



# Data Science for Copernicus

## Jeux de données

- 1) Images Sentinel-2 téléchargées depuis un DIAS (démonstration).
- 2) Jeu de données vecteur issu de la donnée RPG de l'IGN (Registre Parcellaire Graphique).

## Prérequis

- 1) Notions de base de programmation Python.
- 2) Connaissance des Systèmes d'Information Géographique.
- 3) Notions de base de télédétection optique.

## Environnement technique – Principaux modules

- Ubuntu
- Python 3.8
- GDAL
- Scikit-learn
- Tensorflow
- Matplotlib
- Tiff file
- Geopandas
- Pandas
- Plotly
- Dash
- XGBOOST
- Rasterstats
- Requests
- Jupyter
- Ipykernel



## Cas d'usages

### 1) Jeux de données de parcelles cultivées

#### **Contexte:**

Ce cas d'usage a pour objectif d'initier les participants aux techniques de traitement de jeux de données structurés par machine learning. Le jeu de données structuré est construit à partir d'une donnée vecteur (shapefile) issue du RPG et d'une série temporelle d'images Sentinel 2. Le jeu de données vecteur contient les contours de parcelles graphiques du RPG pour différents types de cultures.

Différentes techniques de machine learning seront appliquées à ce jeu de données structuré, dans le but :

- de détecter des anomalies sur un type de culture donné
- de distinguer par classification plusieurs types de cultures.

#### **Objectifs pédagogiques:**

A l'issue de ce cas d'usage, les participants seront capables :

- de préparer des jeux de données issus de sources multiples, tels que des jeux de données vecteur et des images satellites,
- d'extraire des signatures/caractéristiques à l'aide d'une connaissance thématique à priori du cas d'étude ou à l'aide d'outils de machine learning,
- de choisir une méthode de prétraitement adaptées pour nettoyer le jeu de données,
- de distinguer les apports et limites des différents algorithmes de machine learning afin de faire le bon choix selon la problématique,
- de préparer un jeu de données d'entraînement adapté à un algorithme de machine learning,
- de détecter des anomalies dans un jeu de données à l'aide du machine learning,
- d'entraîner des modèles de machine learning pour mettre en œuvre, paramétrer et valider des processus de classification et de régression
- d'appliquer les modèles issus de la phase d'apprentissage dans le but d'obtenir un produit final exploitable.

### 2) Traitement d'images optiques Sentinel-2

#### **Contexte:**

Ce cas d'usage a pour objectif d'initier les participants au traitement d'images satellites par machine learning et deep learning. Les traitements appliqués seront particulièrement ciblés sur la segmentation sémantique d'images Sentinel 2.

#### **Objectifs pédagogiques:**

A l'issue de ce cas d'usage, les participants seront capables :

- de distinguer les apports et limites des différents algorithmes de machine learning et deep learning dans l'objectif de segmenter des images,
- de préparer des jeux de données adaptés pour l'apprentissage de modèles de machine learning et de deep learning,



- d'entraîner des modèles de machine learning et deep learning pour mettre en œuvre, paramétrer et valider des processus de segmentation d'image,
- d'appliquer les modèles issus de la phase d'apprentissage dans le but d'obtenir un produit final exploitable.

### **Grandes étapes de la démarche :**

Cette rubrique liste l'ensemble des techniques et algorithmes qui seront présentés aux participants. Le panel présenté vise à donner une vue d'ensemble des diverses opérations pouvant être mises en œuvre, même si certaines d'entre elles sont moins communément appliquées que d'autres à des cas pratiques

- **Préparation des données (téléchargement, conversion de format...)**
- **Nettoyage des jeux de données et reconstruction de données manquantes**
  - Calcul de valeurs moyennes et gestion des nodata à l'aide de l'algorithme image inpainting,
  - Reconstruction et interpolation avec l'algorithme K-nearest neighbors (KNN) (gestion des no data),
- **Visualisation de données (Matplotlib, Plotly, Dash)**
- **Extraction de signatures à l'aide de technique de réduction de dimension supervisées et non supervisées:**
  - Analyse en Composantes Principales (ACP)
  - Auto-encodeur variationnel
- **Segmentation de données (non supervisée):**
  - Segmentation par l'algorithme K-Means
  - ⊖ Segmentation par approche SuperPixel avec l'algorithme Mini-Batch K-Means
- **Classification et Segmentation (supervisées):**
  - Algorithme Random forest
  - Algorithme XGBoost
  - Algorithme LSTM
- **Réseau neuronal convolutif (démonstration)**