



PROJET

# CLASSIFICATION DES TYPES DE CONSOMMATION DES MÉNAGES HAÏTIENS

PROJET RÉALISÉ EN BINÔME  
**SAINT GERMAIN EMODE ET DARLENS DAMISCA**

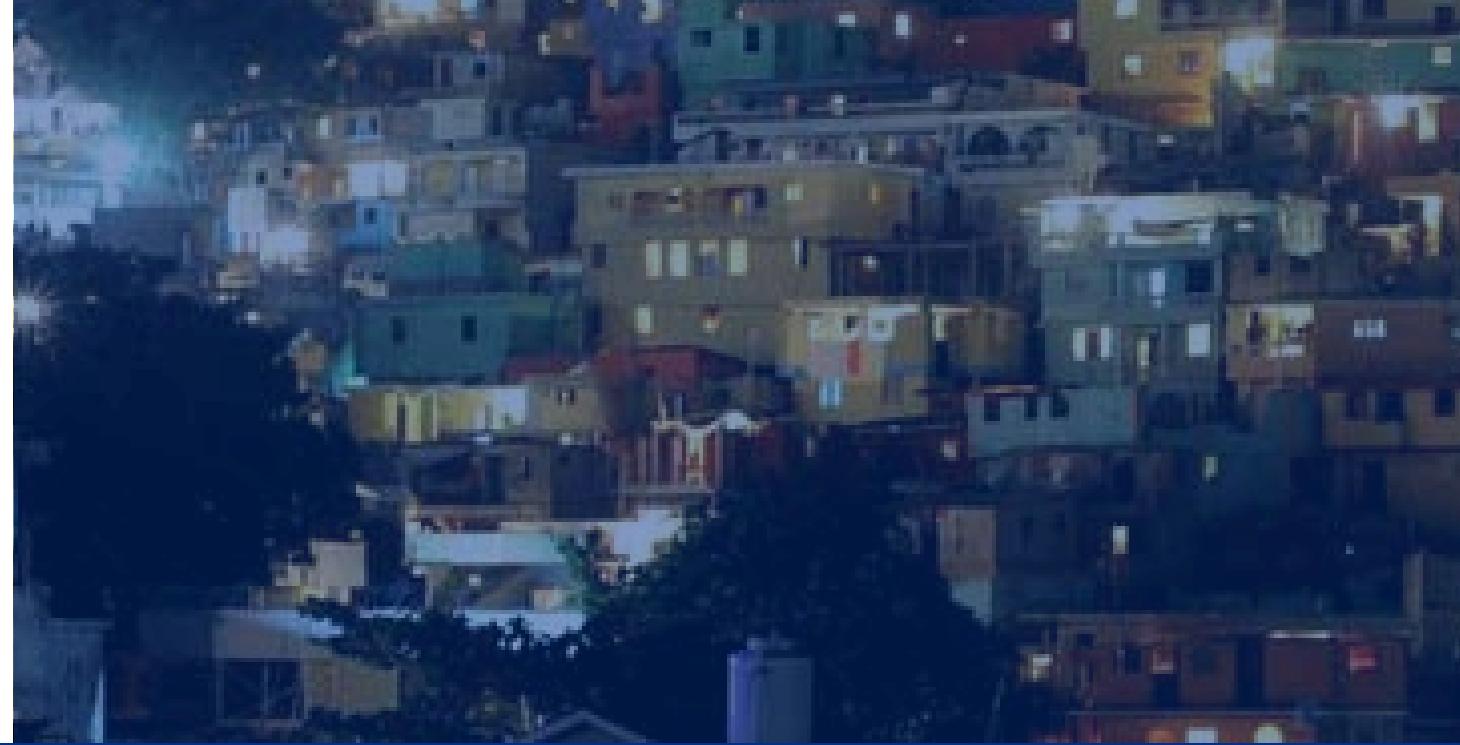
# CONTEXTE DU PROJET

En Haïti, la consommation d'électricité varie fortement selon les foyers.

Sans classification claire, il est difficile d'évaluer les besoins réels et de planifier efficacement la distribution énergétique.

Les données issues de compteurs intelligents permettent aujourd'hui d'analyser ces comportements et d'appliquer l'intelligence artificielle pour regrouper les ménages selon leur niveau de consommation.

