 

# Proposition de stage au LIFO – Université d’Orléans avec le BRGM:

# Interprétation automatique de coupe géologique de forage par apprentissage automatique semi-supervisé

Encadrant : P. Marcel (LIFO – Univ. Orléans) ; C. Loiselet (BRGM)

Contexte : Le BRGM, service géologique national, a la charge de mettre à disposition une connaissance géologique du sous-sol de notre territoire. Pour ce faire, nous collectons les données d’ouvrages souterrains (forages, sondages) qui sont structurées dans une base de données nommée la Banque du Sous-Sol (BSS), organisée et gérée par le BRGM. Cette bancarisation permet de favoriser les applications en ressources naturelles (ressources fossiles et énergétiques), en géotechniques (travaux d'infrastructure et d'aménagement, etc.). Cette base de données contient 700 000 ouvrages et travaux souterrains qui, pour près de la moitié, contiennent des données et informations sur la géologie du sous-sol et plus particulièrement la description géologique le long des logs de sondages/forages. Ces données sont actuellement mises à disposition sur notre plateforme de diffusion InfoTerre ([http://inforterre.brgm.fr](http://inforterre.brgm.fr/)).

Ces logs géologiques sont par la suite interprétés afin de fournir une information géologique cohérente de notre sous-sol selon une coupe géologique. Aujourd’hui, environ 20% des logs possèdent une coupe géologique interprétée. Nous cherchons ici à accélérer le travail d’interprétation de ces données afin de fournir le plus d’informations possible par l’application d’une méthode d’Intelligence Artificielle.

Description du stage : L'objectif de ce stage est d'étudier une méthode d'apprentissage automatique semi-supervisée pour prédire des données de forage. Cette méthode est inspirée de méthode de recommandation par apprentissage de similarité [1,2]. Dans le cas des données de forage, il s'agira d'utiliser une approche basée sur les K plus proches voisins (KNN), où le voisinage est constitué par des forages existants. Un point clé est donc la définition d'une similarité entre forages, qui sera apprise à partir des caractéristiques des forages.

La méthode envisagée repose sur les étapes suivantes :

1. Extraction de caractéristiques pertinentes des forages existants
2. Définition de similarités simples propres à chaque caractéristique
3. Constitution d'un ensemble d'apprentissage par labellisation faible [3]
4. Entrainement d'un classifier linéaire permettant de pondérer les similarités entre caractéristiques [1]
5. Définition des paramètres du KNN
6. Définition de la méthode de prédiction à partir des forages voisins

L'approche sera comparée aux méthodes de l'état de l'art pour prédire les données de forage, notamment le Krigeage ([https://fr.wikipedia.org/wiki/Krigeage](https://urldefense.com/v3/__https%3A/fr.wikipedia.org/wiki/Krigeage__;!!KbSiYrE!idjtB57mv8d534SF7kJeYfKRMxHvEbrcRbiq0fJYYTRIJVNVJ9HGcTd3WjC1AhZ00if2xmpC8Qo6FsPT7pAA0-M3-D_4SFyLBg$)).

Le déroulement du stage comprendra :

* une étude bibliographique,
* la proposition d'une approche de prédiction de données de forage,
* l'implémentation et le test de l'approche proposée.

Le profil recherché est un(e) étudiant(e) ayant des connaissances solides en géologie, programmation, et mathématiques, et une appétence pour la recherche. Le stage pourra déboucher sur une thèse de doctorat financée.

### Spécificités du poste :

* Stage en fin d'études (Ingénieur ou Master 2)
* Niveau d'études requis : bac+4
* Durée du stage : 6 mois
* Période de stage : Avril – Septembre

Le stage sera encadré par Christelle Loiselet (BRGM c.loiselet@brgm.fr ) et Patrick Marcel (LIFO -Univ. Orléans patrick.marcel@univ-orleans.fr ). Il se déroulera **au LIFO, campus Universitaire de la Source à Orléans.**

Votre lettre de motivation et un CV sont à adresser par email jusqu’au **31/12/2024** aux tuteurs du stage.

Références :

[1] Ramanathan V. Guha, Vineet Gupta, Vivek Raghunathan, Ramakrishnan Srikant: User Modeling for a Personal Assistant. WSDM 2015: 275-284

[2] Krista Drushku, Julien Aligon, Nicolas Labroche, Patrick Marcel, Verónika Peralta: Interest-based recommendations for business intelligence users. Inf. Syst. 86: 79-93 (2019)

[3] Alexander Ratner, Stephen H. Bach, Henry R. Ehrenberg, Jason A. Fries, Sen Wu, Christopher Ré: Snorkel: rapid training data creation with weak supervision. VLDB J. 29(2-3): 709-730 (2020)