

# Projet 4

**Anticipez les besoins en consommation de bâtiments**

# Aider la ville de Seattle à atteindre l'objectif ville neutre

**Prédire les émissions et consommations des bâtiments non résidentiels**

- 1 - Sélection et nettoyages des variables
- 2 - Features engineering et modèles
- 3 - Approche méthodologique : optimisation et sélection des modèles
- 4 - Résultats
- 5 - Etude de l'ENERGYSTARScore
- 6 - Conclusion

# 2- Sélection et nettoyage des variables

Relevés des bâtiments effectués en 2016.

## Données Structurelles

- Nombre de bâtiments
- Nombre d'étages
- Année de construction
- Type d'utilisation
- Surfaces

## Données de localisation

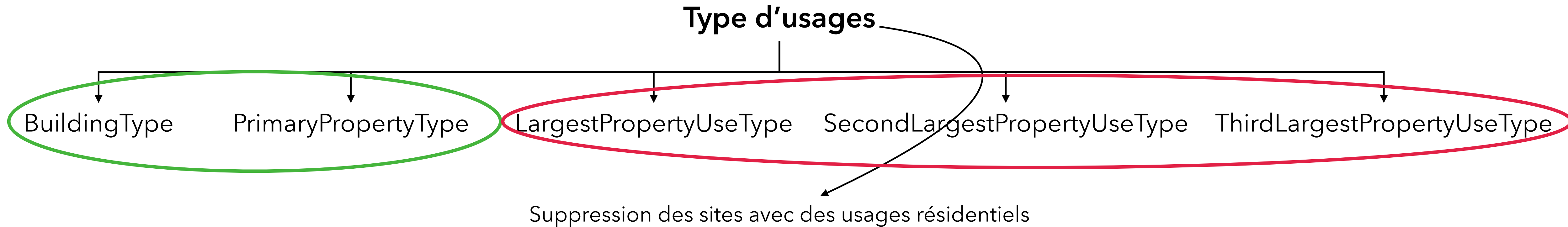
- Adresse
- ZipCode
- Quartier
- Coordonnées Lat./Long.
- Numéro de District
- ....

## Données énergétiques

- Consommation électrique
- Consommation de gaz
- Consommation de vapeur
- Consommation énergétique totale
- Emissions de CO2
- ....

# 2- Sélection et nettoyage des variables

## Données Structurales



### Nombre de bâtiments

Suppression des sites  
> 40 bâtiments

Surface totale bâtiment  
Nbre étages

Complétion des  
valeurs manquantes

### Nombre d'étages

Suppression des sites  
avec plus de 77 étages

### Année de construction

# 2- Sélection et nettoyage des variables

Données de localisation

Latitude / Longitude

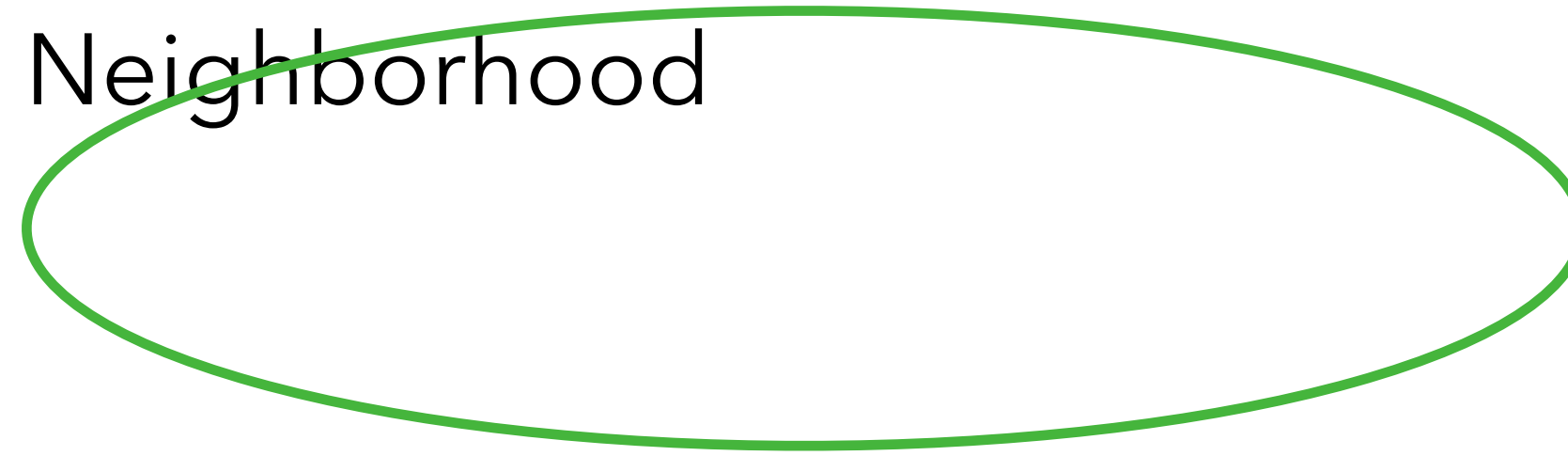
ZipCode

Neighborhood

Ville

CouncilDistrictCode

Adresse



# 2- Sélection et nettoyage des variables

## Données énergétiques

TotalGHGEmissions

Electricity

NaturalGas

SiteEnergyUse (kBtu)

SteamUse

Plusieurs unités possibles

→ Relation entre les unités

→ Choix de l'unité commune  
à la cible (kBtu)

Valeurs manquantes

→ Suppression des sites sans valeurs

→ SiteEnergyUse manquant :  
 $SEU = E + G + S$

Suppression des sites avec  
une incohérence énergétique

$95\% < SEU = E + G + S < 105\%$

Création de variables : ratio d'énergie

Electricity\_ratio

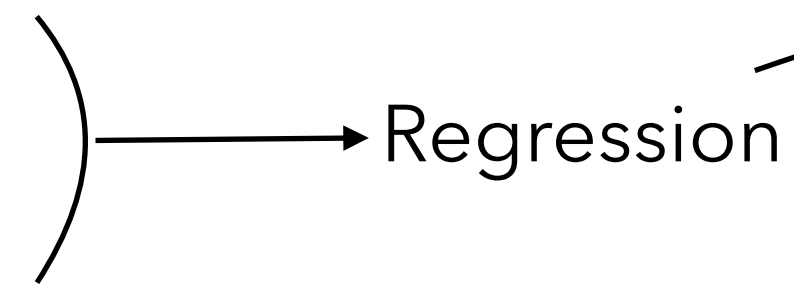
NaturalGas\_ratio

SteamUse\_ratio

# 3 - Features engineering et modèles

Identification des variables cibles :

- TotalGHGEmissions
- SiteEnergyUse



Choix de modèles :

- LinearRegression
- SGDRegressor
- Ridge
- Lasso
- ElasticNet
- SVR
- KNNRegressor
- RandomForestRegressor
- XGBoostRegressor

