

# Rapport Synthétique Métier

## Système de Prédiction de Rendement Agricole

---

### Résumé

Ce rapport présente les résultats d'un système intelligent de prédiction de rendements agricoles, conçu pour aider les agriculteurs à **maximiser la production agricole et sa rentabilité**.

#### A retenir :

1. **Prédiction fiable** : Notre système prédit les rendements avec une précision de 96%, avec une marge d'erreur moyenne de seulement 1 tonne par hectare.
  2. **Variables clés identifiées** : Le type de culture et le pays sont les facteurs les plus déterminants, suivis par la température, l'utilisation de pesticides et les précipitations.
  3. **Impact mesurable des pratiques agricoles** : L'irrigation et la fertilisation peuvent augmenter les rendements de 1.2 à 1.5 tonnes par hectare respectivement.
- 

### Objectifs du Projet

Les agriculteurs font face à plusieurs défis :

- **Incertitude sur les rendements** : Difficultés à prévoir la production à l'avance.
- **Choix des cultures** : Quelle type de culture planter pour maximiser les profits ?
- **Investissements technologiques** : Faut-il investir dans l'irrigation et/ou la fertilisation pour améliorer les rendements ?
- **Volatilité des prix** : Comment s'adapter aux fluctuations des prix des matières premières sur le marché ?

Nous avons développé un outil intelligent d'aide à la décision qui :

- Prédit les rendements attendus pour chaque culture
  - Recommande les cultures les plus productives
  - Optimise la rentabilité selon les prix du marché
  - Simule l'impact des investissements agricoles sur le rendement.
-

# Résultats Principaux

## 1. Performance du Système de Prédition

Notre système atteint une **précision de 95.76%** dans ses prédictions, ce qui signifie :

Indicateur	Signification Concète
<b>Précision : 96%</b>	Sur 100 prédictions, 96 sont très proches de la réalité
<b>Erreur moyenne : 1.07 t/ha</b>	Si on prédit 40 t/ha, le résultat réel serait entre 39 et 41 t/ha dans la majorité des cas
<b>Fiabilité élevée</b>	Le système ne "sur-apprend" pas : il fonctionne aussi bien sur de nouvelles données

**En pratique :** Un agriculteur peut planifier avec précision sa production, ses ventes, estimer sa rentabilité et ses besoins en stockage grâce à l'outil développé.

## 2. Exemple Concret : Culture de Maïs en France

**Scénario :** Un agriculteur français veut estimer son rendement de maïs pour 2026

**Conditions :**

- Pays : France
- Culture : Mais
- Année : 2026
- Précipitations moyennes : 650 mm
- Pesticides utilisés : 5,000 tonnes
- Température moyenne : 15°

**Résultats :**

Scénario	Rendement Prédit	Revenu Potentiel (à 180€/t)
<b>Rendement sans irrigation ni fertilisation</b>	5.05 t/ha	910 €/ha
<b>Avec irrigation</b>	6.25 t/ha	1126 €/ha
<b>Avec fertilisation</b>	6.55 t/ha	1180 €/ha
<b>Avec irrigation + fertilisation</b>	7.75 t/ha	1396 €/ha

## **Recommandation :**

Le rendement agricole est estimé à 5 t/ha pour cet agriculteur, avec un revenu de 910 €/ha. Notre agriculteur pourrait augmenter son rendement pour atteindre plus de 7 t/ha et près de 1400 € de revenu par ha en investissant dans des moyens technologiques avancés tels que la fertilisation et l'irrigation. (Soit près de + 40%).

---

## **Explication des Variables Clés**

Nous avons analysé l'importance de chaque facteur présent dans le dataset sur les rendements. Voici les résultats tirés de notre analyse :

### **1- Type de Culture :**

C'est **le facteur le plus important** dans la prédiction des rendements. Certaines cultures produisent naturellement plus que d'autres comme les pommes de terre qui sont connues pour leur fort rendement agricole.

### **2- Pays :**

Le pays (climat, qualité des sols, savoir-faire local, avancée technologiques...) joue un rôle important dans les rendements. Un même type de culture peut avoir des rendements très différents selon le pays.

**Exemple :** Le rendement du blé en France est généralement supérieur à celui en Afrique subsaharienne, principalement en raison du climat et des techniques agricoles.

### **3- Température Moyenne :**

Chaque culture a une plage de température optimale :

- **Cultures tempérées** (blé, orge) : 10-20°C
- **Cultures tropicales** (manioc, igname) : 25-30°C
- **Cultures intermédiaires** (maïs, riz) : 20-25°C

### **4- Utilisation de Pesticides**

Les pesticides protègent les cultures contre les ravageurs et maladies, ce qui se traduit par de meilleurs rendements.

