

# **Proposition de stage de master 2ème année**

## **Apprentissage semi- et auto-supervisé exploitant du clustering avec connaissances. Application à la cartographie du potentiel minéral.**

**Equipes partenaires** : LIFO, Université d'Orléans et BRGM, Orléans, France.

**Encadrement** : Vincent Nguyen, Christel Vrain, Thi-Bich-Hanh Dao (LIFO), Hugo Breillard, Christel Loiselet (BRGM)

**Gratification** : environ 600€/mois selon la réglementation

**Durée** : 6 mois

**Date** : début 2023

### **CONTEXTE**

La cartographie du potentiel minéral consiste principalement à déterminer le lien statistique entre un ensemble d'occurrences minérales (points) et une carte géologique (polygones) dans le but d'indiquer des zones favorables en termes de potentiel minier. La majorité de ces cartes de favorabilité sont générées par des techniques numériques, et notamment depuis quelques années par des méthodes d'intelligence artificielle. Cependant les méthodes actuelles comportent des défauts liés à la qualité des données, à la présence d'approximations infondées et à l'utilisation de méthodes supervisées sur des données majoritairement non annotées [1,2]. Nous nous intéressons à explorer de nouvelles pistes pour améliorer la précision de la cartographie prédictive du potentiel minéral par le couplage d'une part de l'intégration de connaissances géologiques dans des algorithmes de clustering afin d'obtenir des clusters mieux fondées géologiquement, et d'autre part de l'utilisation de méthodes d'apprentissage semi-supervisé ou auto-supervisé pour la construction de carte.

Ce sujet s'intègre dans le cadre du projet ANR Contrats doctoraux IA.iO Artificial Intelligence in Orléans: Learning from heterogeneous data and expert knowledge. Applications in geological and environmental sciences. Le stage sera encadré conjointement par des membres du LIFO et des membres du BRGM. Un financement de thèse est prévu suite à ce sujet.

### **DESCRIPTION DU TRAVAIL**

L'objet du stage est d'étudier les deux volets :

- utilisation de méthodes de clustering existantes et intégration de connaissances [3,4,5] en vue d'améliorer les données d'entrées des modèles de prédictivité minérales. Les connaissances géologiques sont des ontologies développées au BRGM.
- étude de méthodes semi-supervisées ou auto-supervisées pour détecter des anomalies sur la carte géologique en tirant partie de l'apprentissage non-supervisé (clustering). Le résultat du clustering sera utilisé avec peu de données labellisées pour améliorer le modèle de prédiction d'apprentissage semi-supervisé ou auto-supervisé [6, 7].

Les solutions à étudier incluent l'apprentissage non-supervisé, et l'apprentissage profond avec faible supervision.

**Travail attendu :**

- Rapport d'étude des méthodes d'apprentissage non-supervisé sur les données géologiques.
- Rapport d'étude des méthodes auto- et semi- supervisées existantes tirant partie de l'apprentissage non-supervisé.
- Etude et implémentation d'une nouvelle méthode (ou amélioration d'une méthode existante)

**Compétences demandées :**

Deep learning, programmation.

- Profil Informatique, machine learning ou mathématique appliquée.
- Capacité de synthèse et de rédaction permettant une restitution régulière, claire et efficace du travail effectué.

**CANDIDATURE**

Les candidats sont encouragés à nous contacter au plus tôt possible. Envoyez nous votre candidature à Thi-Bich-Hanh Dao ([thi-bich-hanh.dao@univ-orleans.fr](mailto:thi-bich-hanh.dao@univ-orleans.fr)) en joignant votre dossier en un seul fichier pdf, comprenant un CV, une lettre de motivation et vos relevés de notes depuis le baccalauréat. Les documents peuvent être en anglais ou en français.

**RÉFÉRENCES**

- [1] J. Torppa, M. Middleton, E. Hyvönen, J. Lerssi, S. Fraser, A novel spatial analysis approach for assessing regional-scale mineral prospectivity in northern Finland. *Novel technologies for greenfield exploration*, Geological Survey of Finland, Special Paper, 57 (2015), pp. 87-120
- [2] Wang, J., Zuo, R. & Xiong, Y. Mapping Mineral Prospectivity via Semi-supervised Random Forest. *Nat Resour Res* 29, 189–202 (2020)
- [3] K. Wagstaff, C. Cardie, Clustering with instance-level constraints, in: *Proceedings of the 17th International Conference on Machine Learning*, 2000, pp. 1103–1110.
- [4] Dao T.-B.-H., Duong K.-C., Vrain C.: Constrained clustering by constraint programming. *Artificial Intelligence Journal*, 244: 70-94, 2017
- [5] Nghiem N.-V.-D, Vrain C., Dao T.-B.-H.: Knowledge integration in deep clustering, *ECMLPKDD*, 2022
- [6] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. 2020. Unsupervised learning of visual features by contrasting cluster assignments. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*. Curran Associates Inc., Red Hook, NY, USA, Article 831, 9912–9924.
- [7] Ciortan, M., Defrance, M. Contrastive self-supervised clustering of scRNA-seq data. *BMC Bioinformatics* 22, 280 (2021). <https://doi.org/10.1186/s12859-021-04210-8>