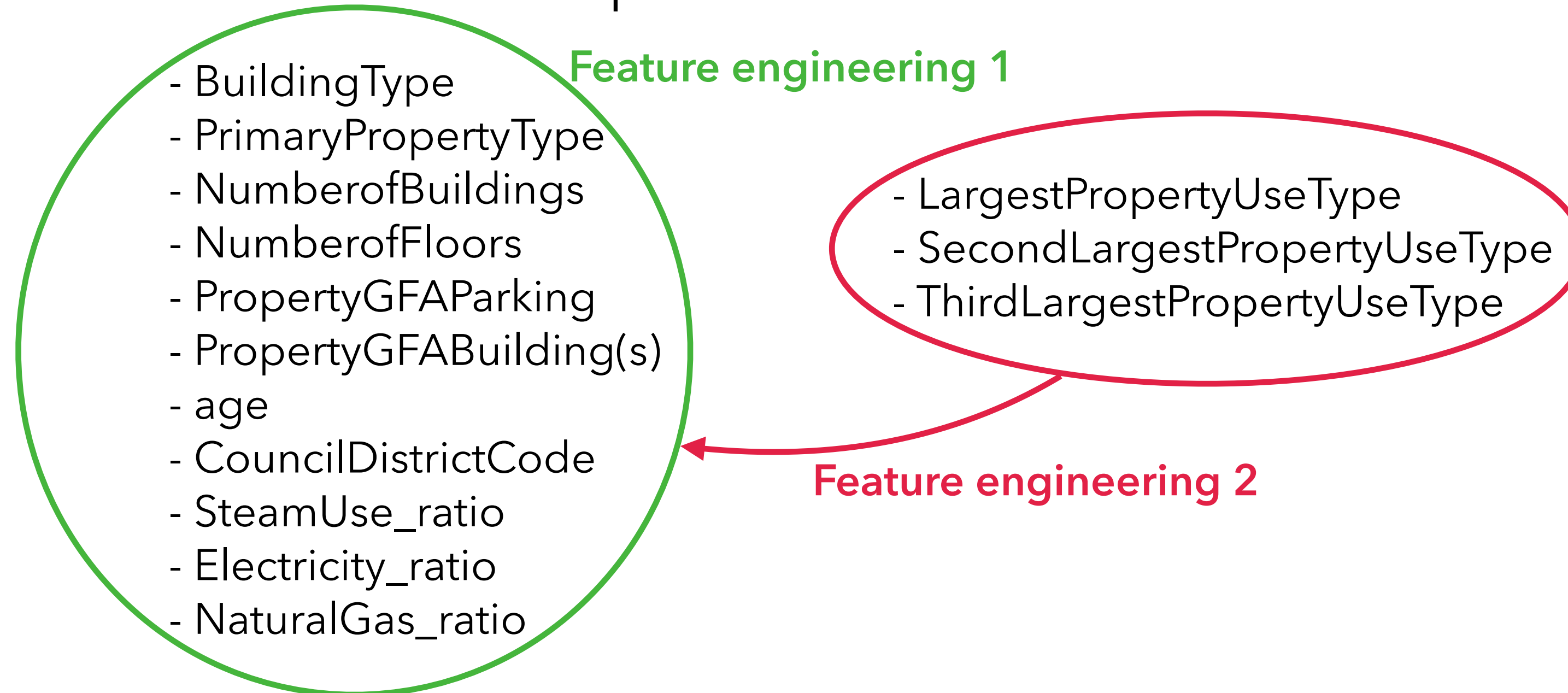
Identification des variables prédictives :

- BuildingType
- PrimaryPropertyType
- NumberofBuildings
- NumberofFloors
- PropertyGFAParking
- PropertyGFABuilding(s)
- age
- CouncilDistrictCode
- SteamUse\_ratio
- Electricity\_ratio
- NaturalGas\_ratio

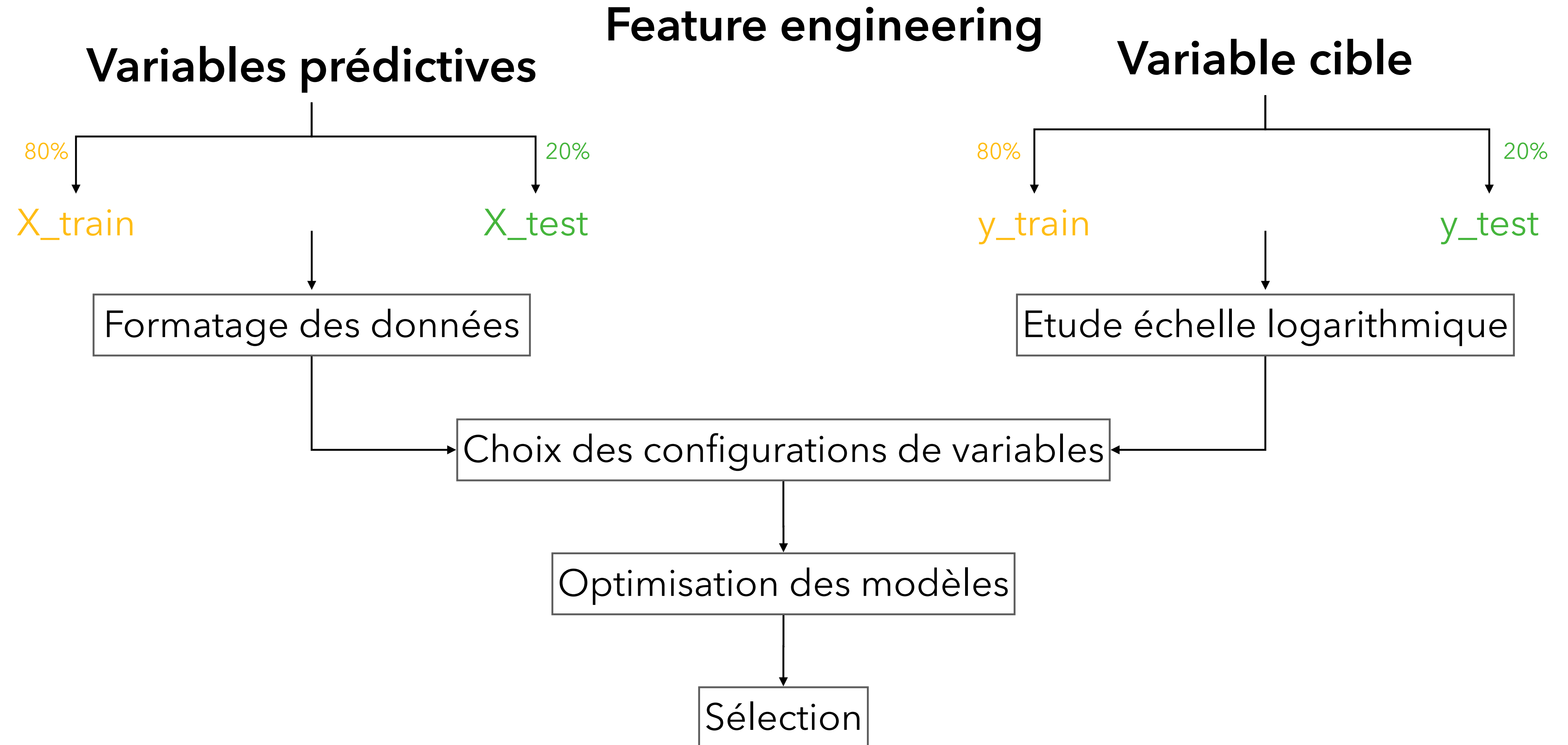
Feature engineering 1

- LargestPropertyUseType
- SecondLargestPropertyUseType
- ThirdLargestPropertyUseType