### **5- Précipitations**

L'eau est essentielle en agriculture, mais le niveau d'eau nécessaire dépend de la culture :

- **Riz** : Besoin de beaucoup d'eau (1,200-1,500 mm/an)
  - **Maïs** : Besoins modérés (500-800 mm/an)
  - **Sorgho** : Résistant à la sécheresse (400-600 mm/an)
- 

## Recommandations Agronomiques

La maximisation des rendements repose sur :

### 1. Le choix de la culture :

Notre système permet de comparer toutes les cultures possibles pour les mêmes conditions climatiques, afin de choisir la meilleure.

**Exemple** : Pour la France en 2026 (conditions moyennes) :

Classement	Culture	Rendement Prédit
1er	Pommes de terre	19.3 t/ha
2ème	Patate douce	13 t/ha
3ème	Ignames	11.5 t/ha
4ème	Plantains	10.8 t/ha
5ème	Manioc	8.2 t/ha

### 2. L'usage de l'Irrigation :

L'utilisation de l'irrigation a un impact moyen de : +12,000 kg/ha (+1.2 t/ha), mais elle a également un coût d'investissement à prendre en considération.

L'utilisation de l'irrigation serait utile dans les cas suivants :

- Précipitations naturelles insuffisantes (<500 mm/an)
- Cultures à haute valeur ajoutée
- Pays souffrant de sécheresse durable.

### 3. L'usage de la Fertilisation

L'utilisation de la fertilisation a un impact moyen de +15,000 kg/ha (+1.5 t/ha), mais dépend grandement de la qualité du sol. Le recours à la fertilisation pourrait être utile pour :

- Les cultures à fort rendement
- Les sols appauvris

#### 4. Combiner Irrigation + Fertilisation

C'est l'option la plus performante +27,000 kg/ha (+2.7 t/ha) car les deux pratiques se renforcent mutuellement. L'eau permet aux nutriments d'être mieux absorbés, et les nutriments permettent à la plante de mieux utiliser l'eau. Mais cette option engendre un cout élevé à prendre en considération, et correspond principalement à des exploitations de grande taille, ou à des cultures à forte valeur ajoutée.

---

## 💰 Optimisation de la Rentabilité

Un rendement élevé ne garantit pas systématiquement une bonne rentabilité car :

- **Les prix de vente** varient selon la demande sur le marché
- **Les coûts de production** : Semences, intrants, main d'œuvre
- **La demande sur le marché** : Demande locale, export, congélation...

### Exemple d'Optimisation : Comparaison Rendement vs Profit

Scénario : France 2026, sans irrigation ni fertilisation

Culture	Rendement (t/ha)	Prix (eur_per_t)	Rentabilité (€/ha)
Sweet potatoes	13.02	1400	18234.45
Cassava	8.24	1300	10707.44
Yams	11.47	600	6883.28
Plantains and others	10.87	450	4893.58
Potatoes	19.28	250	4821.12
Rice, paddy	5.6	520	2911.02
Soybeans	1.91	950	1809.99
Maize	5.06	188	951.12
Wheat	3.22	250	804.51
Sorghum	2.83	150	424.19

💡 Irrigation : ✗ Non | 🌱 Fertilisation : ✗ Non

Nous constatons que certaines cultures sont plus rentables financièrement malgré le faible rendement agricole comme le Manioc.

---

# Choix Techniques

## 1- Stratégie de Fusion des Données

Nous avions plusieurs sources de données à notre disposition :

- crop\_yield.csv : Ce dataset contient différentes informations agricoles.
- yield.csv : Rendement historique annuel par pays, et culture, entre 1961 et 2016, avec un volume de données de 56708
- yield\_df.csv : Rendement historique annuel par pays, et culture (avec pesticides, niveau de pluie et de température) entre 1990 et 2013 avec un volume de données de 28242.
- pesticides.csv : Usage de pesticides annuels par tonne et par pays entre 1990 et 2016, pour un volume de données de 4349.
- rainfall.csv : pluviométrie annuelle par pays entre 1985 et 2017, avec un volume de données de 6603.
- temperatures.csv : température annuelle par pays entre 1743 et 2013, avec un volume de données de 71311.

Nous avons fait un travail d'harmonisation de ces dataset qui consiste à :

- Éliminer les valeurs incohérentes (précipitations ou rendements négatifs par exemple.)
- Neutraliser les valeurs aberrantes.
- Suppression des informations, colonnes inutiles et des doublons.
- Choix de la période d'analyse (de 1990 à 2013), car c'est sur cette période que nous disposons d'un plus grand nombre de data utile pour notre analyse.
- Gestion des valeurs manquantes.

Puis nous avons procédé à la fusion de ces fichiers par le couple de critères : pays + année, en s'assurant de ne pas avoir de valeurs différentes de températures, pluie, et pesticides pour la même année x pays.

Nous avons obtenu un dataset final de 28242 observations, portant sur 101 pays, sur une période allant de 1990 à 2013, et sur 10 différents types de culture.

Nous avons également généré deux variables proxy, à savoir la fertilisation et l'irrigation, que nous intégrons au calcul du rendement financier de chaque culture.

