

VILLE DE SEATTLE

*Prédiction de la consommation d'énergie
& des émissions de CO2*

*Marion Dedieu
10/2023*

01. PROBLÉMATIQUE & JEU DE DONNÉES

02. FEATURE ENGINEERING

03. APPROCHE DE MODÉLISATION

04. RÉSULTATS

SOMMAIRE

01.

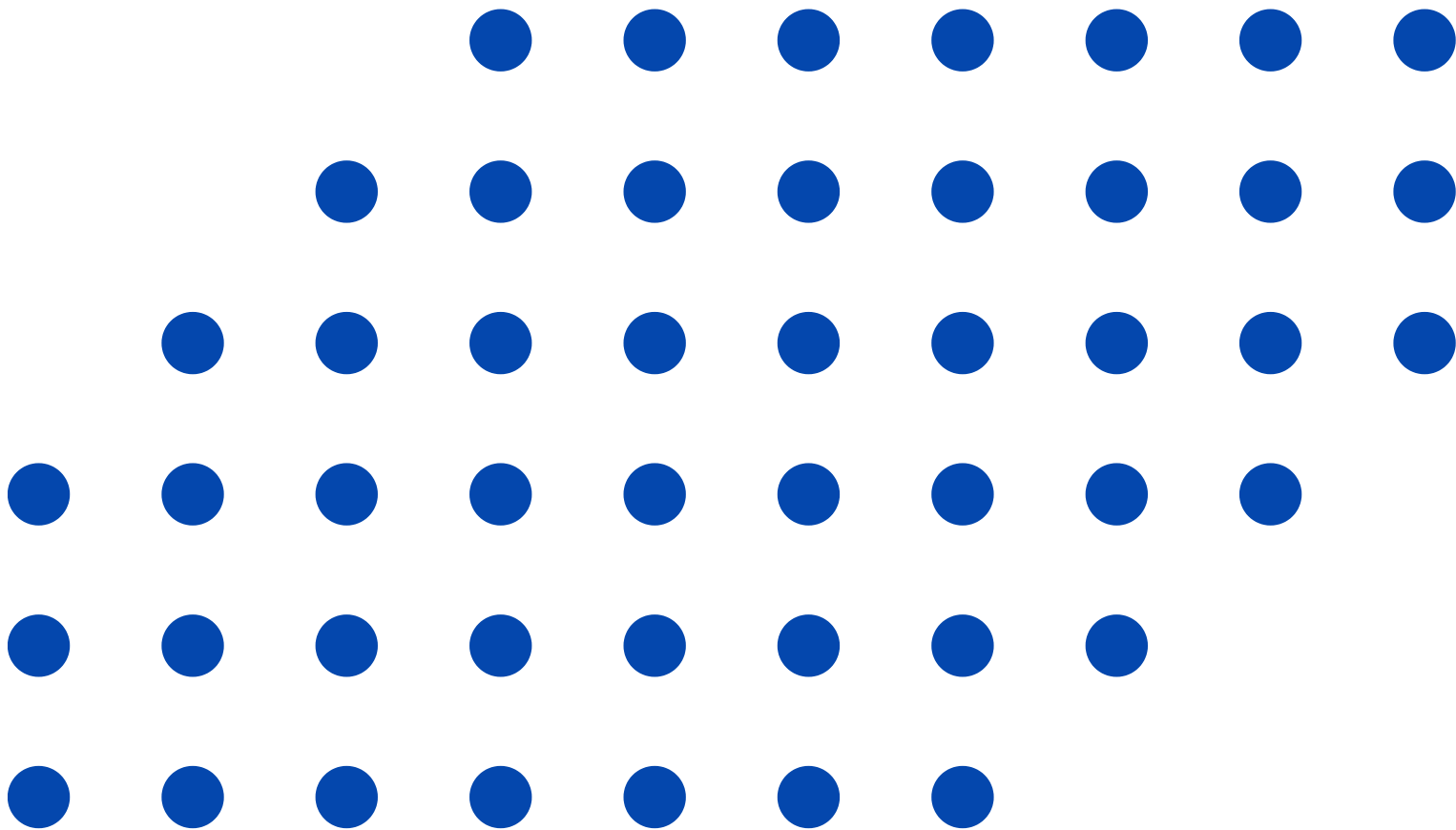
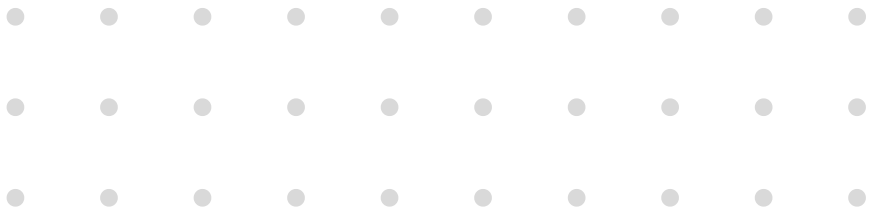
PROBLÉMATIQUE & JEU DE DONNÉES

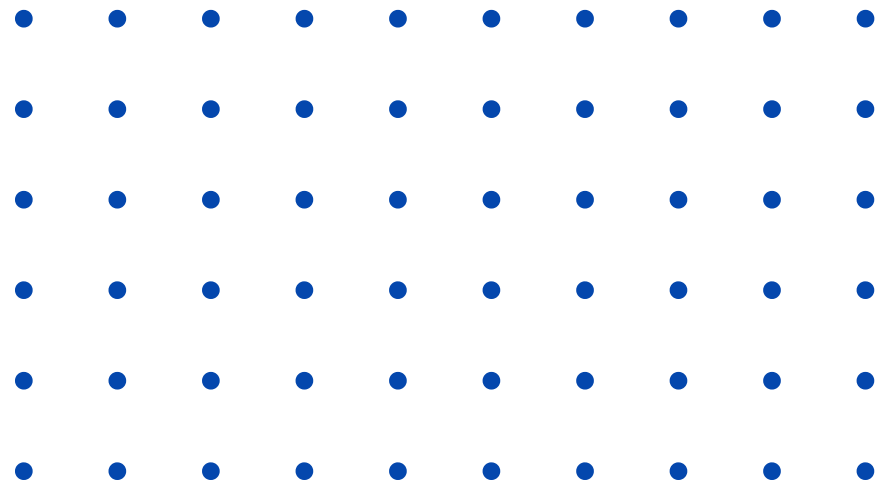
CONTEXTE

Démarche de devenir
neutre en émissions de
carbone d'ici **2050**



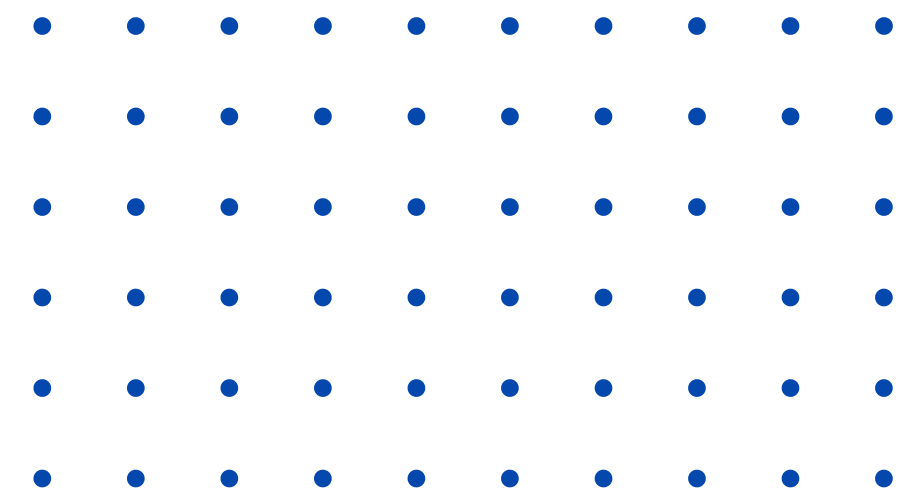
Seattle





PROBLÉMATIQUE

- À partir des relevés faits en 2016, prédire pour les bâtiments non destinés à l'habitation :
 - la **consommation d'énergie**
 - les **émissions de CO2**
- Évaluer l'intérêt de l'**Energy Star score** pour les prédictions



JEU DE DONNÉES

2016 Building Energy Benchmarking

- Caractéristiques énergétiques des bâtiments non résidentiels et multifamiliaux de 20 000 pieds carrés ou plus à Seattle.
- Créé dans le cadre du programme de Benchmarking et de Reporting énergétique des bâtiments de Seattle.