Feature engineering 2



# 4 - Approche méthodologique : optimisation et sélection des modèles

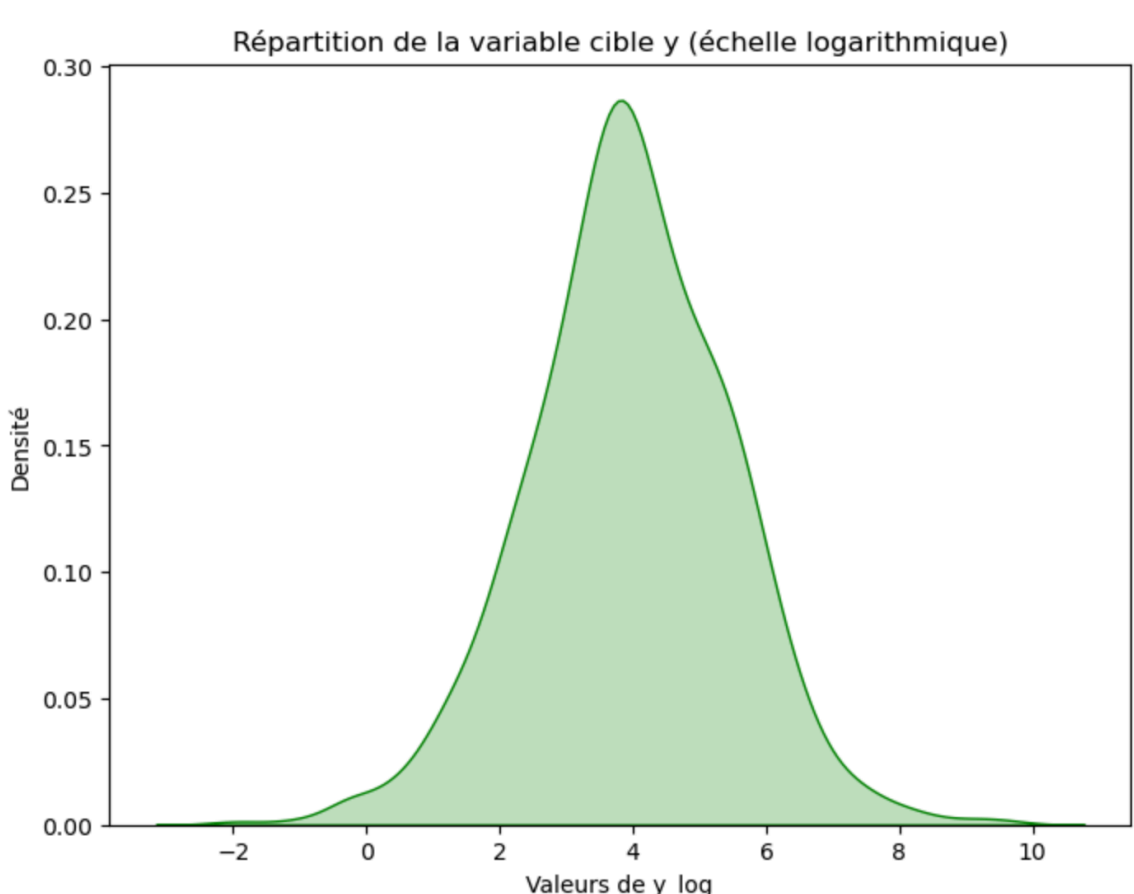
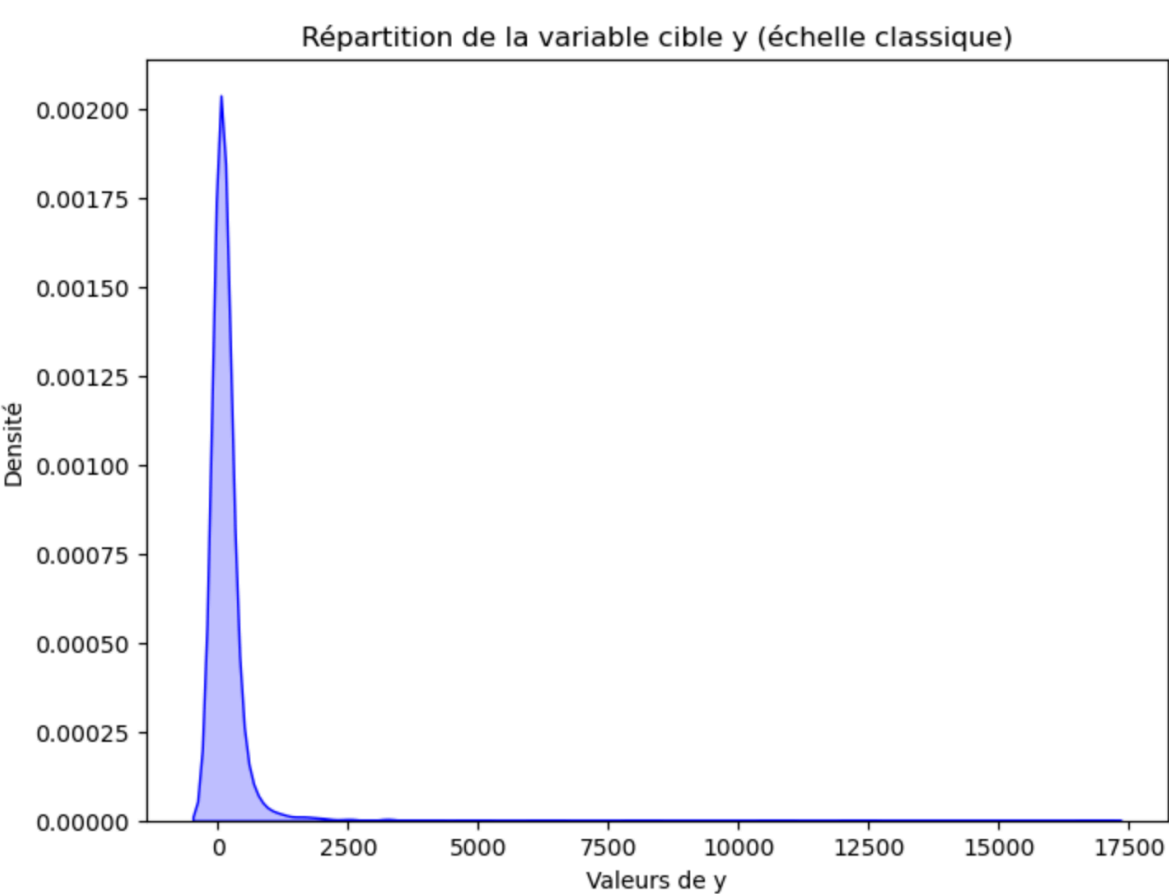




