

# Cahier des charges du dossier

Mathématiques appliquées aux Data Sciences niv. 1

YNOV - Filière Informatique - Spé DataIA - B3

Année scolaire 2024 / 2025

Intervenant : Nicolas Miotto

## Modalités

- ❖ Le projet est à rendre sous un format dossier pdf à rendre le 10 / 03 / 2025 lors de la dernière séance du module.
- ❖ Individuel ou en binôme.
- ❖ Il n'y a pas de présentation orale à préparer.
- ❖ Le dossier doit comporter 10 pages au maximum.
- ❖ Le dossier doit comporter un lien github de vers vos différents scripts python.
- ❖ Le dossier, incluant l'ensemble des scripts, sera noté /20 et représente 50% de la moyenne finale du module (les autres 50% pour le partiel écrit aussi noté /20).

## Partie I : La théorie

1. Qu'est-ce que le machine learning ?
2. Quelles sont les dates-clés du développement de l'IA ?
3. Pourquoi utiliser le machine learning ?
4. Citez des exemples concrets et précis de la vie quotidienne.
5. Quelle est la différence entre machine learning et deep learning ?
6. Expliquez les caractéristiques de ces sous-domaines du ML :
  - 6.1. Apprentissage supervisé : régression
  - 6.2. Apprentissage supervisé : classification
  - 6.3. Apprentissage non supervisé : clustering
  - 6.4. Apprentissage par renforcement

## Partie II : La structure d'un projet de ML

Détaillez chaque étape d'un projet de ML :

1. Définition du problème à résoudre

2. Acquisition des données d'apprentissage
3. Appel aux Data Sciences
  - 3.1. Préparer les données
  - 3.2. Analyser les données : les statistiques
  - 3.3. Nettoyer les données : traitement des données extrêmes et manquantes
4. Choix d'un modèle de prédiction
5. Optimisation du modèle
6. Évaluation du modèle
  - 6.1. Courbes d'erreur
  - 6.2. Courbe d'apprentissage

### Partie III : Les algorithmes à connaître

Pour chacun des algorithmes ci-dessous, vous devez fournir un script en langage python (sauf \*), une explication concise de son fonctionnement, ainsi que les savoirs mathématiques sous-jacents :

1. Algorithme de descente de gradient
2. Méthode des moindres carrés
3. Apprentissage supervisé pour la régression
  - 3.1. Régression par force brute
  - 3.2. La régression linéaire
  - 3.3. Régression polynomiale
  - 3.4. Arbre de décision
  - 3.5. Random Forest
4. Apprentissage supervisé pour la classification
  - 4.1. Régression logistique
  - 4.2. Méthode des plus proches voisins (K-Nearest Neighbours KNN)
  - 4.3. Machines à vecteurs de support et méthodes à noyaux (\*)
  - 4.4. Méthode Bayésienne (\*)
5. Apprentissage non supervisé par clustering
  - 5.1. Méthode des k-moyennes (K-Means)
  - 5.2. Mean-shift
  - 5.3. Modèles de mélange gaussien (GMM) (\*)

### Compétences évaluées (chacune évaluée sur 2 points)

1. Identifier les différentes formes d'apprentissage.
2. Concevoir et appliquer une solution algorithmique efficace.

3. Identifier des modèles algorithmiques (leurs caractéristiques, leurs différences).
4. Utiliser les outils numériques à bon escient pour résoudre un problème.
5. Optimiser un modèle et visualiser des courbes de performances.
6. Effectuer une prédiction.
7. Connaître les différents algorithmes de Machine Learning.
8. Structurer un projet de Machine Learning.
9. Analyser de façon mathématique un modèle algorithmique.
10. Rédiger un dossier complet, précis, propre et structuré.