# OBJECTIF GÉNÉRAL



Développer un modèle d'apprentissage automatique capable de classifier les ménages haïtiens selon leur niveau de consommation électrique moyenne, en trois catégories :

- Faible consommation
- Moyenne consommation
- Forte consommation

2

Ce modèle permettra d'améliorer la planification énergétique et de promouvoir une gestion durable de l'électricité.

# OBJECTIFS SPÉCIFIQUES



- Analyser les profils de consommation à partir des données collectées.
- Créer des indicateurs statistiques et temporels pour décrire la consommation des foyers.
- Construire un modèle prédictif fiable malgré le déséquilibre des classes.
- Fournir un outil exploitable pour les entreprises et les institutions énergétiques.

# DESCRIPTION DU DATASET

Dans le cadre de ce projet, nous avons eu accès à un jeu de données privé fourni par la société Sigora, contenant des informations détaillées sur les compteurs des clients situés dans plusieurs communes du Nord-Ouest d'Haïti, telles que Môle Saint-Nicolas, Jean Rabel, Bombardopolis et Môle Rouge.

Ce jeu de données, couvrant la période de janvier 2023 à septembre 2025, permet d'analyser les profils réels de consommation et de développer un modèle d'intelligence artificielle capable de classifier les ménages en trois catégories : faible, moyenne et forte consommation.

# PRÉPARATION DU DATASET

- Nettoyage et imputation des valeurs manquantes (moyenne ou mode).
- Suppression des doublons et des variables fortement corrélées à la cible.
- Standardisation des variables numériques.
- Encodage One-Hot des variables catégorielles.

|   | numero_compteur | nom_complet              | mean_amperage | std_amperage | max_amperage | min_amperage | count_readings | argent_total | capacity_mode | tension_v | prix_mode | wifi_mode | total_transactions | sum_tx_amount | sum_tx_in | sum_tx_out | zon         |
|---|-----------------|--------------------------|---------------|--------------|--------------|--------------|----------------|--------------|---------------|-----------|-----------|-----------|--------------------|---------------|-----------|------------|-------------|
| 0 | #0054           | Jean-Baptiste jeandeline | 0.036000      | 0.004899     | 0.04         | 0.03         | 5              | 0.001800     | 120           | 120.0     | None      | 31        | 1                  | 25.0          | 0.0       | 25.0       | Zon Inconnu |
| 1 | 0010713         | PLANCHER Rochenel        | 0.148571      | 0.103982     | 0.50         | 0.08         | 14             | 0.018360     | 120           | 120.0     | None      | 31        | 1                  | 25.0          | 0.0       | 25.0       | Zon Inconnu |
| 2 | 001jp           | Jp                       | 1.388255      | 0.671161     | 2.82         | 0.00         | 4091           | 53.629710    | 120           | 120.0     | None      | 31        | 53                 | 4525.0        | 0.0       | 4525.0     | Zon Inconnu |
| 3 | 010591          | Paul Camie Ulysse        | 0.325500      | 0.189907     | 0.96         | 0.00         | 80             | 0.303800     | 120           | 120.0     | None      | 31        | 1                  | 25.0          | 0.0       | 25.0       | Zon Inconnu |
| 4 | 010796          | Fortunat Renet           | 0.071765      | 0.122294     | 0.64         | 0.01         | 34             | 0.028467     | 120           | 120.0     | None      | 31        | 1                  | 25.0          | 0.0       | 25.0       | Zon Inconnu |

# CRÉATION DU LABEL

Le label est défini selon la puissance moyenne (kW) :