# 4 - Approche méthodologique : optimisation et sélection des modèles

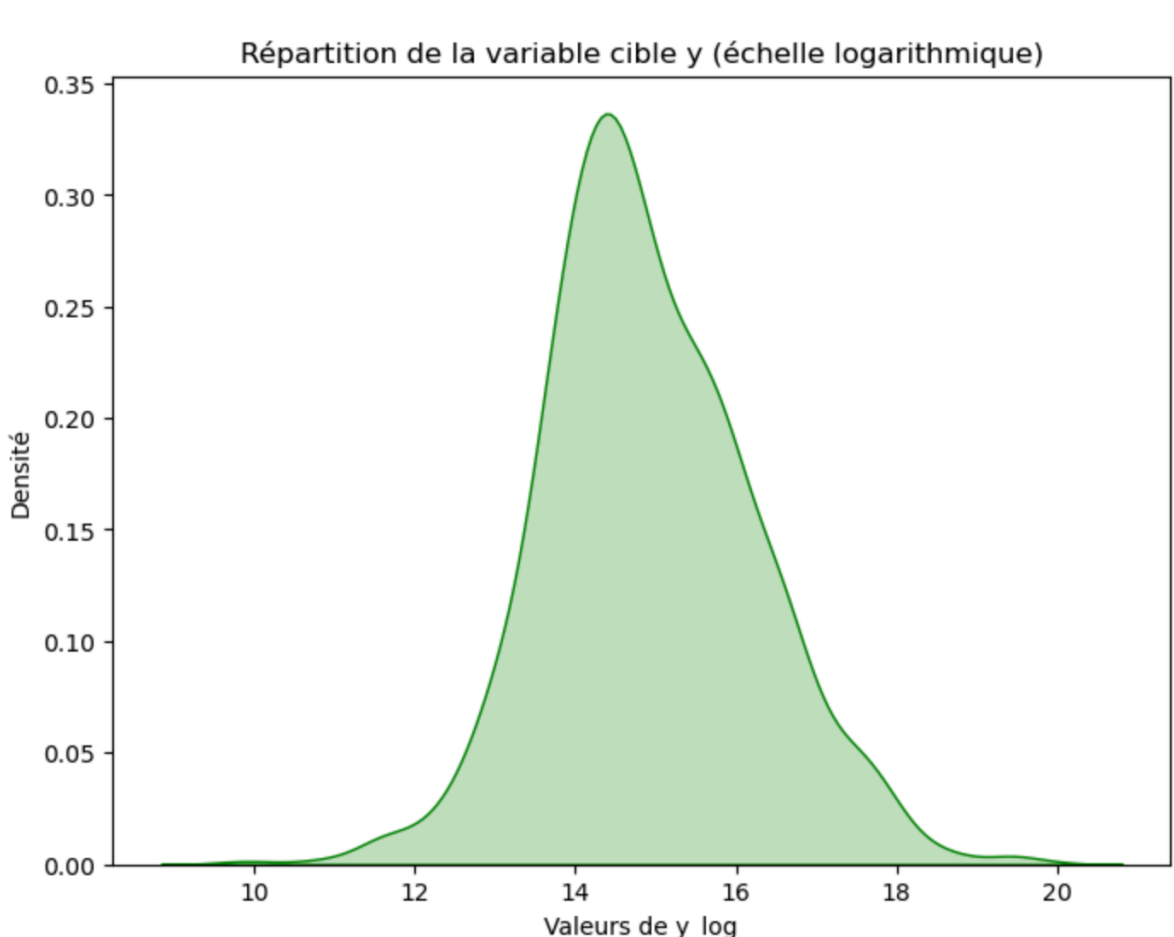
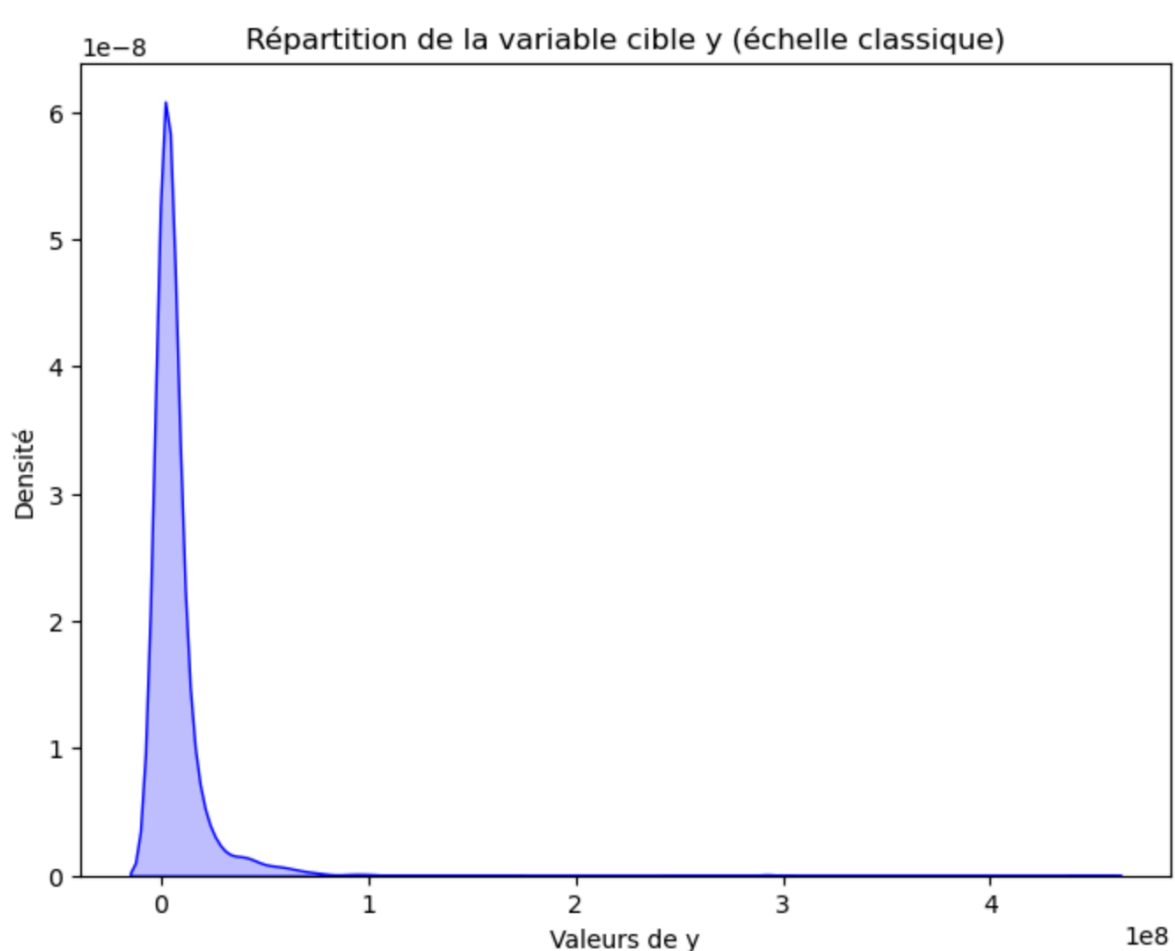
Etude échelle logarithmique

