

# 1. Analyse exploratoire : structuration des données (ACP)

L'Analyse en Composantes Principales (ACP) réalisée sur le dataset *crop\_yield* montre que la variabilité du jeu de données est avant tout structurée par deux dimensions majeures :

## 1 Les conditions météorologiques encodées

Les modalités *Sunny* et *Rainy* dominent la première composante (PC1), créant un axe de variation principalement lié à la météo.

Cette forte influence provient du fait que les variables catégorielles encodées (One-Hot Encoding) introduisent des contrastes nets entre observations.

## 2 La localisation géographique (Nord, Sud, Ouest)

La seconde composante (PC2), puis PC3, mettent en évidence des clusters géographiques distincts.

Les régions apparaissent donc comme un facteur structurant important du dataset.

---

## Faible contribution des variables continues

Les variables quantitatives telles que :

- **Rainfall\_mm**
- **Temperature\_Celsius**
- **Fertilizer\_Used**
- **Irrigation\_Used**

n'apparaissent presque pas dans les premières composantes.

Cela s'explique par la nature **synthétique** du dataset, où ces variables sont réparties de manière homogène, ce qui réduit leur pouvoir de structuration dans une analyse non supervisée.

---

## Interprétation métier

L'ACP ne cherche pas à expliquer le rendement, mais à comprendre comment les **variables explicatives se structurent entre elles**.

Ainsi :

- Les **variables catégorielles météo et région** constituent les axes majeurs d'organisation des données.
- Les **variables agronomiques continues**, pourtant essentielles pour le rendement, ne ressortent pas ici — ce qui est **normal**, puisqu'un dataset simulé homogénéise ces effets et que l'ACP ignore la variable cible.

---

L'ACP met en évidence les **dimensions structurantes** du dataset, sans identifier les facteurs explicatifs du rendement.

Cette distinction est essentielle :

- **ACP = variabilité interne des données (météo, région)**
- **Modélisation supervisée = identification des variables qui expliquent réellement le rendement**

C'est cette seconde étape qui permet d'identifier les variables clés influençant la performance agricole.

## 2. Résultats du modèle et analyse des facteurs influençant le rendement

### 2.1 Performances du modèle choisi

Plusieurs modèles ont été entraînés et comparés, dont une Régression Linéaire et un Random Forest optimisé avec Optuna.

Les résultats montrent clairement la supériorité du modèle linéaire :

Modèle	RMSE ↓	MAE ↓	R <sup>2</sup> ↑	Paramètres	Commentaire
Linear Regression	0.499	0.398	0.913	<code>fit_intercept=True</code> (autres paramètres par défaut)	🎯 <b>Meilleur modèle global</b> – précis, robuste et très interprétable
Random Forest (Optuna)	0.553	0.441	0.894	<code>n_estimators=75</code> <code>max_depth=15</code> <code>max_features="sqrt"</code> <code>min_samples_split=3</code> <code>min_samples_leaf=3</code>	Bon modèle non linéaire, mais moins performant sur ce dataset

## Interprétation métier

- La **Régression Linéaire** explique **plus de 91 % du rendement observé**, ce qui en fait un modèle fiable et très transparent.
- De manière surprenante, elle **surpasse le Random Forest**, pourtant plus complexe. Cela indique que les relations entre les variables agronomiques et le rendement sont **principalement linéaires** dans notre dataset.

## Pourquoi un modèle linéaire bat un Random Forest ?

1. Le dataset contient de nombreuses variables catégorielles encodées (One-Hot).  
→ Les modèles linéaires sont très adaptés à ces représentations haute dimension.
2. Les relations entre variables et rendement semblent globalement **simples et linéaires**.
3. Le Random Forest lisse certaines variations fines... qui étaient utiles ici.

---

## 2.2 Variables clés influençant le rendement agricole

L'analyse des coefficients du modèle linéaire révèle **quatre variables principales**, avec un poids très nettement supérieur aux autres.

### 1 Pluviométrie (Rainfall\_mm) — La variable la plus déterminante

C'est le facteur ayant l'impact le plus fort sur le rendement.

Plus il pleut (ou plus l'accès à l'eau est assuré), meilleur est le rendement.

**Impact métier :**

- Les régions naturellement arrosées disposent d'un avantage productif.
  - En zone sèche, l'irrigation devient un levier stratégique indispensable.
- 

**2 Utilisation d'engrais — Gain de rendement majeur**

Les parcelles fertilisées affichent systématiquement les meilleurs rendements.

**Impact métier :**

- Un plan de fertilisation bien dimensionné améliore fortement la productivité.
  - C'est l'un des investissements les plus rentables pour l'agriculteur.
- 

**3 Irrigation — Un soutien crucial, surtout en climat sec**

L'irrigation contribue presque autant que les engrais.

**Impact métier :**

- Permet de stabiliser les rendements malgré les aléas climatiques.
  - Très efficace dans les zones arides ou en années de faible pluviométrie.
- 

**4 Température — Effet modéré mais réel**

L'impact est moins important, mais une température adéquate favorise la croissance.

