

Durée : 24 H

1. Objectifs du Projet

L'objectif de ce mini-projet est de consolider les bases de l'apprentissage supervisé à travers l'implémentation et l'analyse des deux modèles fondamentaux :

1. **La Régression Linéaire** pour la prédition de variables continues.
2. **La Régression Logistique** pour les problèmes de classification.

2. Partie 1 : Régression Linéaire

2.1. Données

Utilisez un jeu de données de régression au choix (ex : *Medical Insurance Cost* ou *House Prices* disponibles sur Kaggle). La variable cible doit être numérique.

2.2. Tâches

- Analyser les corrélations via une matrice de chaleur (Heatmap).
- Formaliser le modèle : $y = \beta_0 + \sum \beta_i x_i + \epsilon$.
- Évaluer la performance avec le coefficient de détermination R^2 et l'erreur quadratique moyenne (MSE).
- Interpréter l'importance des variables selon les coefficients β_i .

3. Partie 2 : Régression Logistique

3.1. Données

Utilisez le dataset **Iris** (natif dans *scikit-learn*). Pour cet exercice, vous transformerez le problème en classification binaire (ex : Classe 0 vs Autres).

3.2. Tâches

- Préparer les données et normaliser les variables d'entrée.
- Modéliser la probabilité d'appartenance à l'aide de la fonction sigmoïde : $P(y = 1|x) = \frac{1}{1+e^{-z}}$.
- Évaluer le modèle via une **Matrice de Confusion**.
- Calculer l'Accuracy, la Précision et le Rappel (Recall).

4. Livrables

Les étudiants devront soumettre une archive contenant :

1. **Notebook Python (.ipynb)** : Un code propre, commenté et structuré.
2. **Rapport (PDF)** : Une synthèse de 2 à 3 pages expliquant les choix méthodologiques et l'interprétation des résultats obtenus.