

TP Analyse de Données - Projet final

1 Evaluation

L'évaluation sera faite sur la qualité du code de démonstration ainsi que la qualité d'un rapport de deux (2) pages maximum. Ce rapport devra contenir une présentation du code ainsi qu'une analyse des résultats et une conclusion générale. La présentation du code sera la partie principale de l'évaluation !

La date limite de rendu est le **10 juin 2022** (deadline stricte). Les rapports ainsi que le code devront être transmis à l'adresse mail suivante : `c.lemaitre58@gmail.com`.

2 Travaux demandés

1. Récupérer les images à l'adresse `https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php` et en particulier la version du dataset traitée (*processed*).

2. Construire un DataFrame avec les valeurs des 7 moments de Hu (calculées sur les éléments du dataset) ainsi que le numéro de classe (contenu dans les noms des fichiers).

Indication : des fonctions permettant d'ouvrir les images et calculer les 7 moments de Hu sont disponibles dans le paquet `scikit-image`.

3. Sauvegarder les données en CSV.

4. Entraîner un classificateur SVM noyau RBF (avec les paramètres par défaut) sur le DataFrame défini précédemment. Interpréter les résultats obtenus.

Indication : Vous pourrez utiliser l'implémentation des SVM de `scikit-learn`.

5. Réaliser une recherche des paramètres optimaux (C et γ (gamma)) pour le noyau RBF à l'aide du GridSearch.

6. Analyser et conclure à partir des résultats obtenus.