TotalGHGEmissions



y\_train      y\_train\_log      y\_test      y\_test\_log

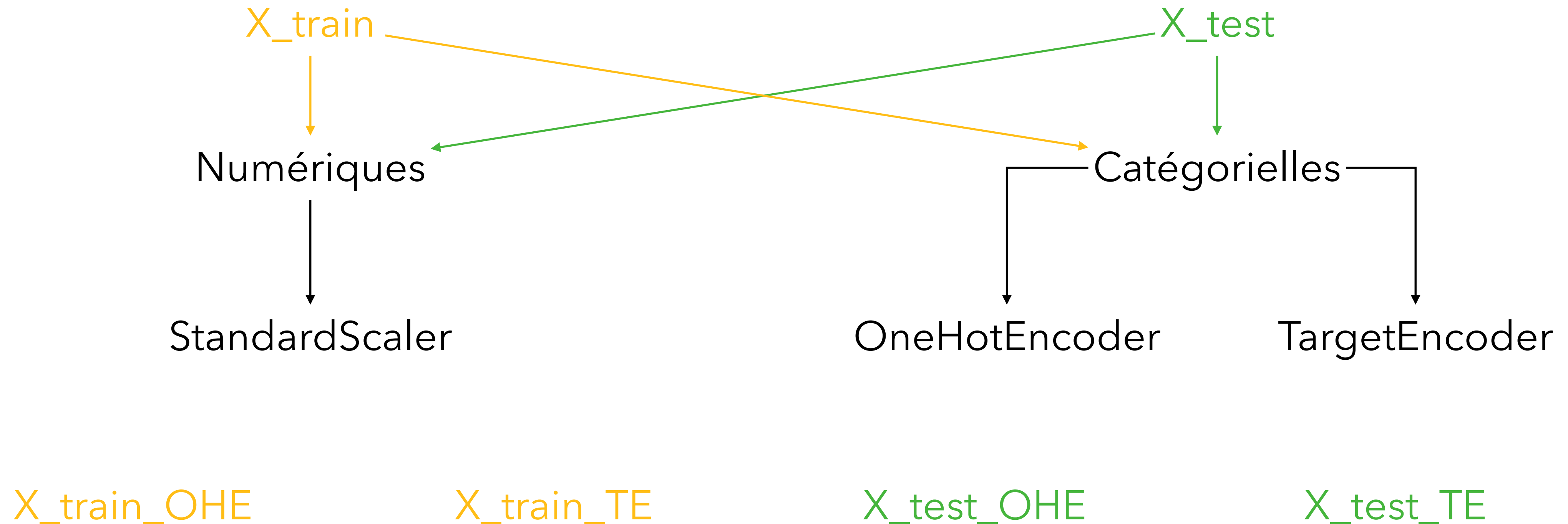
SiteEnergyUse



y\_train      y\_train\_log      y\_test      y\_test\_log

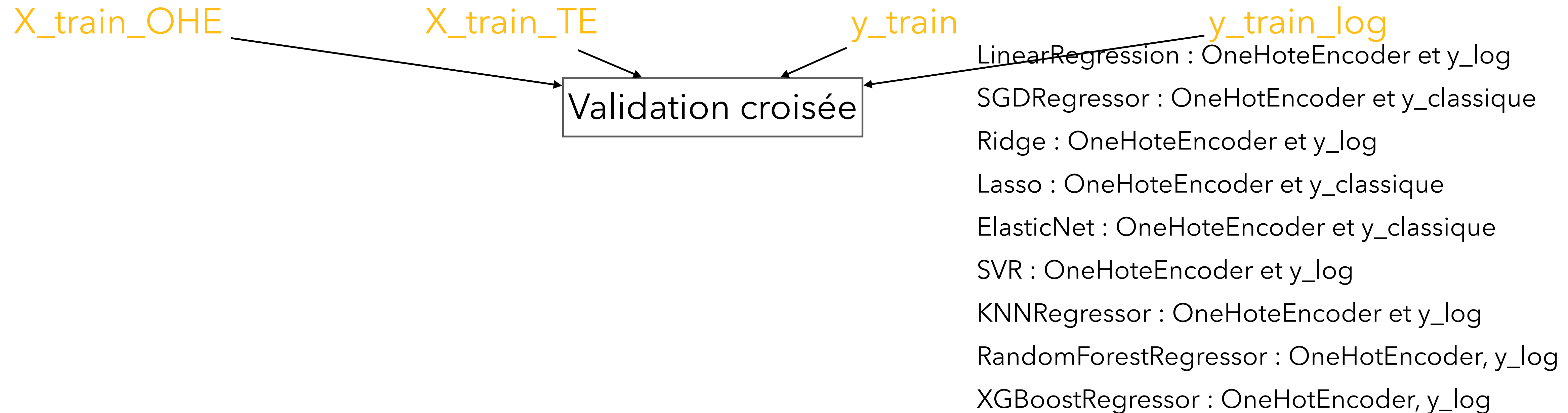
# 4 - Approche méthodologique : optimisation et sélection des modèles

