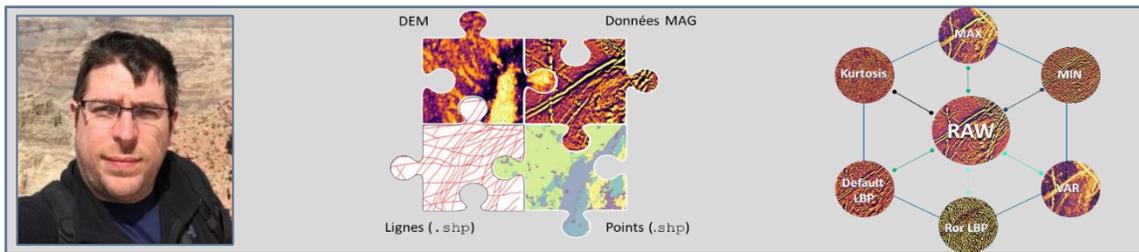


## Projet 2020-01: Comparaison de la performance des algorithmes d'apprentissage machine (IA) appliqués à l'exploration minière



**Dominique Genna, géo, Ph.D. – CONSOREM**

**Collaborateurs : Martin Blouin, Ph.D., GéoLearn, Jean-Philippe Paiement géo, M. Sc., Mira Geosciences et Sylvain Trépanier, géo, M.Sc., chercheur associé CONSOREM**

L'Intelligence Artificielle (IA) est de plus en plus utilisée afin d'optimiser plusieurs aspects de l'exploration minière. Son utilisation est encouragée par la croissance continue des banques de données et par le nombre grandissant d'algorithmes d'Apprentissage Machine (AM) disponibles. Ce projet visait à poser les lignes directrices de l'utilisation de l'AM pour une application en cartographie prédictive. Les objectifs inclus : 1) l'identification du pouvoir de discrimination des couches de données brutes du SIGEOM et des produits dérivés et 2) la comparaison de plusieurs algorithmes d'AM.

Le territoire choisi pour cette étude (Figure 1a) se situe à cheval sur les sous-provinces de LaGrande et d'Opinaca (SNRC33G06 et 33G11) et couvre aussi bien des lithologies ignées, sédimentaires qu'une ceinture volcano-sédimentaire. Les données d'entrée brutes sélectionnées pour cette étude comprennent : le MNE (Modèle Numérique d'Élévation), les images LANDSAT, le levé magnétométrique et la géochimie des sédiments de fond de lac. Les produits dérivés des couches brutes incluent une sélection de variables statistiques et texturales sur des fenêtres mobiles de 3x3, 5x5 et 10x10. Trois jeux de données d'apprentissage ont été bâtis : 1) MNE, et levé magnétométrique; 2) MNT, levé magnétométrique et les produits dérivés; 3) MNT, levé magnétométrique et sédiments de fond de lac. Deux types de points d'entraînement ont aussi été testés : les affleurements de géofiches et les stations lithogéochimiques de distribution très hétérogène. Un protocole de simplification et d'unification des codes lithologiques des géofiches du SIGEOM a été développé et permet une reclassification objective de la couche d'entrée lithologique (ici en 8 classes). Au total, 6 algorithmes d'AM non-supervisé (*K-Means* et *segGAN*), semi-supervisé (*segGAN* conditionnel) et supervisé (*Random Forest*, réseau de neurones et *Extreme Gradient Boosting*) ont été utilisés et ont mené à plus de 45 simulations.

Les résultats indiquent que, sur notre territoire, les données brutes (levé magnétométrique et le MNE) sont suffisantes pour prédire les grands ensembles géologiques avec un algorithme relativement simple et non supervisé, comme *K-Means* (Figure 1b). L'utilisation des produits dérivés (variables statistiques et texturales) apporte un pouvoir discriminant supplémentaire, mais fonctionne seulement avec des algorithmes plus complexes et paramétrables comme le *Random Forest* et le *Extreme Gradient Boosting* (Figure 1c et d). Ces algorithmes permettent en plus de générer des cartes de probabilités par lithologie qui peuvent ensuite être valorisées. Il est délicat d'évaluer la performance des algorithmes pour la cartographie prédictive.