46 colonnes
3 376 lignes

SOURCE

Ville de Seattle

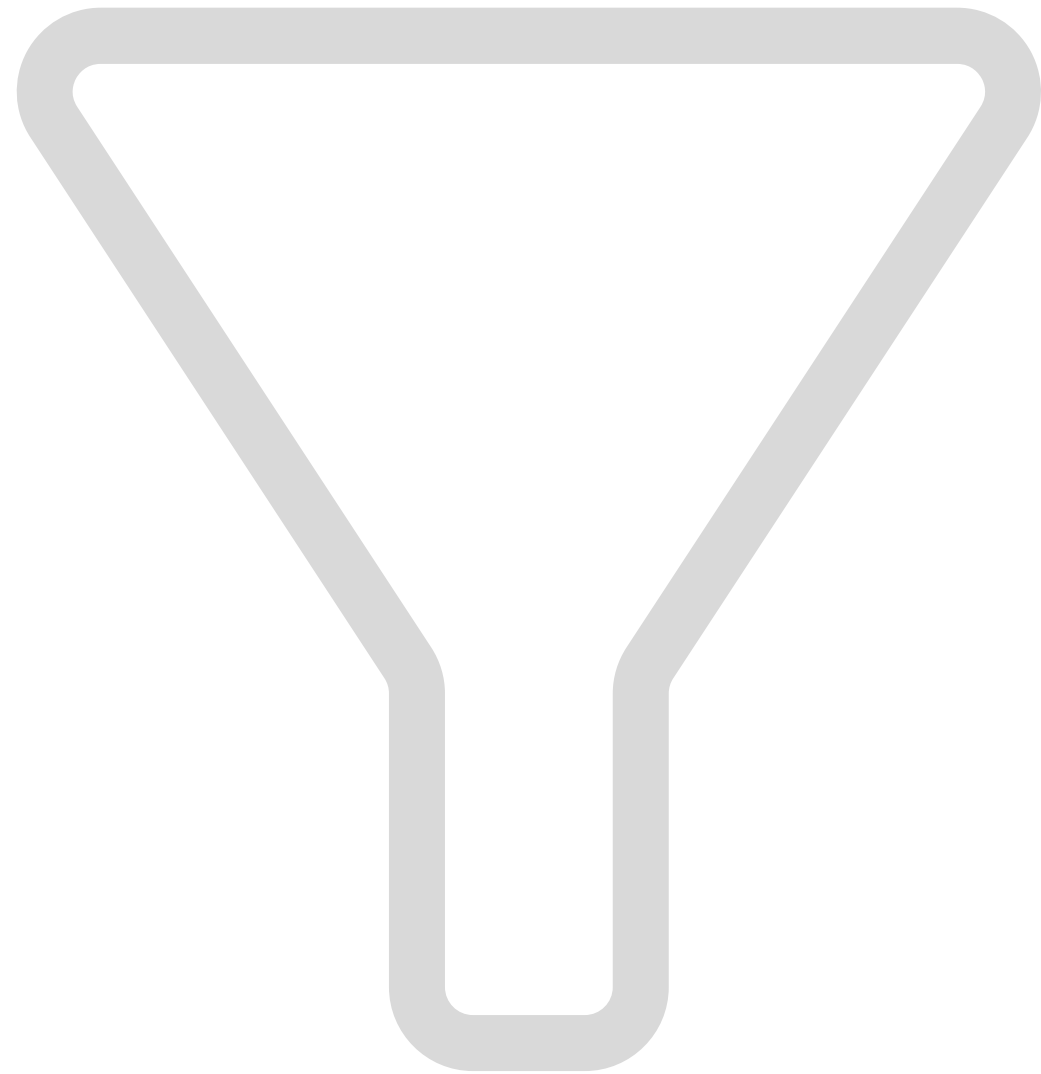
COMPLÉTUDE

87,15 %

02.

FEATURE ENGINEERING

FEATURE ENGINEERING



Définition

Création, sélection et transformation des variables à partir des données brutes pour améliorer les performances des modèles d'apprentissage automatique

Sélection de variables

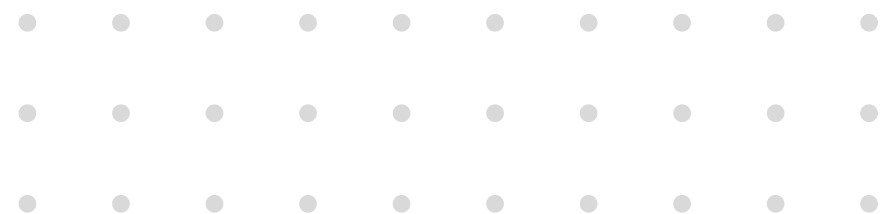
Gestion des valeurs manquantes, doublons, outliers

Encodage de variables

Transformation de variables

SÉLECTION DES FEATURES

- **BuildingType** : Type de bâtiment
- **LargestPropertyUseType** : L'usage le plus important du bâtiment (par exemple, bureaux, magasins de détail) en termes de surface brute de plancher
- **Latitude**
- **Longitude**
- **YearBuilt** : Année de construction du bâtiment ou année de la rénovation complète
- **NumberofBuildings** : Nombre de bâtiments
- **NumberofFloors** : Nombre d'étages
- **PropertyGFATotal** : Surface brute de plancher totale du bâtiment et du parking
- **SteamUse(kBtu)** : Quantité annuelle de vapeur urbaine consommée par le bien
- **Electricity(kBtu)** : Quantité annuelle d'électricité consommée par le bien, mesurée en milliers d'unités thermiques britanniques (kBtu)
- **NaturalGas(kBtu)** : Quantité annuelle de gaz naturel, mesurée en kBtu
- **Outlier** : Si un bien contient des valeurs aberrantes élevées ou faibles (Yes or No)



SÉLECTION DES TARGETS

- **SiteEUIWN(kBtu/sf)** : Consommation d'énergie annuelle du bâtiment divisée par sa surface brute de plancher, normalisée en fonction des conditions météorologiques, c'est-à-dire la consommation d'énergie du site que le bien aurait consommée dans des conditions météorologiques moyennes sur 30 ans.
- **GHGEmissionsIntensity** : Total des émissions de gaz à effet de serre divisé par la surface brute de plancher du bien

ENERGY STAR SCORE

- **Objectif** : évaluer l'efficacité énergétique des bâtiments commerciaux et industriels.
- **Notation** : 1 à 100.
- **Facteurs** pris en compte : taille du bâtiment, le type d'activité, les heures d'exploitation, etc.
- **Interprétation** : Score élevé = une meilleure efficacité énergétique.



TRAITEMENTS

Valeurs manquantes

Suppression valeurs nulles
(sauf Energy Star Score)

Doublons

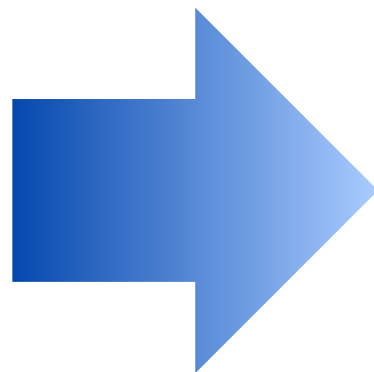
Pas de doublons

Outliers

Suppression des 17 lignes
taguées "outlier"