Formatage des données

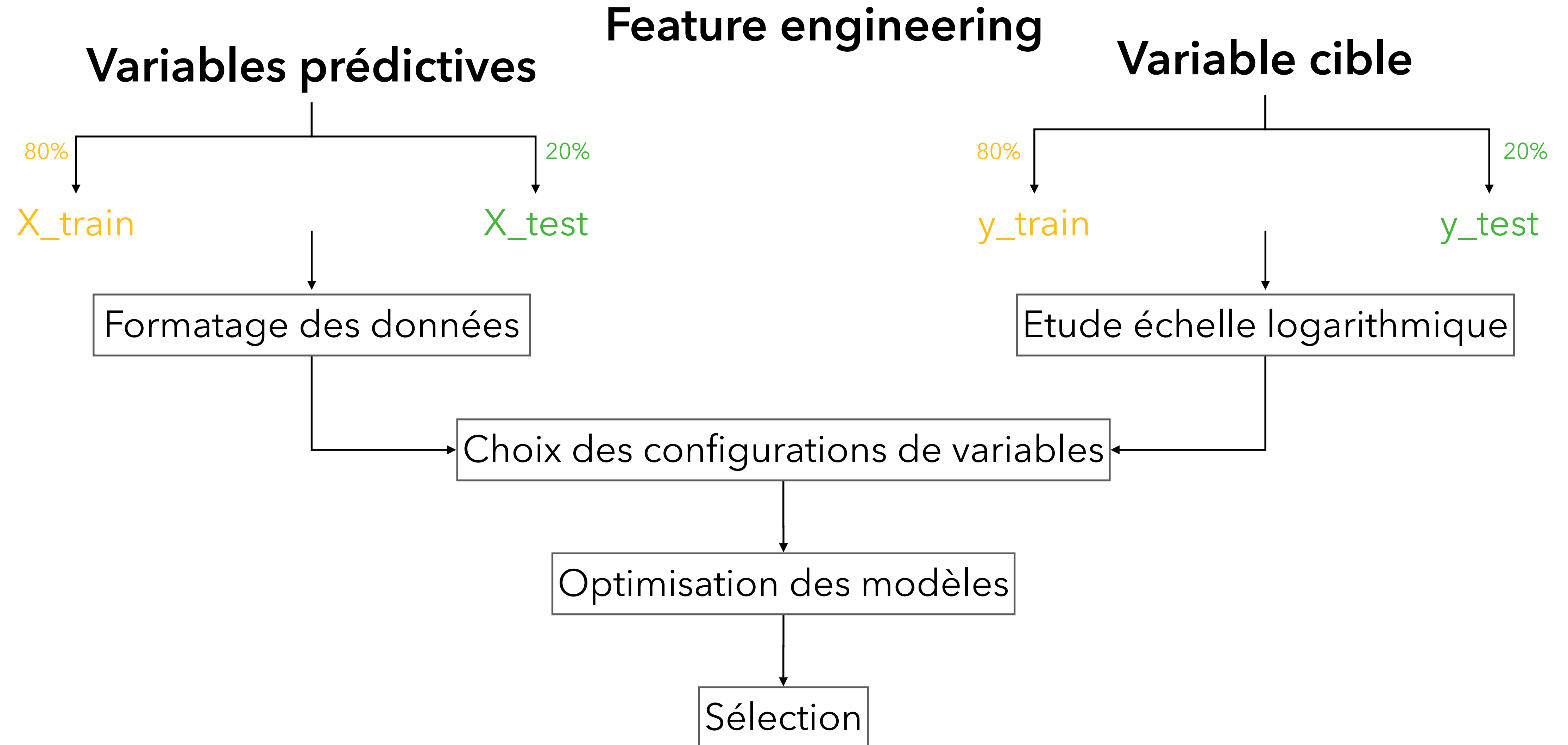


# 4 - Approche méthodologique : optimisation et sélection des modèles

Choix des configurations de variables



# 4 - Approche méthodologique : optimisation et sélection des modèles

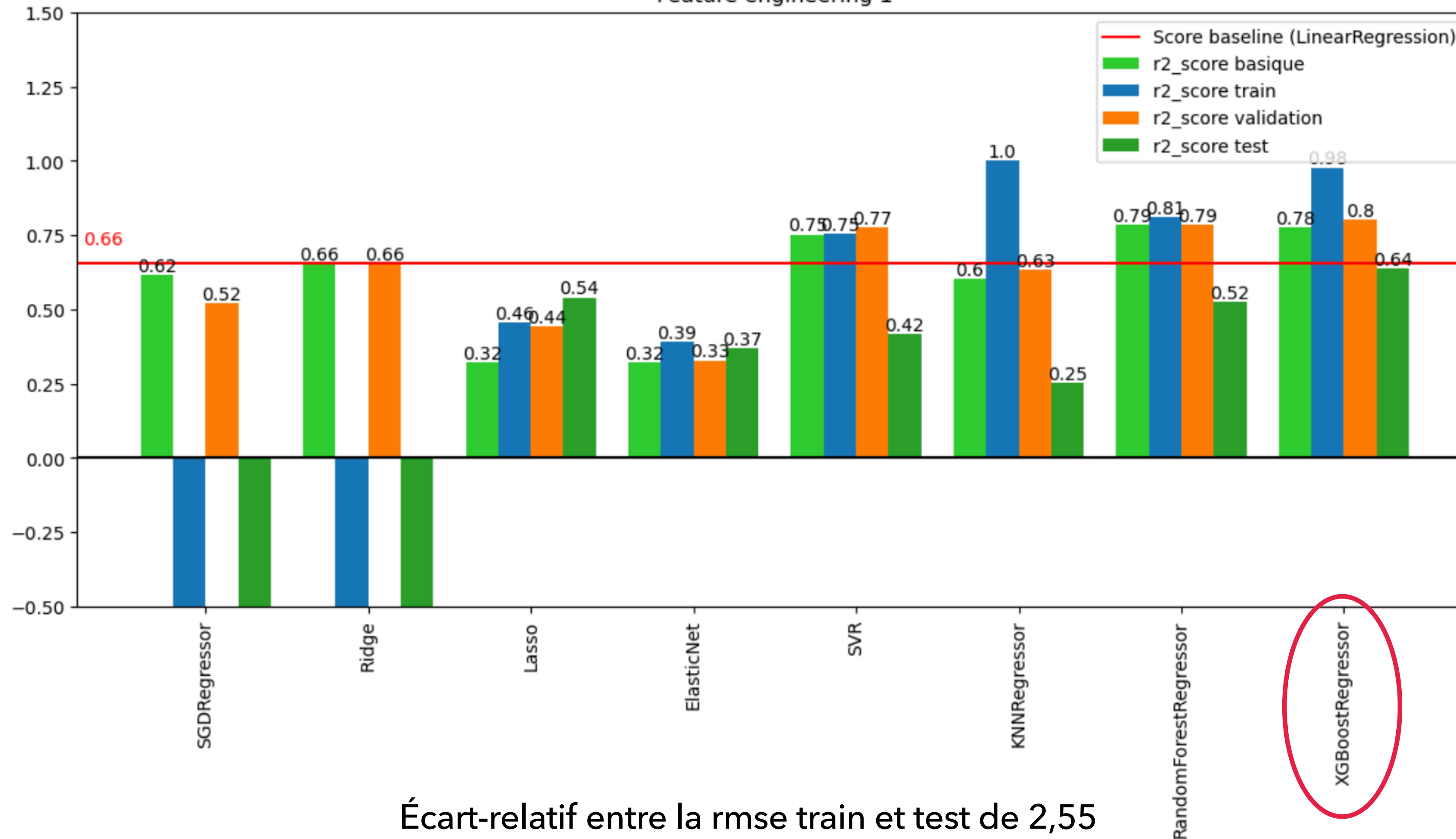


# 5 - Résultats

## Feature engineering 1

Utilisation principale

Prédiction des émissions de CO2  
Feature engineering 1



Consommations énergétiques

Sélection du modèle XGBoostRegressor

R2 train : 0.85

R2 validation : 0.75

R2 test : 0.72

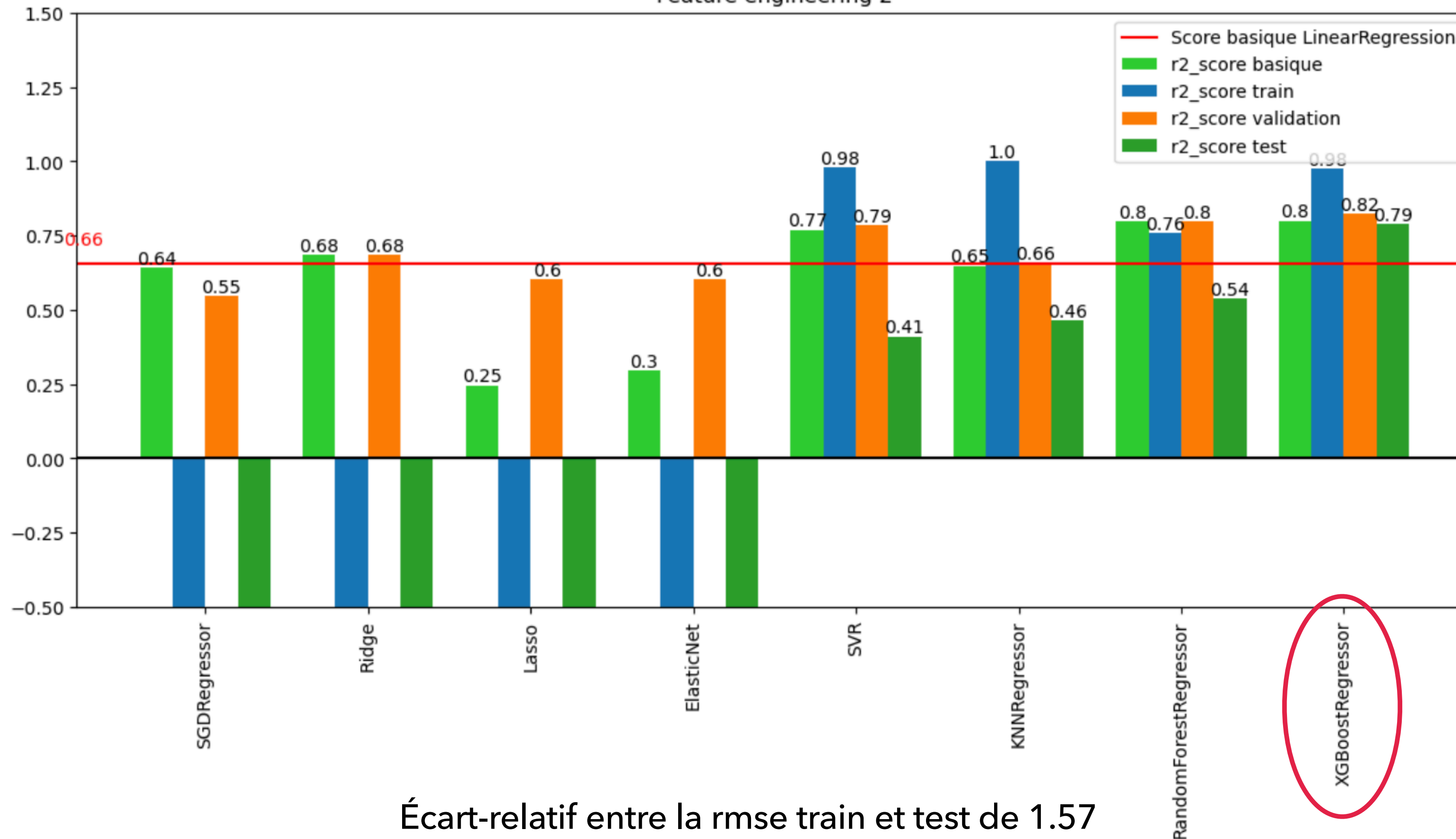
Ecart-relatif rmse train et test : 0.18

# 5 - Résultats

## Feature engineering 2

Toutes utilisations

Prédiction des émissions de CO2  
Feature engineering 2



Consommations énergétiques

Sélection du modèle XGBoostRegressor

R2 train : 0.96

R2 validation : 0.77

R2 test : 0.72

Ecart-relatif rmse train et test : 1.4

# 5 - Retour sur le feature engineering

## Feature engineering 2

Basé sur l'ensemble des types d'utilisations

$X_{\text{train}}$   $\longleftrightarrow$   $X_{\text{test}}$

## Feature engineering 3

OneHotEncoding « personnalisé »

Usage principal  
BuildingType



Surface totale  
PropertyGFATotal

Usage secondaire  
LargestPropertyUseType  
SecondLargestPropertyUseType  
ThirdLargestPropertyUseType