## 2- Modélisation :

Nous avons entraîné et évalué 4 différents modèles réputé pour être performants pour ce type de problématique (Ridge, Random Forest, XGBoost, HistGradientBoosting), et nous avons également travailler sur un modèle de base pour avoir des valeurs minimums de comparaison (Dummy).

Nous avons utilisé les métriques suivantes pour comparer les modèles : R2, RMSE, et MAE.

Les modèles qui ont montré une forte performance à l'issu de l'évaluation sont :

- Random Forest:
  - R2 = 0.998
  - RMSE = 2935.42 hg/ha
  - MAE = 915.93 hg/ha
- XGBoost:
  - R2 = 0.985
  - RMSE = 9762.46 hg/ha
  - MAE = 6102.46 hg/ha
- HistGradientBoosting:
  - R2 = 0.933
  - RMSE = 14782.71
  - MAE = 9042.23 hg/ha

Puis nous avons effectué une optimisation des paramètres de ces 3 modèles, en testant 50 combinaisons d'hyperparamètres différentes pour trouver les meilleurs paramètres donnant les meilleurs résultats. A l'issu de cette analyse, le meilleur modèle que nous avons retenu est :

- **HistGradientBoosting:**

- R<sup>2</sup> = 0.96
- RMSE = 19665.07 hg/ha
- MAE = 10714.73 hg/ha

Ce modèle offre :

- Une meilleure précision (96%)
- Une Erreur la plus faible (10,715 h/ha)
- Peu de sur-apprentissage (écart train/test de seulement 3.7%)
- Rapide à exécuter : Prédictions en quelques millisecondes
- Robuste : Gère bien les valeurs manquantes et les relations complexes

### **3- Variables nécessaires à la prédiction :**

Voici les variables indispensables pour générer une prédiction :

- Pays (area)
- Type de culture (item)
- Année (year)
- Précipitations moyennes (avg\_rain\_mm)
- Utilisation de pesticides (pesticides\_tonnes)
- Température moyenne (avg\_temp)

### **Variables exclues :**

- Code pays/culture : Redondant avec les noms
- Données manquantes : Imputées avec des valeurs médianes/fréquentes

### **Préparation des données :**

- Variables catégorielles → Encodage one-hot (transformation en 0/1)
  - Variables numériques → Standardisation (mise à l'échelle)
- 

## **Système de prédiction :**

Nous avons mis en place 3 modes de prédiction pour répondre aux besoins des agriculteurs :

### **1- Prédiction Simple :**

Permet de calculer le rendement pour une culture précise (en hg/ha et t/ha) et sa rentabilité financière si le prix est fourni.

#### **Étapes :**

1. Sélectionnez votre culture (ex: maïs)
2. Indiquez vos conditions (pays, année, climat, pesticides)
3. Cochez irrigation/fertilisation si applicable
4. Obtenez votre prédiction en hg/ha et t/ha
5. Saisissez le prix de vente (optionnel)
6. Obtenez votre prédiction de rentabilité en euros

## **2- Recommandation par Rendement :**

Permet d'obtenir un classement des prédictions des cultures ayant le plus fort rendement.

### **Étapes :**

1. Indiquez vos conditions (pays, climat, etc.)
2. Le système compare toutes les cultures
3. Vous obtenez un classement des 5-10 meilleures cultures en termes de rendement.

## **3- Recommandation par Rentabilité**

Permet d'obtenir un classement des prédictions des cultures ayant la plus forte rentabilité financière.

### **Étapes :**

1. Indiquez vos conditions
2. Saisissez les prix de vente actuels pour chaque culture
3. Le système calcule le revenu potentiel par hectare
4. Vous obtenez un classement par rentabilité

---

## **⚠️ Limites du système**

### **Avantages :**

- Prédictions des rendements moyens attendus
- Comparaisons des cultures entre elles
- Simulation de l'impact de l'irrigation/fertilisation
- Optimisation selon les prix du marché

### **Limites :**

- Les prédictions ne tiennent pas compte des événements exceptionnels (sécheresse extrême, inondations...)
- Le système ne tient pas compte des maladies/ravageurs spécifiques
- Le système ne prend pas en considération les spécificités liées à la qualité du sol.
- Le rendement financier est brut de toutes charges liées à l'exploitation.
- Tenir compte des éventuelles fluctuations des prix.

Exemples de cas d'usage :

- Planification de la saison : Que planter cette année pour maximiser les profits.
- Décision d'investissement : Agriculteur envisageant l'irrigation.
- Adaptation au changement climatique : Quelles cultures recommander face à la hausse des températures ?

---

**Version du rapport :** 1.0

**Dernière mise à jour :** Janvier 2026

**Auteur :** Fatima-Zahra BARHOU

---

*Ce système est un outil d'aide à la décision. Les résultats doivent être interprétés en complément de votre expertise agronomique et de votre connaissance du terrain. Les prédictions sont basées sur des données historiques et des modèles statistiques, et ne peuvent garantir les rendements futurs.*