Suppression des 52 lignes où le
nombre bâtiments est à 0

Suppression des minimums
négatifs



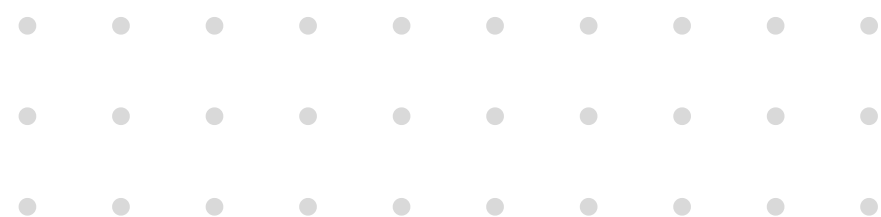
14 colonnes
1 591 lignes

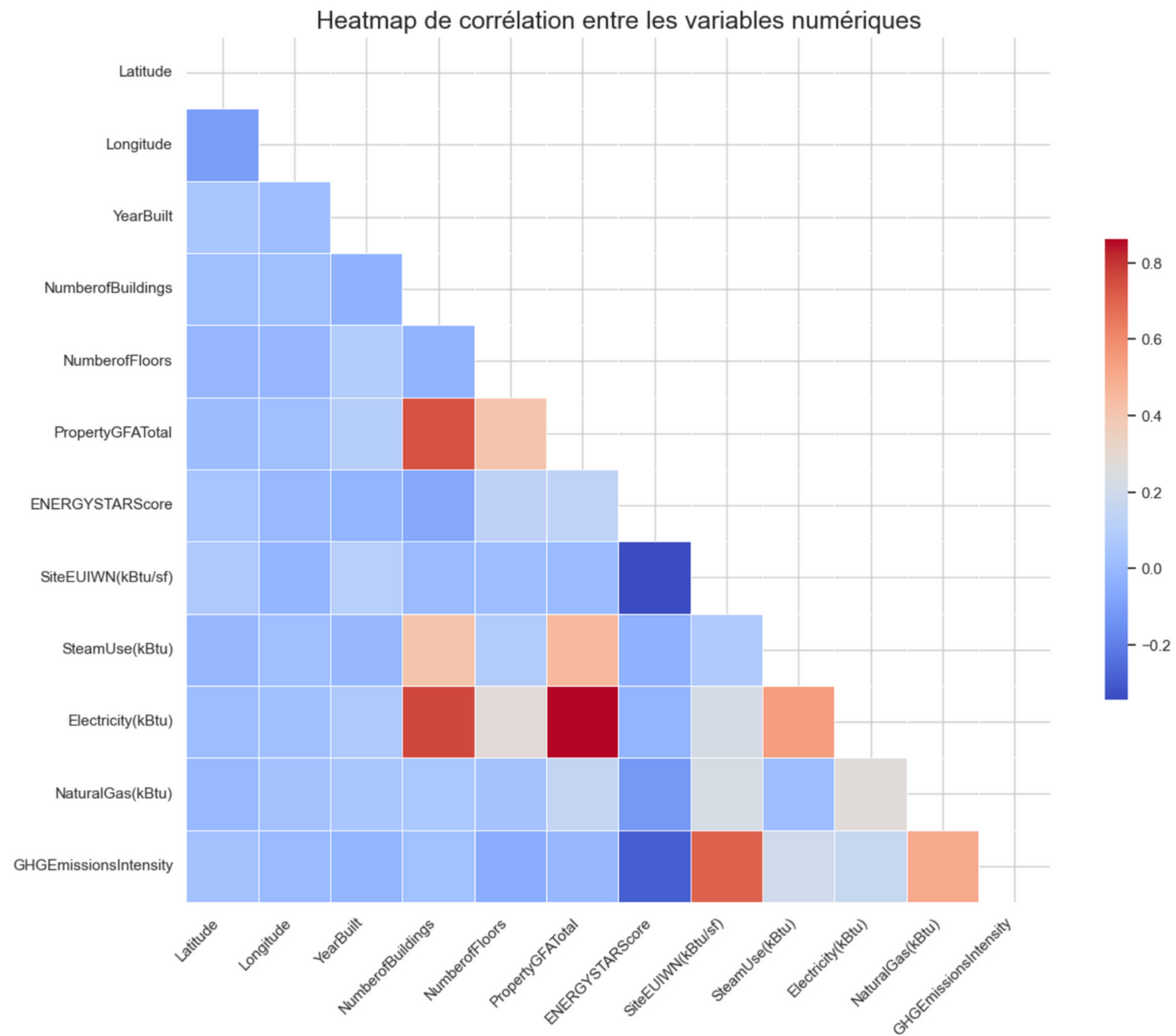
ENCODAGE : ONE HOT ENCODER

- **Objectif** : Créer des variables indicatrices à partir de variables catégorielles.
- **Comment ça marche** : Pour chaque catégorie unique, une nouvelle colonne est créée, et la colonne correspondante prend la valeur 1 si la catégorie est présente et 0 sinon.
- **Utilisation** : C'est généralement utilisé lorsque l'ordre des catégories n'a pas d'importance, et chaque catégorie doit être traitée de manière équivalente.
≠ LabelEncoder (variables ordinales)

TRANSFORMATIONS

- **Normalisation** : StandardScaler : Ajuste la moyenne des données à 0 et l'écart-type à 1
 - Résultats améliorés nettement pour Gradient Boosting et Random Forest
- **Passage au logarithme** : Réduit l'effet des valeurs extrêmes et rend la distribution des données plus proche d'une forme normale
 - Résultats améliorés pour Ridge, Lasso, ElasticNet, KNN



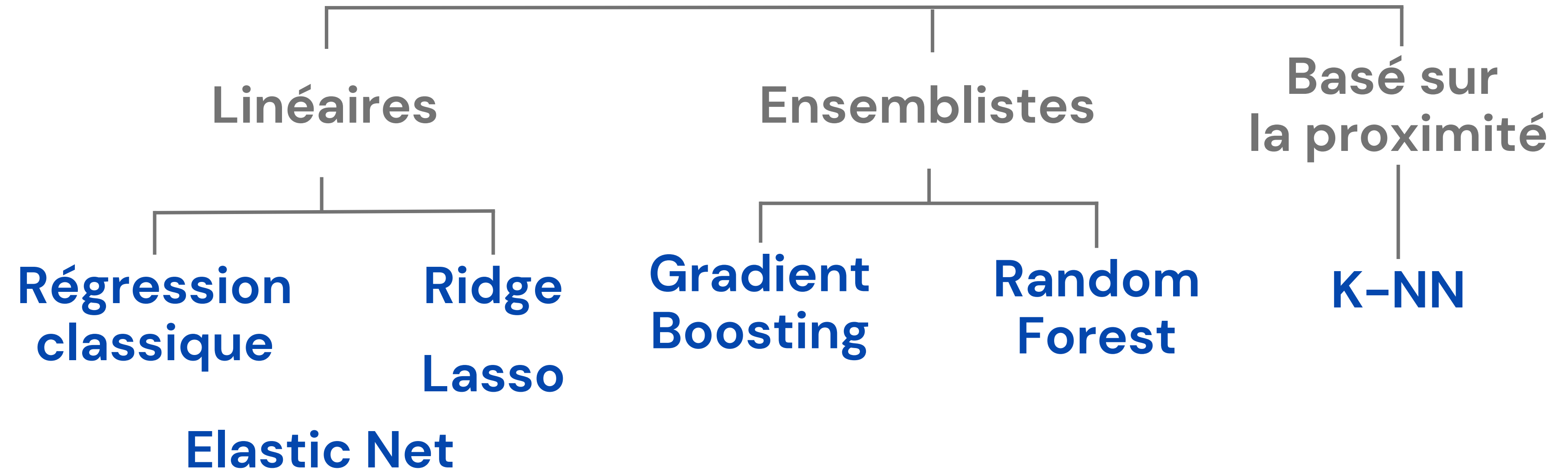


Vérification du
data leakage

03.