Usages > 3



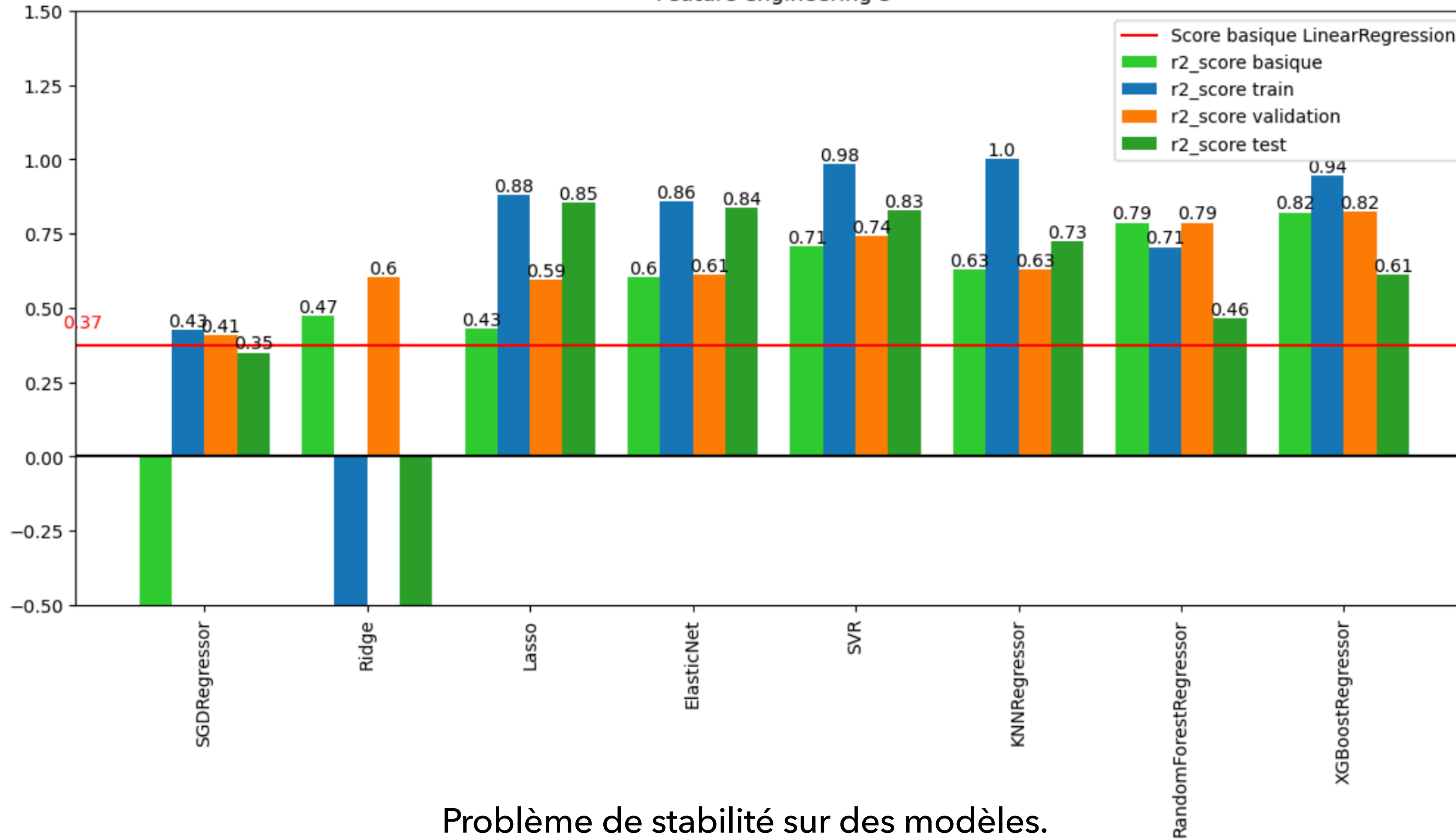
Surface associé  
Somme des surfaces



# 5 - Résultats

## Feature engineering 3 OneHotEncoding « personnalisé »

Prédictions des émissions de CO2  
Feature engineering 3



Problème de stabilité sur des modèles.

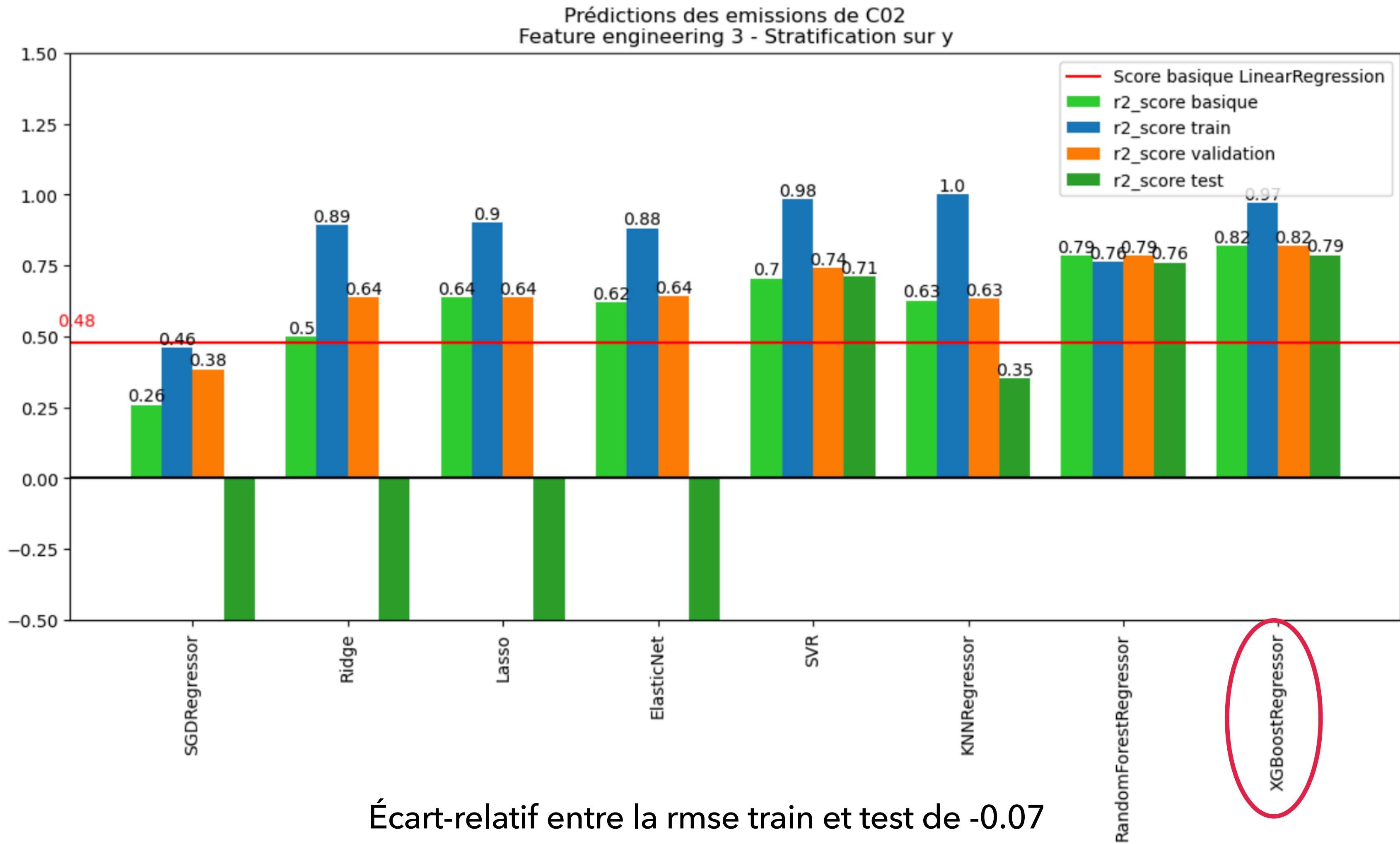
Consommations énergétiques

Problèmes de stabilités également



Feature engineering 3  
OneHotEncoding « personnalisé »  
Et stratification sur la cible

