

ATDN2 TP1

Remy XU M1 OIVM

26 March 2025

Etape 1 : Compréhension du problème

1.1 Décrire les variables disponibles

Nous disposons probablement de données agricoles incluant :

- **Rendement (cible)** : Quantité récoltée par hectare.
- **Précipitations** : Quantité de pluie reçue.
- **Température** : Moyenne des températures pendant la saison.
- **Type de sol** : Catégorisation des sols (ex : sableux, argileux, limoneux).
- **Engrais utilisé** : Quantité et type d'engrais appliqué.

La variable de type de sol est une variable qualitative et les autres variables sont toutes des variables quantitatives.

1.2 Formuler le problème métier

L'objectif est d'optimiser le rendement agricole en identifiant les facteurs qui l'influencent le plus. Cela permettrait d'adapter les pratiques agricoles (type de sol, engrais, irrigation) pour maximiser la production.

1.3 Identifier la variable cible et les variables explicatives

Variable cible : Rendement

Variables explicatives : Précipitations, température, type de sol, engrais utilisé.

1.4 Problématique centrale pour la ferme

Comment la ferme peut-elle ajuster ses pratiques agricoles pour maximiser son rendement tout en minimisant les coûts et en s'adaptant aux conditions climatiques ?

Etape 2 : Analyse statistique descriptive

2.1 Mesures de tendance centrale

```
Moyenne du rendement : 7.38  
Médiane du rendement : 7.35  
Mode du rendement (valeur la plus fréquente) : 3.00
```

Figure 1: Calcul de la moyenne, la médiane, et le mode du rendement.

2.2 Mesures de dispersion

```
Variance du rendement : 6.60  
Écart type du rendement : 2.57  
Étendue du rendement (max - min) : 9.00
```

Figure 2: Calcul de la variance, l'écart-type, et l'étendue du rendement.

2.3 Visualisation des données

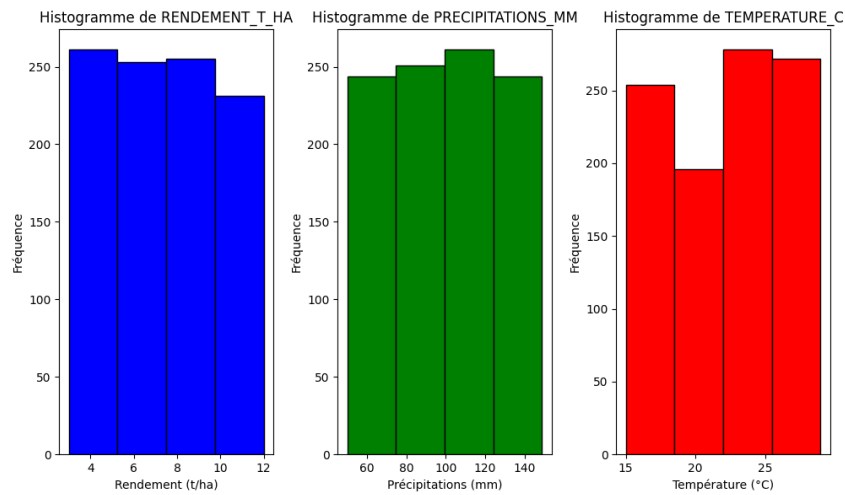


Figure 3: Histogrammes des variables rendement, précipitations, et température.

2.4 Corrélations

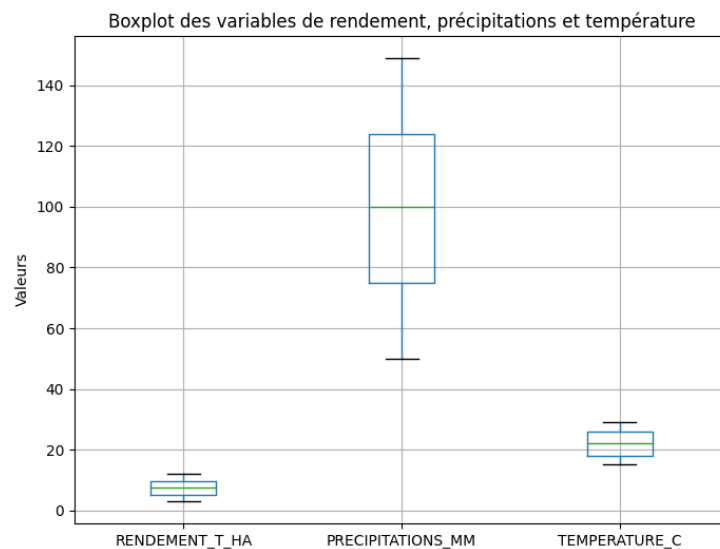


Figure 4: Boîtes à moustaches des variables rendement, précipitations, et température.