APPROCHE DE MODÉLISATION

MODÈLES



GridSearchCV sur chaque modèle

MESURES DE PERFORMANCE

- **MSE (Mean Squared Error)**
 - Mesure de la moyenne des carrés des erreurs
 - Donne plus de poids aux erreurs importantes
 - Meilleur = le score le plus bas
- **RMSE (Root Mean Squared Error)**
 - Racine carrée de la MSE pour des unités de mesure cohérentes
 - Interprétable dans les mêmes unités que la variable cible
 - Meilleur = le score le plus bas
- **MAE (Mean Absolute Error)**
 - Mesure la moyenne des valeurs absolues des erreurs
 - Traite toutes les erreurs de manière égale, sans distinction d'importance (donc robuste aux valeurs aberrantes)
 - Meilleur = le score le plus bas
- **R^2 (Coefficient de Détermination)**
 - Évalue la proportion de la variance expliquée par le modèle
 - Mesure la qualité globale de l'ajustement du modèle
 - Varie de 0 (aucune explication) à 1 (explication parfaite)

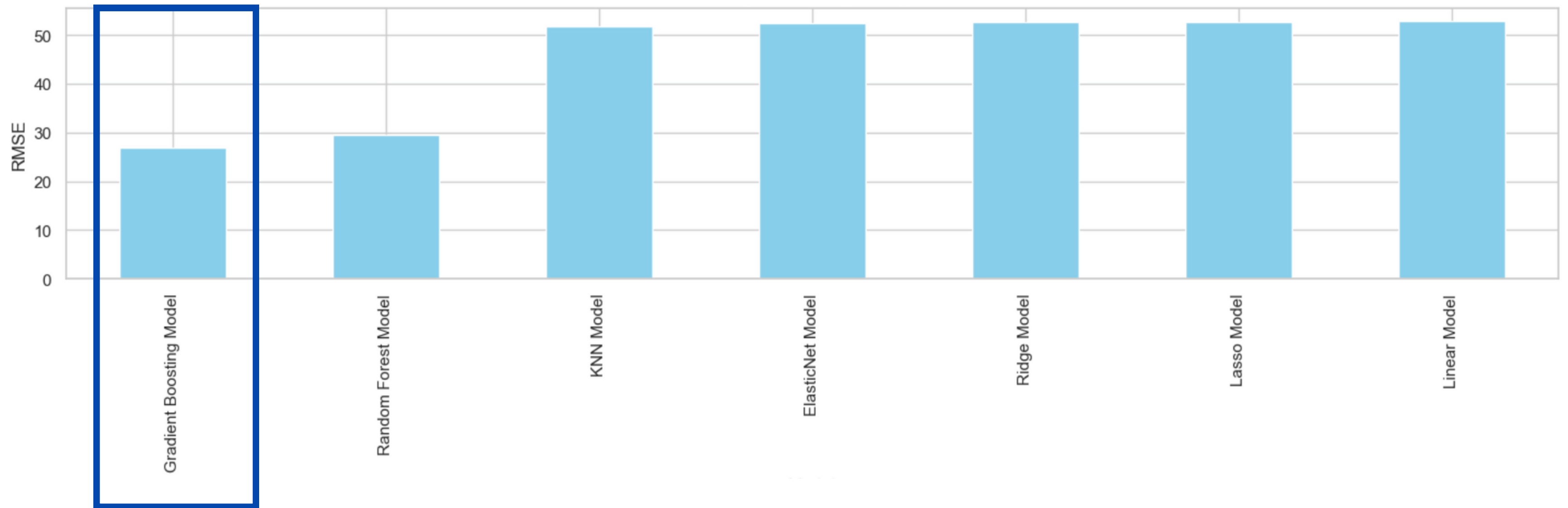
04.