L'évaluation repose ici sur le 1) respect des données d'apprentissage (justesse et précision) et 2) sur la correspondance avec la carte géologique du MERN (bruit, indice de similarité structurelle et évaluation d'image neurale). Il en ressort que le *Random Forest* et le *Extreme Gradient Boosting* fournissent les prédictions les plus justes et les plus précises.

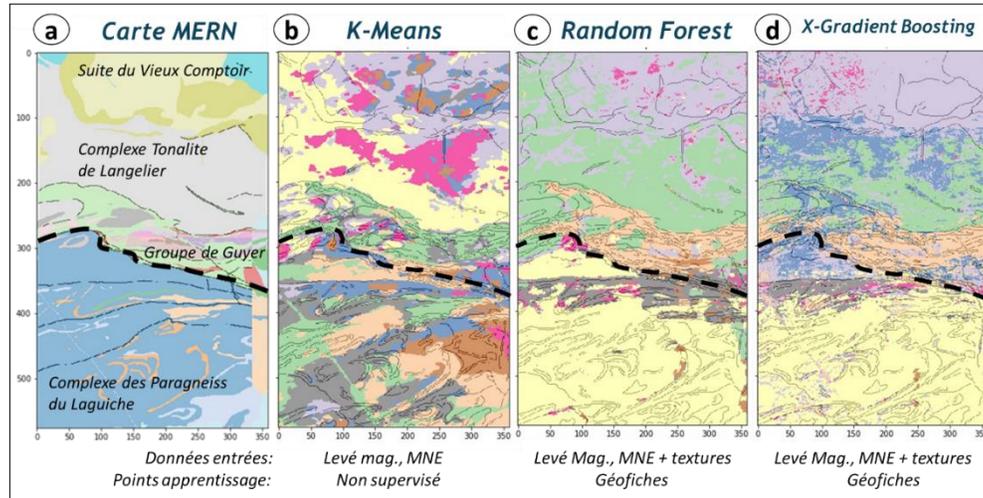


Figure 1 : a-Carte géologique du secteur d'étude (MERN); b-Cartographie prédictive non supervisée utilisant la K-Means; c-Cartographie prédictive supervisée utilisant le Random Forest; d-Cartographie prédictive supervisée utilisant le Extreme Gradient Boosting. La ligne pointillée représente la limite entre la sous-province du LaGrande au nord et d'Opinaca au sud.

## FICHE SOMMAIRE PROJET 2020-01

<b>Objectifs</b>	<ul style="list-style-type: none"> <li>◆ Réaliser une analyse prédictive par apprentissage automatique à partir des données disponibles sur le SIGEOM ;</li> <li>◆ Identifier les signatures statistiques des unités lithologiques à partir des données brutes (sans interprétation) ;</li> <li>◆ Évaluer le pouvoir discriminant des produits dérivés pour la cartographie prédictive ;</li> <li>◆ Comparer la performance des algorithmes d'apprentissage automatique en cartographie prédictive.</li> </ul>
<b>Résultats</b>	<ul style="list-style-type: none"> <li>◆ Les frontières géologiques prédictibles par apprentissage machine sont limitées aux contrastes présents dans les données de départ. Le levé magnétométrique et le MNE sont efficaces comme couches d'entrée de données brutes dans notre étude ;</li> <li>◆ Les méthodes qui produisent des cartes de probabilités par lithologies sont plus intéressantes que les méthodes de <i>clustering</i> ;</li> <li>◆ Utilisation des textures (produits dérivés) apporte un pouvoir de discrimination supplémentaire, mais n'est pas utile dans les méthodes non supervisées de segmentation ;</li> <li>◆ Utilisation de la lithogéochimie génère du bruit si sa distribution n'est pas homogène ;</li> <li>◆ <i>Random Forest</i> et <i>Extreme Gradient Boosting</i> donnent les prédictions les plus justes et les plus précises.</li> </ul>
<b>Innovations</b>	<ul style="list-style-type: none"> <li>◆ Comparaison d'algorithmes pour la cartographie prédictive ;</li> <li>◆ Développement d'une méthode semi-automatique objective d'unification des codes lithologiques des géofiches du SIGEOM.</li> </ul>
<b>Produits livrés</b>	<ul style="list-style-type: none"> <li>◆ Présentations et rapport</li> </ul>