**Impact métier :**

- Importance de choisir des cultures adaptées au climat local (ex. : maïs et riz apprécient des températures élevées).
- 

## **2.3 Variables secondaires (impact faible)**

Le modèle montre que les variables suivantes ont un effet beaucoup plus limité :

- **Type de sol**
- **Région**
- **Type de culture**

Cela signifie que dans ce dataset :

- les sols et régions sont relativement homogènes,
- la **gestion de l'eau et des engrais explique beaucoup plus la variation du rendement** que la localisation,
- les conditions de culture importent davantage que la culture elle-même.

#### **Impact métier :**

Le potentiel de rendement dépend principalement de la maîtrise des ressources (eau, fertilisation), davantage que du type de sol ou de la région.

- 
- Le modèle linéaire offre d'excellentes performances et une forte lisibilité.
  - Trois leviers agronomiques dominant l'explication du rendement :  
**pluviométrie → engrais → irrigation.**
  - Ces résultats permettent de formuler des recommandations concrètes pour guider l'agriculteur vers des décisions optimales.

## **3. Recommandations agronomiques pour optimiser rendement et profit**

À partir des résultats du modèle et de l'analyse des variables clés, plusieurs recommandations concrètes peuvent être formulées pour aider les exploitants à maximiser leur rendement et leur rentabilité.

- 
- 1 Prioriser les parcelles bénéficiant d'un bon accès à l'eau (pluie ou irrigation)**

Les parcelles situées dans des zones **naturellement arrosées** présentent un fort avantage productif.

En conditions sèches, l'installation ou le renforcement d'un **système d'irrigation** devient indispensable pour maintenir un rendement compétitif.

→ **L'eau est le premier facteur de performance agricole.**

---

## **2 Optimiser l'usage des engrais**

L'utilisation d'engrais constitue l'un des **leviers les plus efficaces** pour augmenter le rendement.

Une stratégie de fertilisation adaptée à chaque culture permet d'améliorer significativement la productivité.

→ **Un plan de fertilisation raisonné = gains rapides et mesurables.**

---

## **3 Investir dans l'irrigation pour stabiliser et augmenter les rendements**

L'irrigation apporte un gain important, surtout en climat sec ou variable.

Les investissements réalisés dans ce domaine offrent généralement un **retour sur investissement rapide**, car ils sécurisent la croissance des cultures.

→ **L'irrigation est un amortisseur de risque climatique.**

---

## **4 Choisir des cultures adaptées aux conditions climatiques locales**

- Cultures performantes sous **hautes températures** : *maïs, riz*.
- Cultures adaptées à des climats **plus tempérés** : *blé, orge*.

L'outil développé dans ce projet permet d'automatiser cette analyse en recommandant la **culture la plus rentable** selon les conditions d'une parcelle.

→ **Adapter la culture au climat maximise le rendement prévisible.**

---

## 5 Se concentrer sur la gestion de l'eau et des intrants plutôt que sur le type de sol

Le modèle montre que des facteurs comme le type de sol ou la région ont un **impact bien plus faible** sur le rendement que la pluie, l'irrigation et la fertilisation.

Les efforts d'optimisation doivent donc se concentrer sur :

- la **gestion hydrique**,
- la **fertilisation**,
- le **choix stratégique des cultures**.

→ **Le sol compte, mais l'eau et les intrants comptent beaucoup plus.**

---

Cette étude confirme que les trois facteurs les plus déterminants pour maximiser le rendement agricole sont :

- l'accès à l'eau (pluie + irrigation)
- la fertilisation
- le choix de la culture adaptée.

Les exploitants ont donc tout intérêt à :

- sécuriser l'apport hydrique via l'irrigation,
- optimiser les apports nutritifs (engrais),
- utiliser l'outil d'aide à la décision développé dans ce projet pour sélectionner la culture la plus rentable selon leurs conditions spécifiques.

## Conclusion

Ces recommandations offrent une **base solide pour améliorer la profitabilité**, réduire les risques climatiques et renforcer la résilience des exploitations.

Ce travail a permis de développer une solution complète d'aide à la décision agricole, combinant analyse statistique, modélisation prédictive et déploiement opérationnel via une API et une interface utilisateur.

L'exploration des données et l'Analyse en Composantes Principales ont révélé les

dimensions structurantes du dataset, essentiellement liées à la météo et à la localisation. La modélisation supervisée a ensuite mis en évidence les véritables leviers d'action sur le rendement : la pluviométrie, l'usage d'engrais et l'irrigation, bien loin devant les facteurs tels que le type de sol ou la région.

Le modèle final, une Régression Linéaire, offre une performance solide et interprétable, expliquant plus de 90 % de la variabilité du rendement. Sa simplicité et sa transparence en font un outil fiable pour produire des recommandations exploitables sur le terrain. Grâce à l'application développée, les agriculteurs peuvent désormais estimer rapidement le rendement d'une culture ou identifier celle qui serait la plus rentable selon les conditions spécifiques de leur parcelle.

Au-delà de la performance technique, ce projet apporte une véritable valeur métier : il améliore la prise de décision, réduit l'incertitude et aide à orienter les investissements vers les leviers les plus efficaces, notamment l'eau et la fertilisation.

Il ouvre également la voie à des évolutions futures comme l'intégration de données réelles, la prise en compte des prix de marché ou encore la personnalisation des recommandations à l'échelle de l'exploitation.

En synthèse, cette solution constitue un socle robuste pour accompagner les agriculteurs dans la gestion de leurs cultures et optimiser durablement rendement et rentabilité.