RÉSULTATS OBTENUS

COMPARAISONS DES MODÈLES

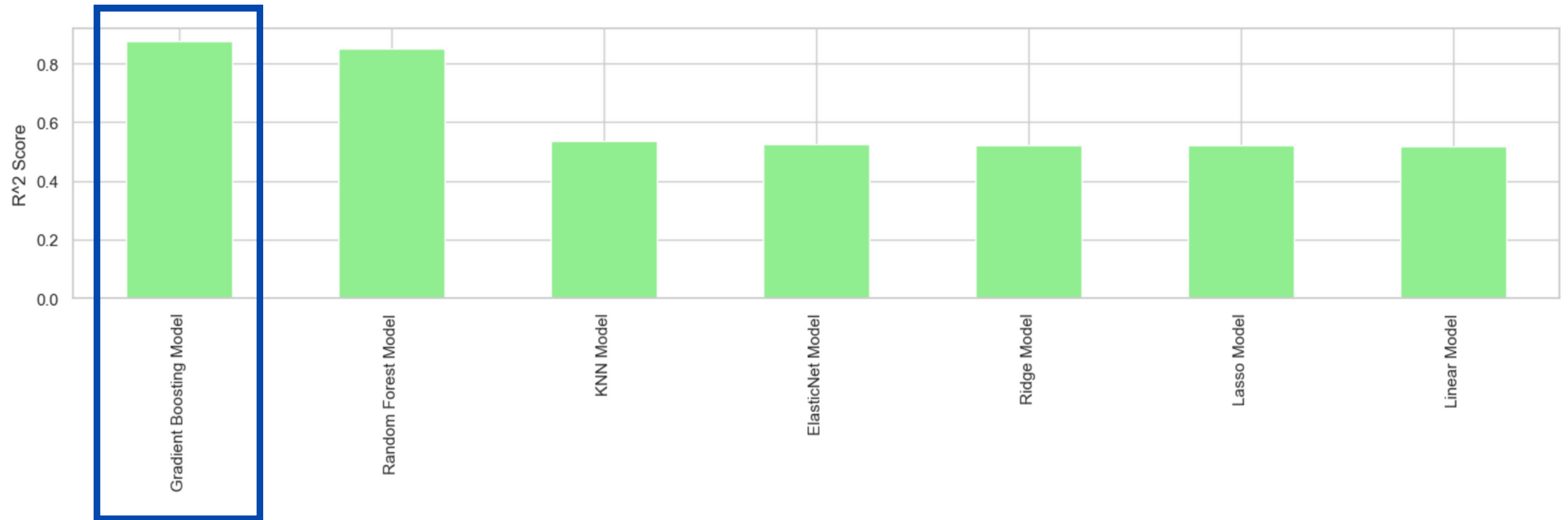
CONSOMMATION D'ÉNERGIE

RMSE



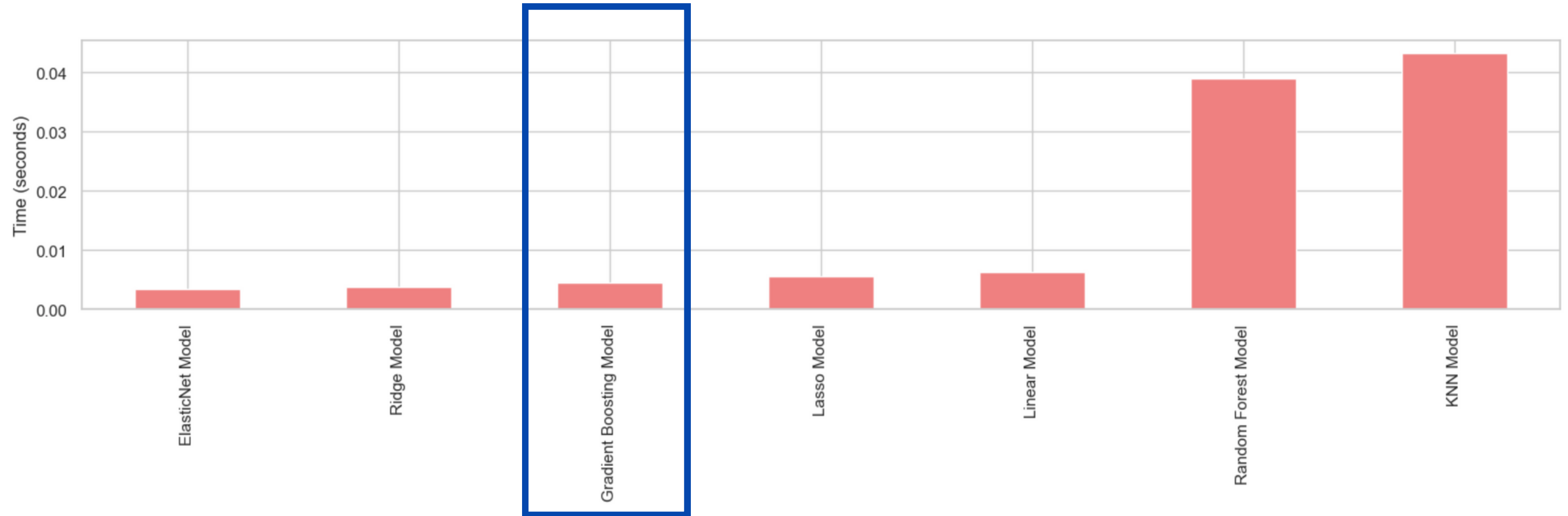
CONSOMMATION D'ÉNERGIE

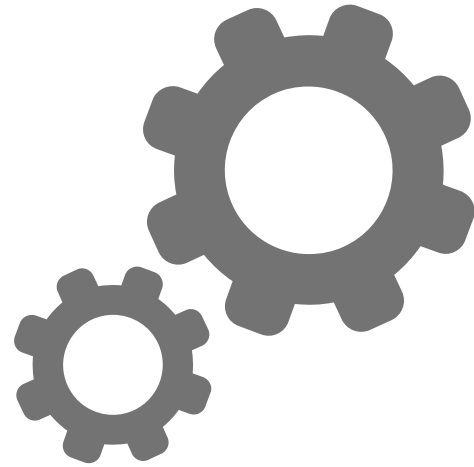
R²



CONSOMMATION D'ÉNERGIE

TEMPS D'ÉVALUATION



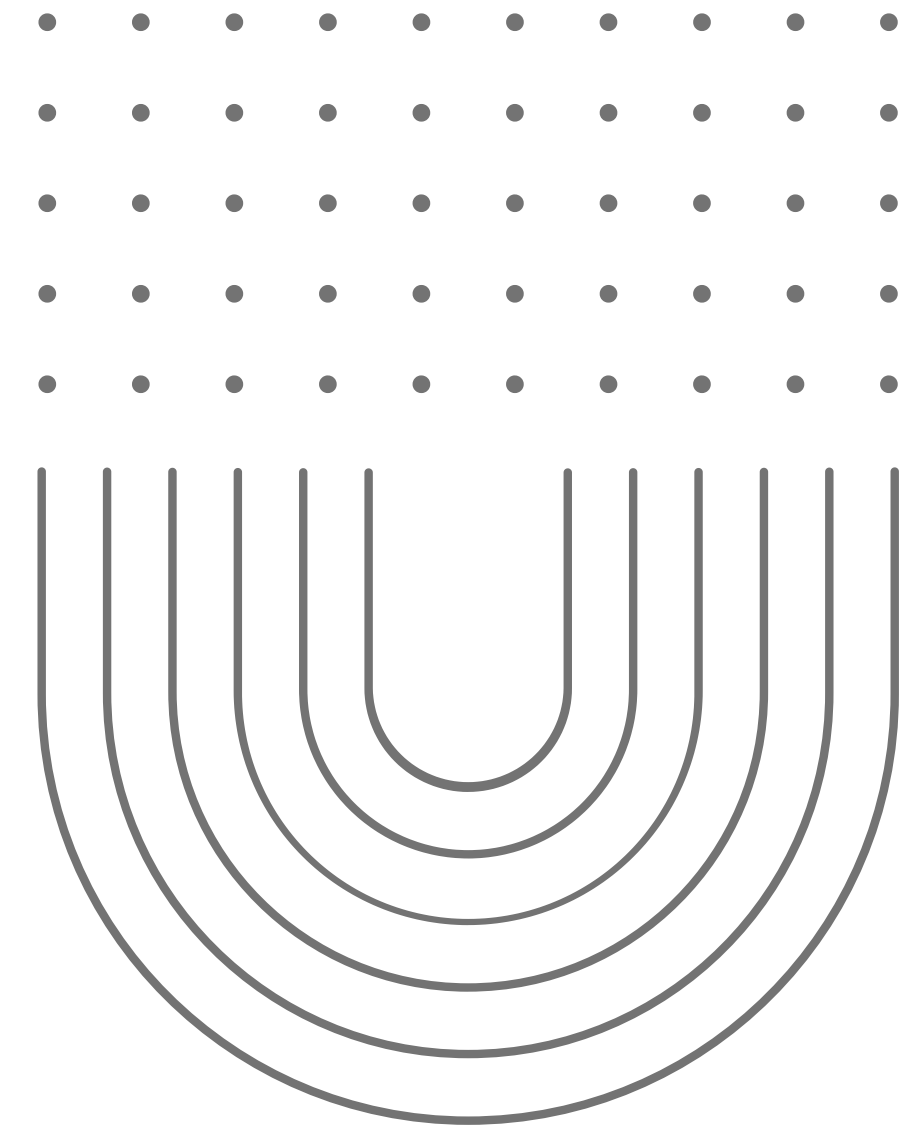
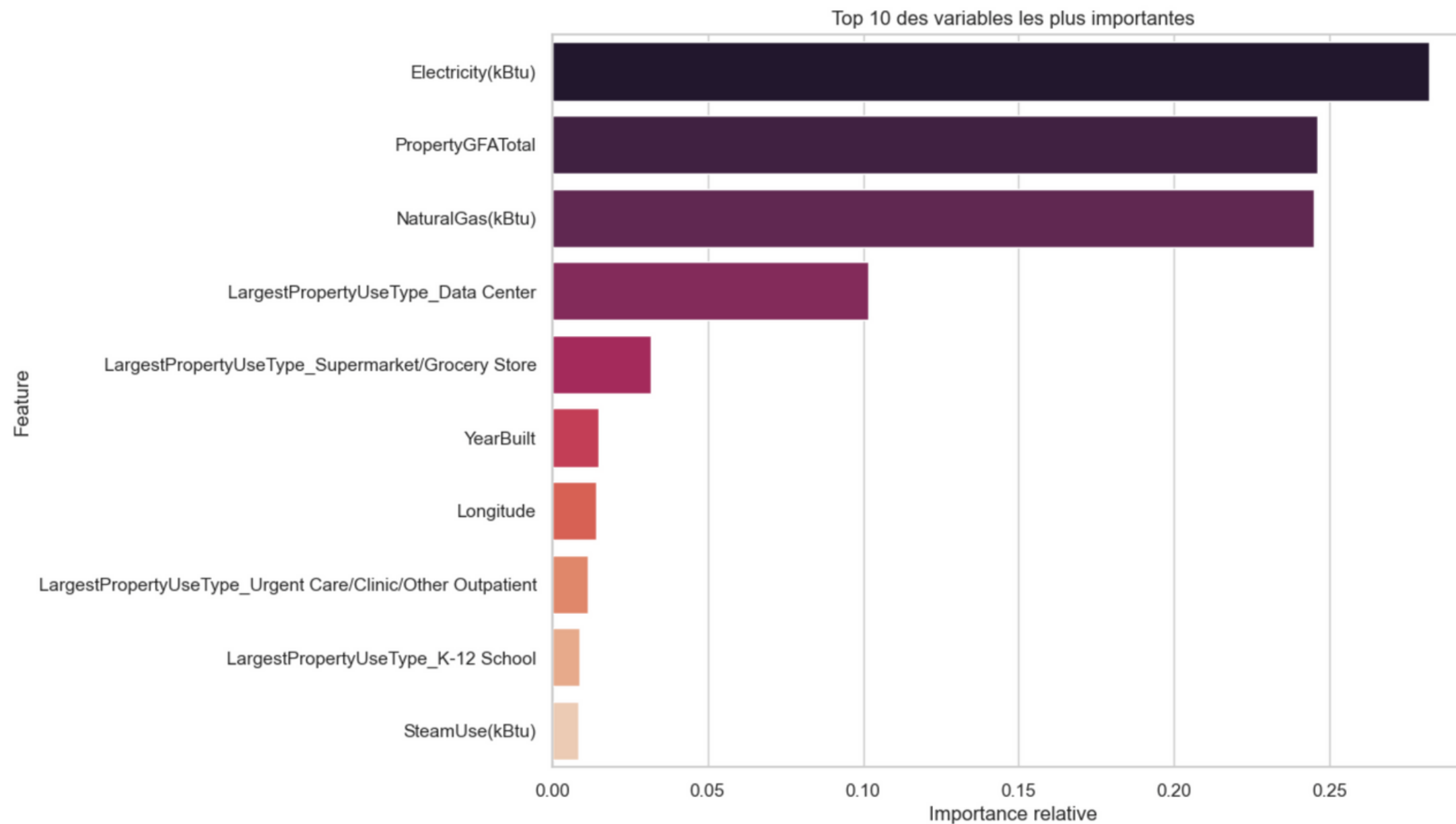