Sur l'ensemble des boîtes à moustaches, nous ne constatons aucun point en dehors des boîtes. Il n'y a donc aucun *outlier* ou valeur aberrante nécessitant un traitement particulier.

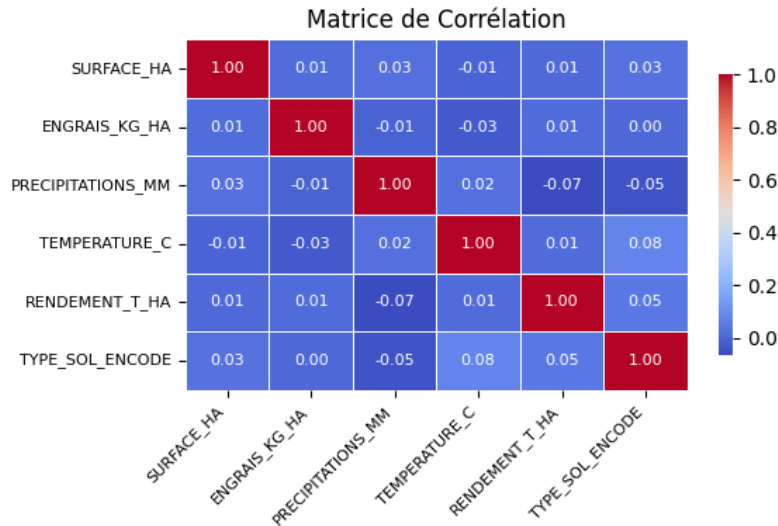


Figure 5: Heatmap de la matrice de corrélation de notre jeu de données.

Les variables qui semblent avoir le plus d'impact sur le rendement sont principalement le type de sol et l'utilisation d'engrais.

Étape 3 : Analyse de la variance (ANOVA)

3.1 Hypothèses

- **H0** : Le type de sol n'influence pas le rendement.
- **H1** : Le type de sol influence le rendement.

3.2 Test ANOVA et interprétation de la p-value

La p-value obtenue lors du test ANOVA indique si le type de sol a une influence significative sur le rendement.

- **Si p-value ≤ 0.05** : On rejette l'hypothèse nulle (H0), ce qui signifie que le type de sol a un impact significatif sur le rendement.
- **Si p-value > 0.05** : On ne rejette pas H0, et on ne peut pas conclure que le type de sol influence le rendement de manière significative.

Étape 4 : Modélisation et évaluation

4.1 Évaluation des modèles

Les modèles sont évalués à l'aide des métriques suivantes :

- **MAE** : Erreur absolue moyenne (plus faible = meilleur modèle).
- **RMSE** : Erreur quadratique moyenne (privilégie les grosses erreurs).
- **R²** : Variance expliquée (plus proche de 1, mieux c'est).

Un modèle performant sera celui avec un faible MAE et RMSE, et un R² élevé. Si les données sont linéaires, une **régression linéaire** peut être efficace. Pour des relations plus complexes, un **Random Forest** ou **XGBoost** peut mieux capturer les interactions.

Étape 5 : Recommandations et amélioration du modèle

5.1 Recommandations pour augmenter le rendement

- **Optimisation du sol** : Privilégier les types de sol les plus performants.
- **Utilisation efficace des engrais** : Tester différents dosages et types.

- **Gestion de l'irrigation** : Ajuster en fonction des précipitations.
- **Adaptation aux conditions climatiques** : Ajuster les périodes de culture.

5.2 Limites du modèle et pistes d'amélioration

- **Données manquantes ou biaisées** : Vérifier leur représentativité.
- **Modèle trop simple** : Essayer des modèles plus complexes si R^2 est faible.
- **Ajout de nouvelles variables** : Qualité des semences, rotation des cultures.
- **Validation croisée** : Vérifier la robustesse du modèle.