

Intelligence artificielle

mini-Projet

Le but de ce projet est de prédire la présence ou de l'absence d'or.

1 Présentation du problème

Le bureau de recherches géologique et minière en charge de la gestion durable des ressources naturelles souhaite mieux comprendre la formation de l'or et mieux prédire la présence ou l'absence de ce minerai. Pour cela, il a constitué à partir d'un système d'information géographique une banque de données constituée de 1000 individus décrits par 24 variables, dont 21 sont quantitatives (20 variables géophysiques et le numéro du site) et 3 sont qualitatives (l'âge de la roche à 3 modalités, le type de la roche à 2 modalités et le type de site à 2 modalités). On cherche en particulier à prédire efficacement la valeur de la variable qualitative site qui informe de la présence ("gisement") ou de l'absence ("stérile") du minerai pour chacun des 1000 sites connus.

2 Travail demandé

Le but du projet est de concevoir et d'entraîner un (ou plusieurs) modèles prédictifs pour prédire au mieux la présence ou de l'absence d'or.

2.1 Réalisation

Les données sont disponibles sur la page Arche dans deux banques de données :

- `gisementLearn.txt` : contient des données d'apprentissage (étiquetées).
- `gisementTestNoLabel.txt` : contient des données à prédire (non étiquetées).

Vous devrez rendre un fichier `gisementTest.txt` qui ajoute comme dernière colonne la prédiction (GISEMENT ou STERILE) de votre meilleur modèle. Le choix du logiciel, de la bibliothèque ou du langage de programmation est libre.

2.2 Rendu

Vous devrez déposer sur la page arche du module avant le vendredi 15 mai une archive contenant un rapport, les sources de vos programmes et le fichier **`gisementTest.txt`**

- Le rapport doit décrire votre démarche en présentant :
 - vos choix (prétraitement des données, paramètres de ou des algorithmes,...) ;
 - les analyses des résultats que vous avez menées ;
 - une conclusion précisant les points forts et les points faibles de l'étude et les améliorations possibles.
- Les sources doivent être accompagnées d'un fichier `README.txt` indiquant comment utiliser vos sources.