Hyperparamètres

learning_rate : 0.1

max_depth : 3

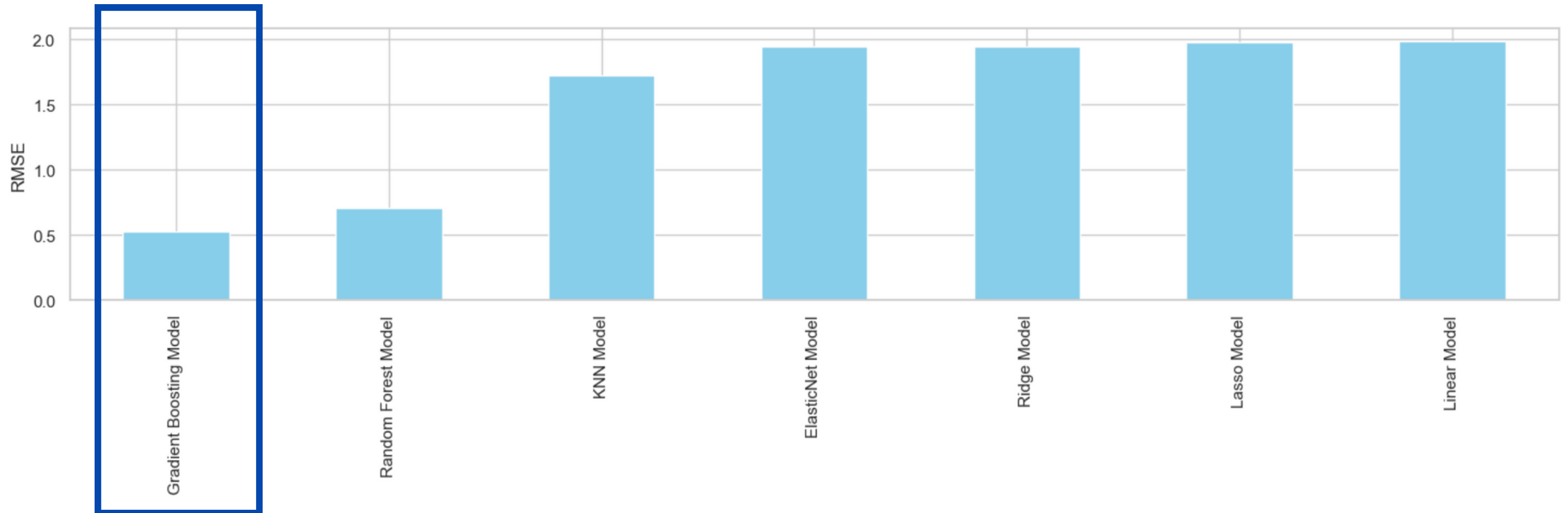
n_estimators : 150



GRADIENT BOOSTING

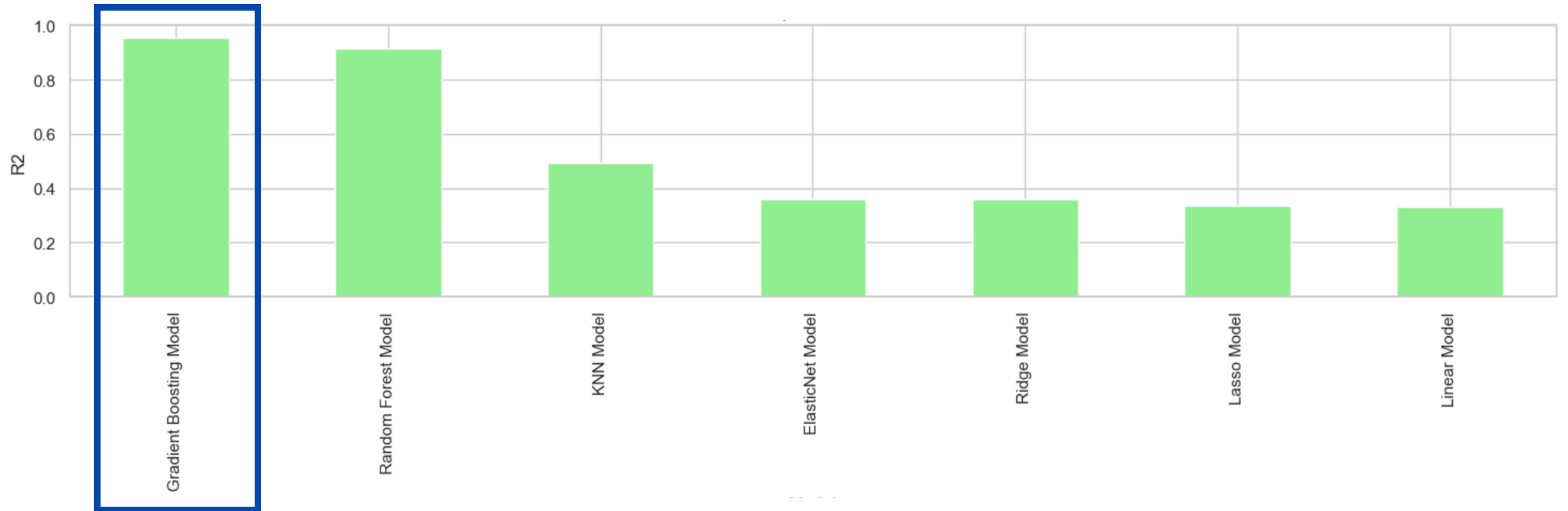
ÉMISSIONS DE CO2

RMSE



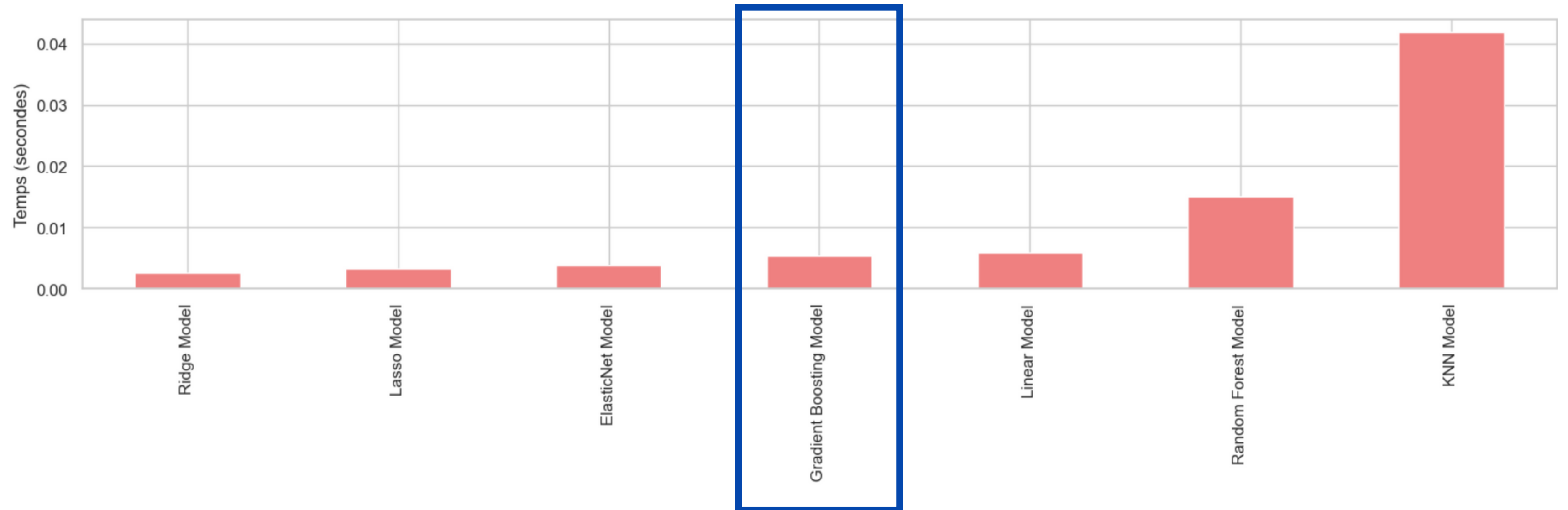
