base d'apprentissage existante.

En option

Le modèle précédent permettait de classifier les effondrements de carrières abandonnées en plusieurs catégories, afin de prédire l'ampleur de ces effondrements. Nous souhaitons désormais développer un second modèle de machine learning capable de calculer la probabilité de survenue d'un effondrement. En fonction de l'avancée des deux premières tâches un travail d'identification et de première mise en œuvre de ce modèle sera réalisé. L'objectif final étant de créer un arbre décisionnel permettant de caractériser la prédisposition aux effondrements, en s'appuyant sur les paramètres et facteurs clés identifiés en amont.

Un fort intérêt pour le développement d'approches scientifiques appliquées est demandé. Des compétences en statistiques, en machine learning et/ou en géostatistique ainsi qu'un goût pour la programmation sont requis. Le stage s'insère dans un contexte d'aide à la décision politique pour les questions environnementales. Un intérêt pour les enjeux environnementaux est souhaité tout comme une appétence pour les géosciences.

PROFIL

Bac+5 (Ingénieur / Master 2). Ce stage s'adresse à un.e étudiant.e ayant suivi des enseignements en statistiques, machine learning et/ou géostatistiques appliquées aux questions environnementales.

Expérience / Compétence

Des compétences en programmation R ou python et SIG sont nécessaires. Une première application de ces compétences en géosciences constituerait un réel atout pour le stage.

DIVERS

Durée : 4 à 6 mois.

Indemnité mensuelle : environ 660 € (stage à temps plein).

Ce poste est ouvert aux personnes en situation de handicap.