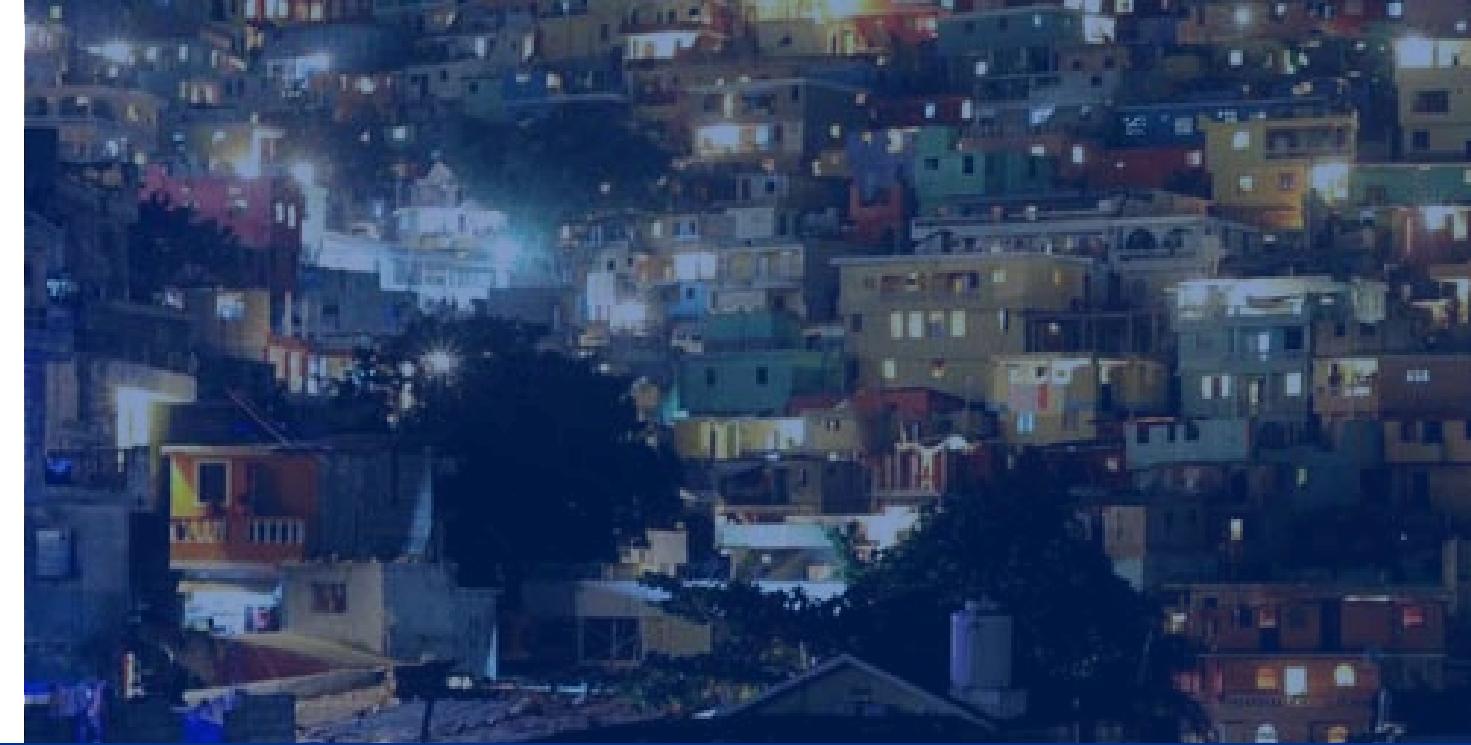
- < 0.05 kW → Faible consommation
- 0.05 à 0.5 kW → Moyenne consommation
- 0.5 kW → Forte consommation

Répartition observée :

- Faible : 1951 foyers
- Moyenne : 748 foyers
- Forte : 17 foyers

Forte consommation = classe rare → nécessite une gestion spéciale.

# GESTION DU DÉSÉQUILIBRE DES CLASSES



Application de pondérations automatiques des classes :

- Forte = 52×
- Moyenne = 1.2×
- Faible = 0.46×

Stratégies utilisées :

- `class_weight` pour Random Forest & Logistic Regression
- `scale_pos_weight` pour XGBoost

7

Cette méthode renforce la détection des cas rares sans fausser le modèle

# MÉTHODOLOGIE DE MODÉLISATION

1-Extraction de caractéristiques (features) :