ÉMISSIONS DE CO2

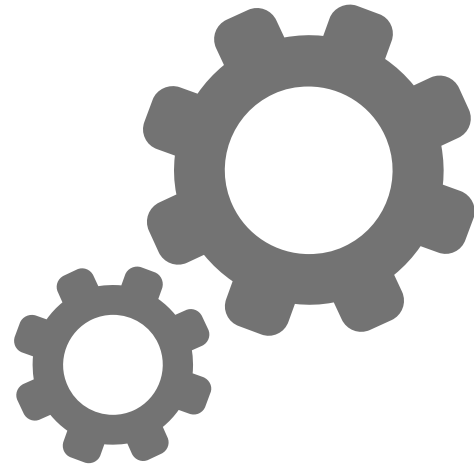
R2



ÉMISSIONS DE CO2

TEMPS D'ÉVALUATION



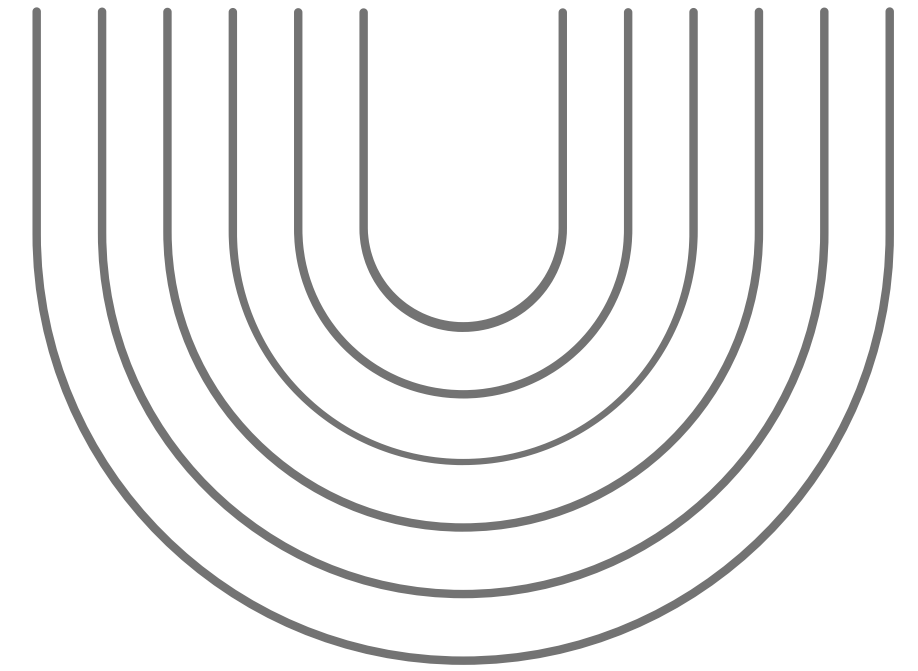
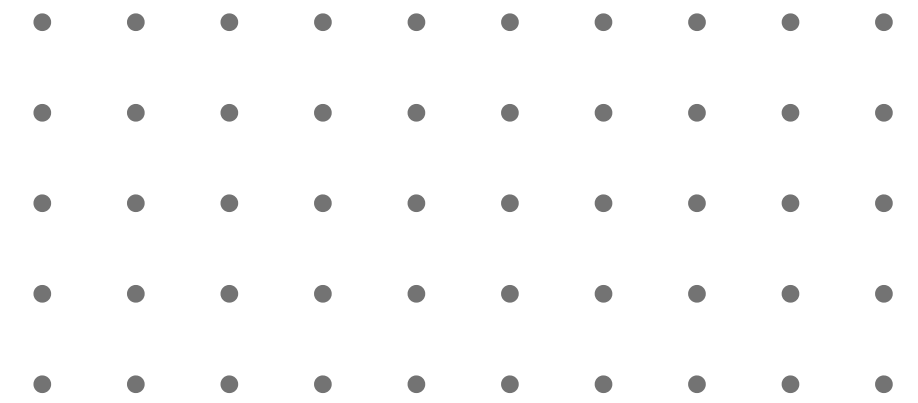
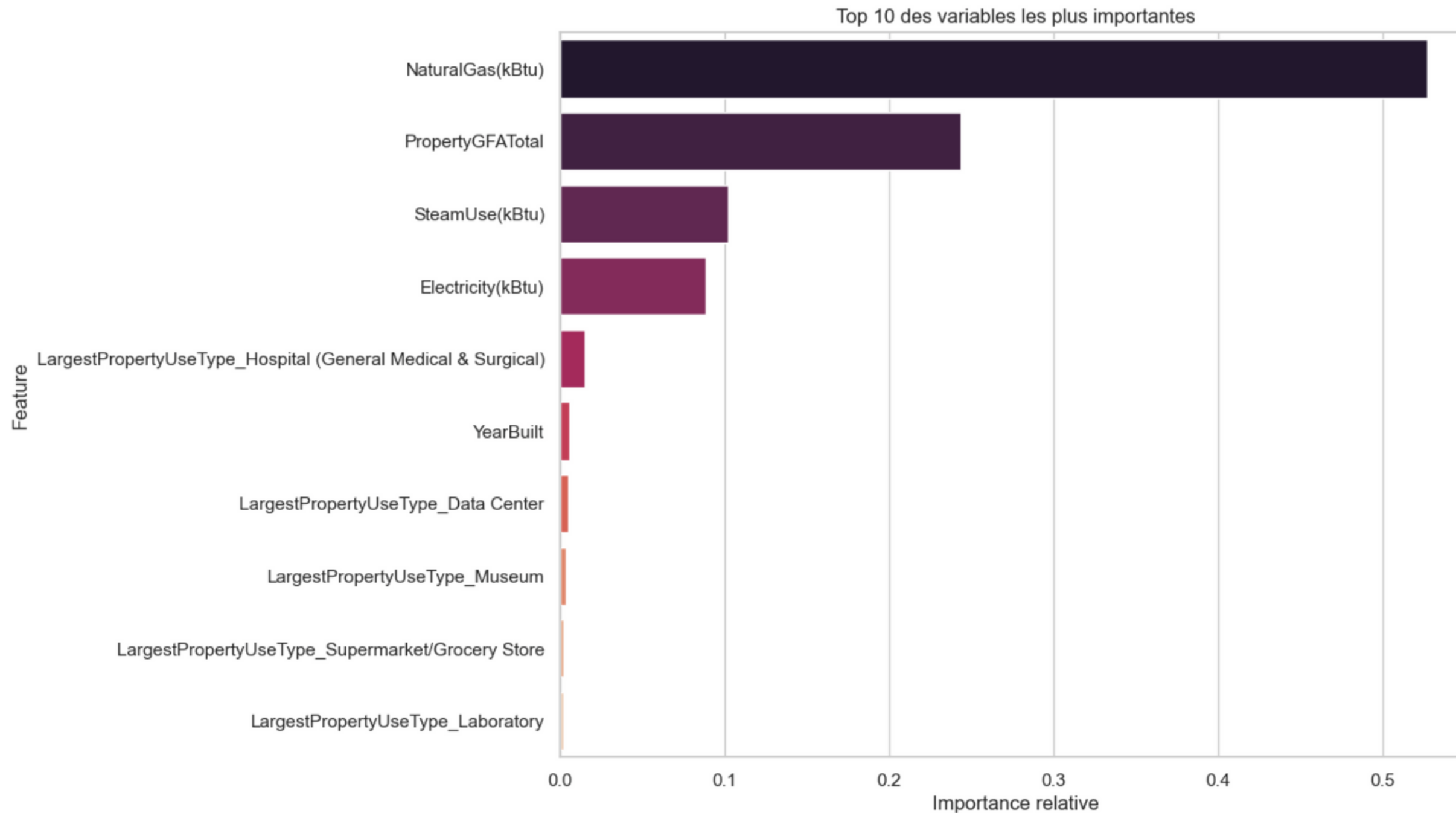


