

# Rapport Général du TP : Analyse de Données et Méthodes d'Ensemble

## 1 Introduction

Ce TP vise à maîtriser les bases de l'analyse exploratoire des données, comprendre les techniques de réduction de dimensionnalité, et implémenter des méthodes d'ensemble pour analyser des données simulées et réelles issues d'un élevage de poulets. Les trois parties du TP — Analyse exploratoire des données, Réduction de dimensionnalité, et Méthodes d'ensemble — sont abordées ci-dessous avec une analyse détaillée des résultats et des interprétations.

## 2 Partie 1 : Analyse exploratoire des données

### 2.1 Objectifs

- Calculer des statistiques descriptives et visualiser les distributions des variables `poids`, `nourriture`, et `température`.
- Détecter les outliers avec différentes méthodes et décider de leur traitement.
- Tester la normalité des variables et comparer les moyennes entre groupes.

### 2.2 Données

Les données simulées sont générées avec la commande suivante :

```
data = pd.DataFrame({  
    'poids': np.random.normal(2.5, 0.5, 100),  
    'nourriture': np.random.normal(1.2, 0.3, 100),  
    'température': np.random.normal(25, 2, 100)  
})
```

## 2.3 Exercice 1 : Statistiques descriptives

Les statistiques descriptives pour les variables sont les suivantes :

- **Poids** : moyenne = 2.448, médiane = 2.437, écart-type = 0.454.
- **Nourriture** : moyenne = 1.207, médiane = 1.225, écart-type = 0.286.
- **Température** : moyenne = 25.130, médiane = 25.195, écart-type = 2.169.

Ces statistiques montrent que les moyennes et médianes sont proches, suggérant des distributions symétriques, probablement normales.

## 2.4 Exercice 2 : Détection des outliers

Méthodes utilisées :

- **IQR** : détecte des outliers pour toutes les variables.
- **Z-Score (seuil = 3)** : détecte des outliers pour la température uniquement.

Les visualisations (Figures 1 à 3) montrent des distributions normales avec des outliers plausibles. Les tests de Shapiro-Wilk confirment la normalité.

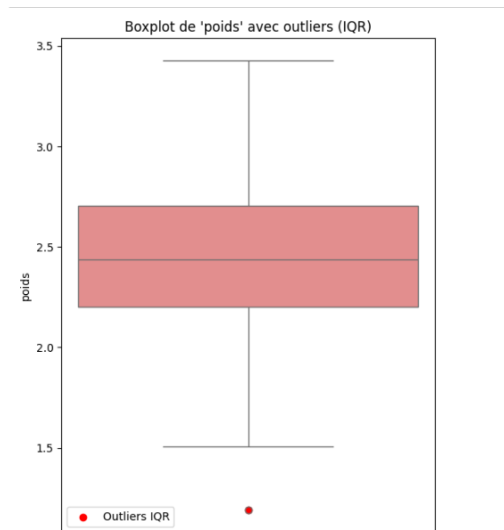


Figure 1: Boxplot du poids avec outliers IQR

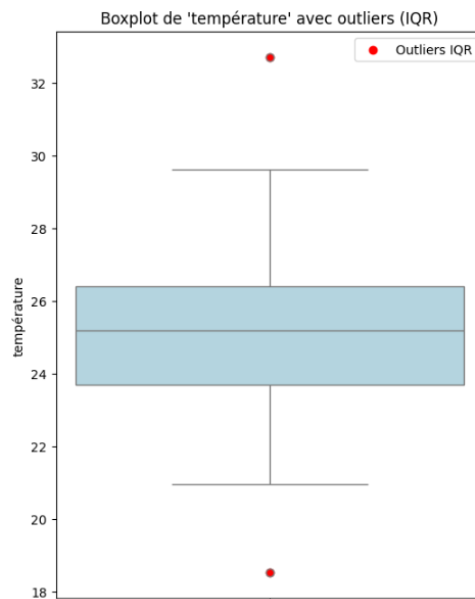


Figure 2: Boxplot de la température avec outliers IQR

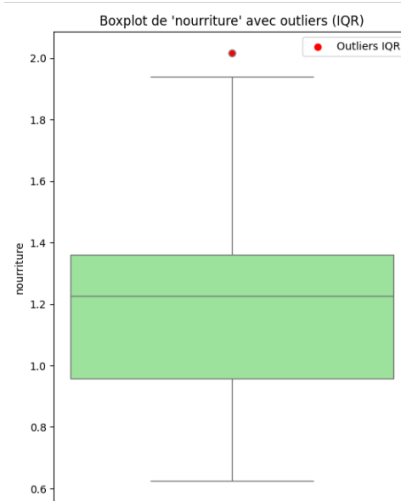


Figure 3: Boxplot de la nourriture avec outliers IQR

## 2.5 Exercice 3 : Tests paramétriques

- **Test de Shapiro-Wilk** : confirme la normalité des trois variables.
- **T-test et ANOVA** : aucune différence significative entre groupes.

## 3 Partie 2 : Réduction de dimensionnalité

### 3.1 Objectifs

- Implémenter une ACP et un KernelPCA.
- Comparer les résultats et justifier le choix des composantes.

### 3.2 Résultats

- **ACP** : Les deux premières composantes capturent 99.97% de la variance totale.
- **KernelPCA** : n'apporte pas d'amélioration significative.

## 4 Partie 3 : Méthodes d'ensemble

### 4.1 Objectifs

- Implémenter une forêt aléatoire et comparer AdaBoost et Gradient Boosting.

### 4.2 Résultats

- **RandomForestClassifier** : Accuracy = 1.000, mais biaisé par un déséquilibre de classe.
- **Boosting** :  $R^2$  négatif, suggérant une faible relation entre variables et cible.

## 5 Conclusion Générale

- **Analyse exploratoire** : les données sont normales avec quelques outliers plausibles.
- **Réduction de dimensionnalité** : ACP est efficace, KernelPCA inutile ici.
- **Méthodes d'ensemble** : La forêt aléatoire est biaisée, et le boosting performe mal.