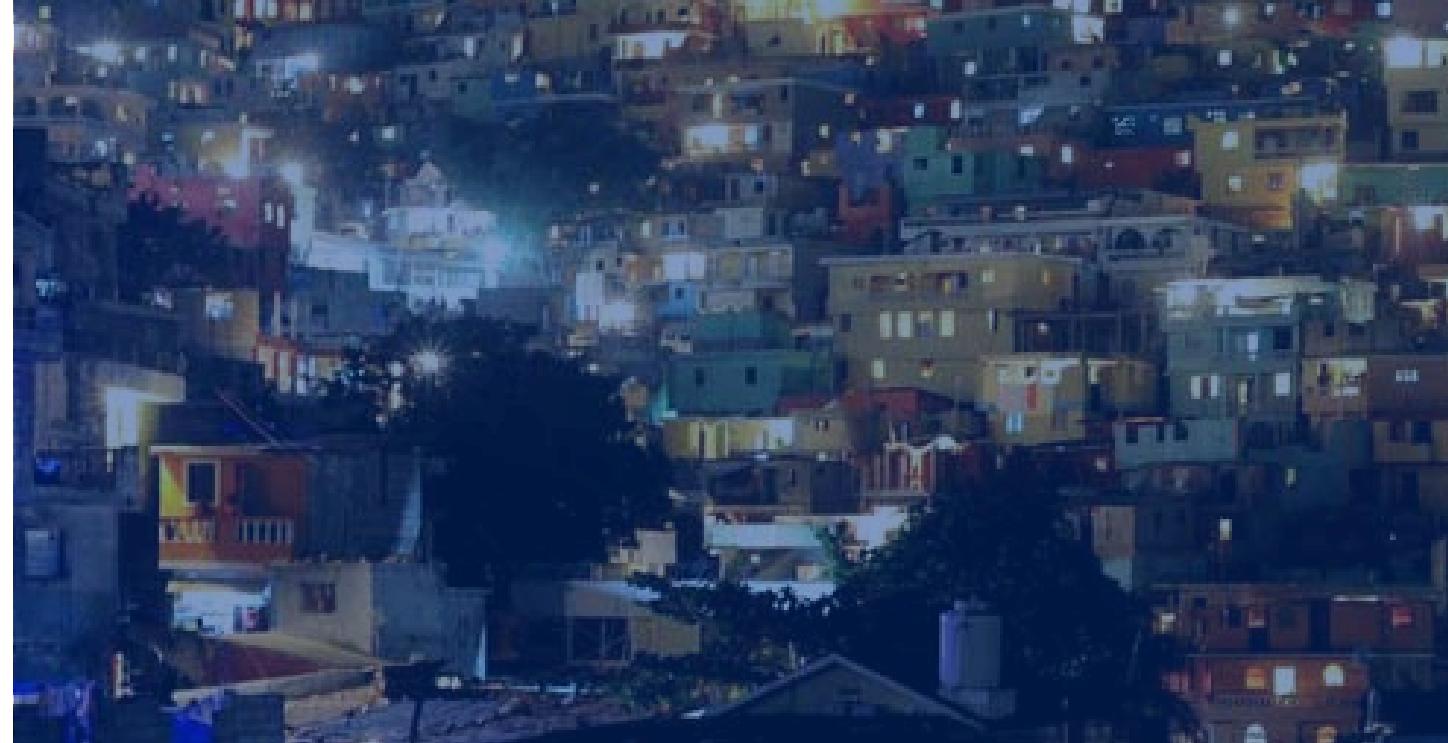
- Moyenne, écart-type, min, max de la puissance.
- Montant total dépensé, fréquence des paiements.
- Métadonnées (type de maison, zone).

2-Prétraitement complet via pipeline automatisé.

3-Division des données :

- 75 % entraînement / 25 % test (avec stratification).
- 2037 échantillons train / 679 test.

# RÉSULTATS CLÉS



Meilleur modèle : xgb\_optimized

Score de performance F1 : 0,9982

Précision équilibrée : 0,9982

Taux d'erreur: 0.2%

## DISTRIBUTION DES PRÉDICTIONS FINALES :

petit : 923 foyers (34,0 %)

moyen : 897 foyers (33,0 %)

grand : 896 foyers (33,0 %)

## CARACTÉRISTIQUES LES PLUS IMPORTANTES :

ratio\_depense\_amperage: 0,0341

avg\_depense\_per\_day: 0,3509

avg\_amperage\_per\_day: 0,6052

# IMPACT DU PROJET

- Pour les entreprises énergétiques : meilleure gestion de la demande.
- Pour les autorités publiques : planification ciblée de l'électrification.
- Pour les citoyens : sensibilisation à la consommation responsable.
- Pour l'environnement : promotion de l'efficacité énergétique et réduction du gaspillage.

# CONCLUSION

Ce projet démontre l'importance de la data science dans le domaine énergétique.

En classifiant les ménages selon leur consommation, il devient possible d'améliorer la distribution de l'énergie, de soutenir les décisions publiques, et de renforcer la durabilité.

## RECOMMANDATIONS:

- Utiliser mon modèle pour segmenter les foyers par niveau de consommation
- Adapter les stratégies tarifaires selon les segments identifiés
- Surveiller les foyers « grand » consommateur pour optimisation énergétique
- Re-entraîner le modèle périodiquement avec de nouvelles données

**MERCI POUR  
VOTRE ATTENTION**

PROJET RÉALISÉ EN BINÔME

**SAINT GERMAIN EMODE ET DARLENS DAMISCA**