Hyperparamètres

learning_rate : 0.2

max_depth : 3

n_estimators : 150



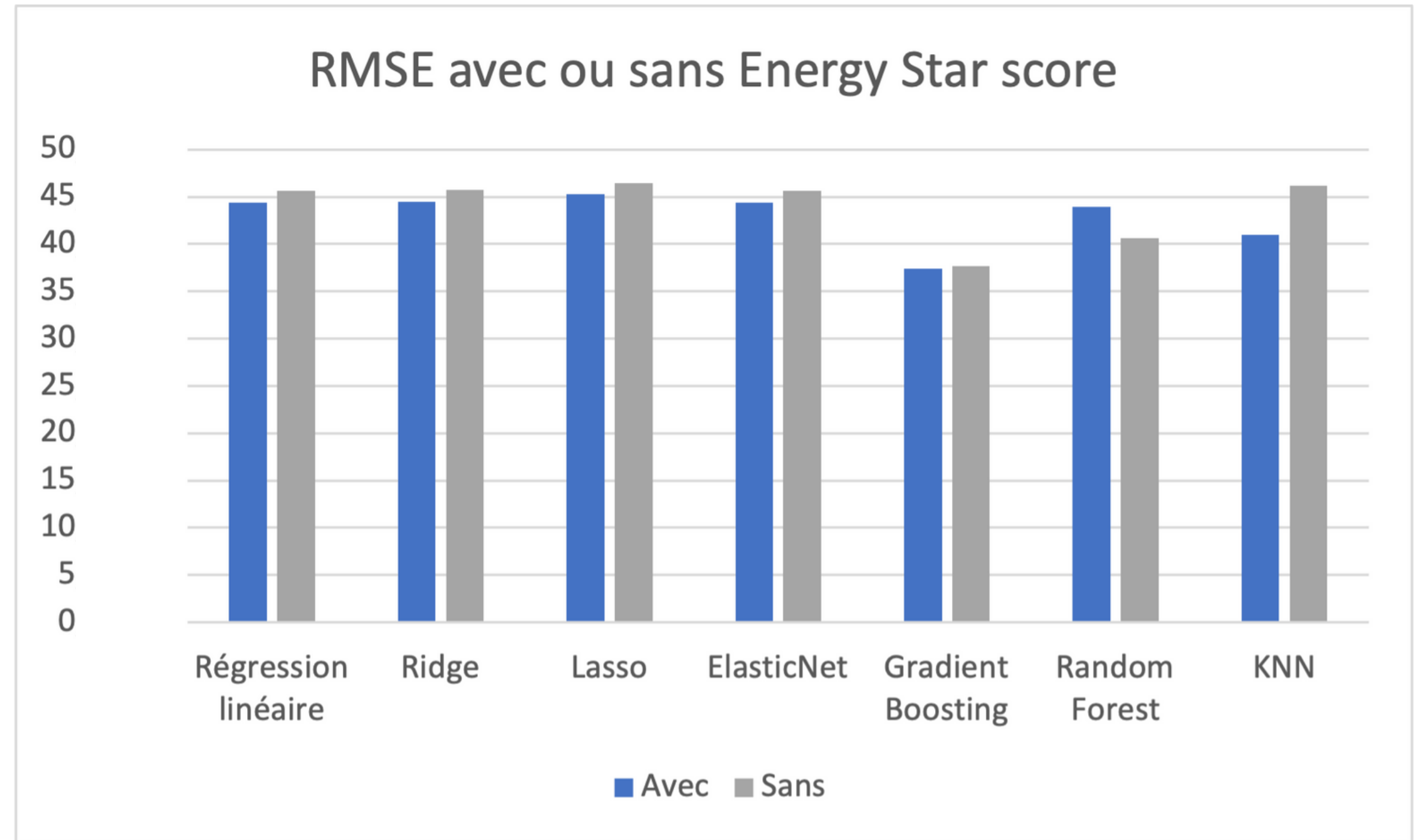
GRADIENT BOOSTING

INTÉRÊT DE L'ENERGY STAR SCORE

CONSOMMATION D'ÉNERGIE

RSME : - 2 %

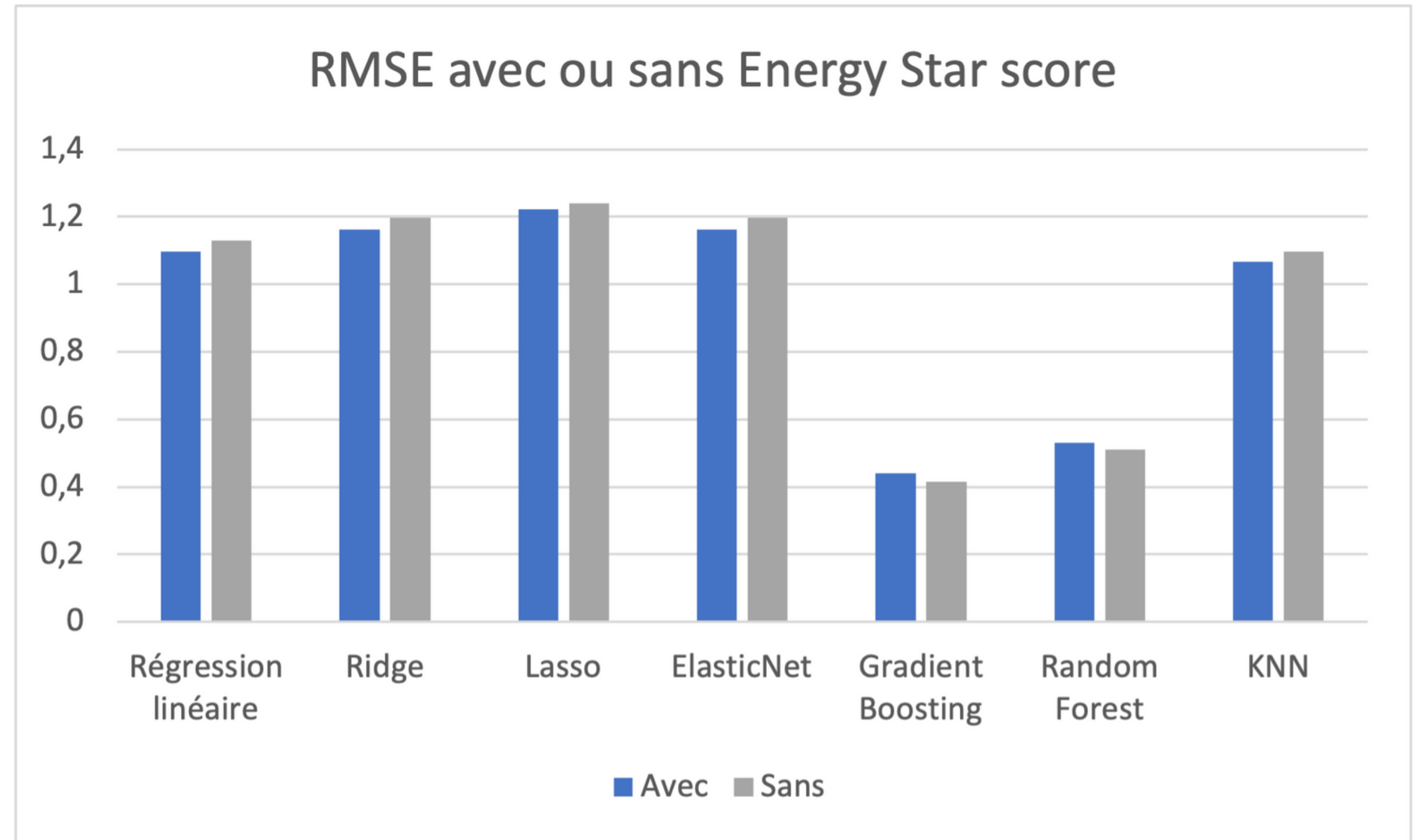
R2 : + 4 %



ÉMISSIONS DE CO2

RSME : - 0,50 %

R2 : + 5 %



05.

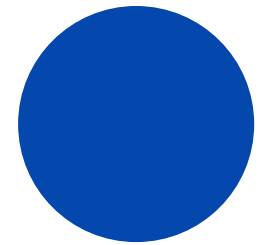
CONCLUSION



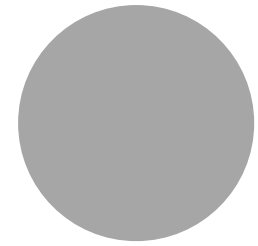
**CHOIX DU MODÈLE
GRADIENT BOOSTING**



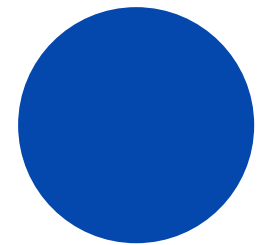
**ENERGY STAR SCORE
PERTINENT**



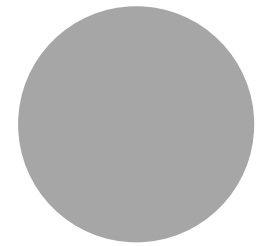
CHOIX DES FEATURES



TRANSFORMATIONS APPLIQUÉES NON
INDIVIDUELLEMENT AUX MODÈLES



**CHOIX DES ALGORITHMES
DE MACHINE LEARNING**



PEU DE DONNÉES POUR TESTER
L'INTÉRÊT DE L'ENERGY STAR SCORE

**NUANCES
SUR L'ANALYSE**

MERCI !

Des questions ?