# 5 - Résultats



Consommations énergétiques

Sélection du modèle Lasso

R2 train : 0.91

R2 validation : 0.81

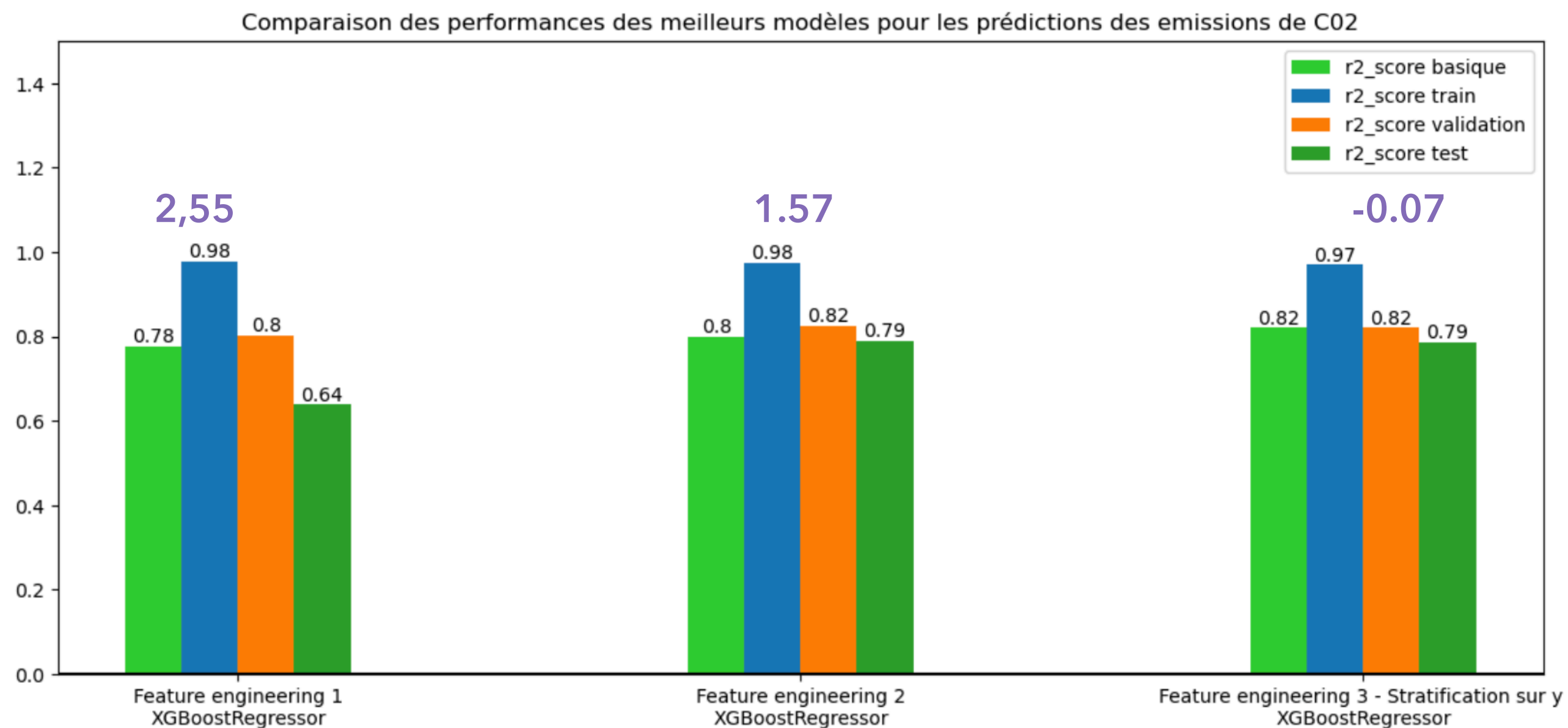
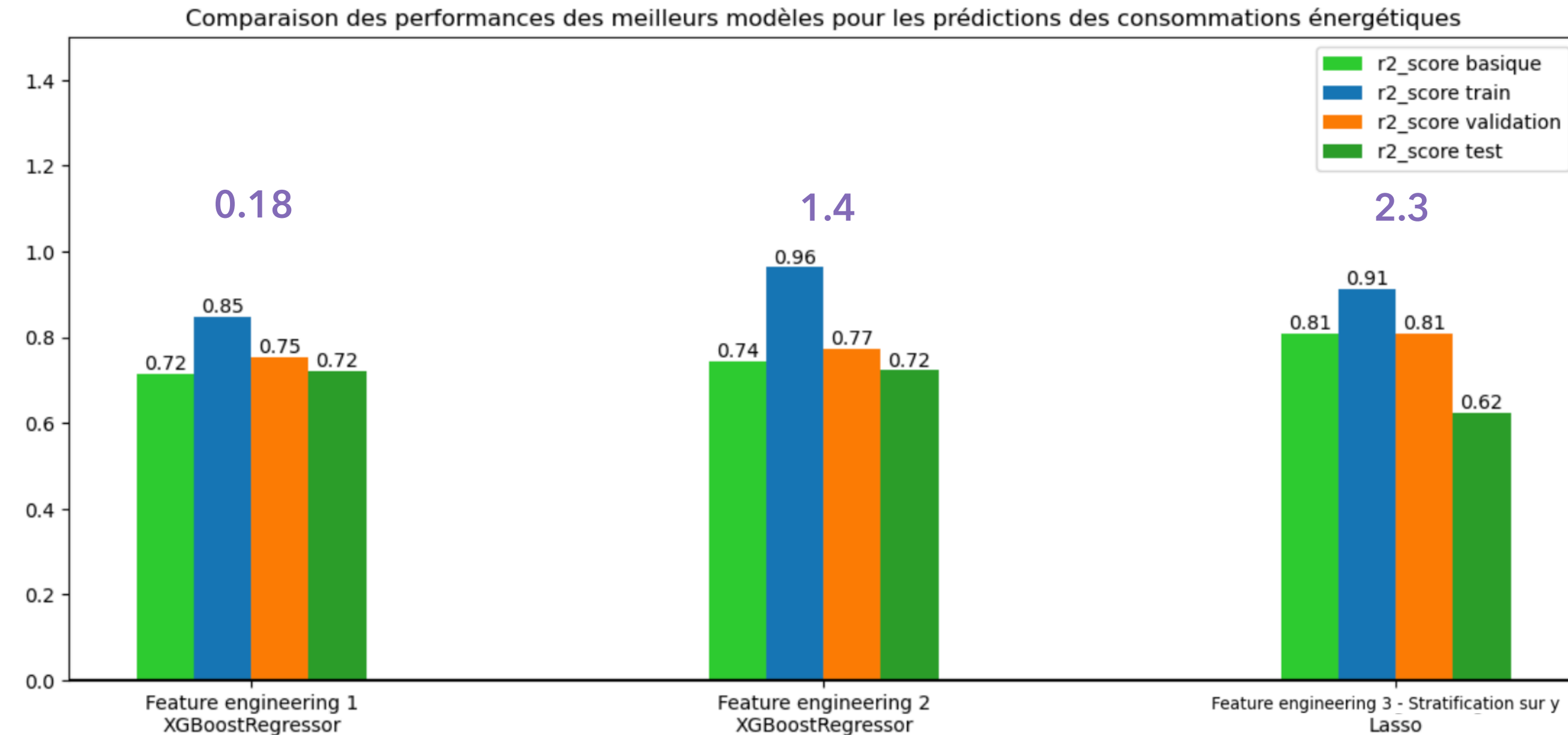
R2 test : 0.62

Ecart-relatif rmse train et test : 2.3

# 5 - Résultats

Prédictions des consommations énergétiques

- > Feature engineering 3 avec stratification
- > Modèle Lasso



Prédictions des émissions de C02

- > Feature engineering 3 avec stratification
- > Modèle XGBoostRegressor

# 6 - Etude de l'ENERGYSTARScore

Identifier l'apport de l'ENERGYSTARScore pour la prédiction des émissions de CO2

Meilleure combinaison

→ feature engineering 3 + stratification

→ modèle XGBoostRegressor

## Scénario 1



Optimisation du modèle

Uniquement sur les sites  
où l'ENERGYSTARScore est présent

## Scénario 1\_bis



Optimisation du modèle

Sur ces mêmes sites  
mais sans prendre la variable de l'ENERGYSTARScore

## Scénario 2



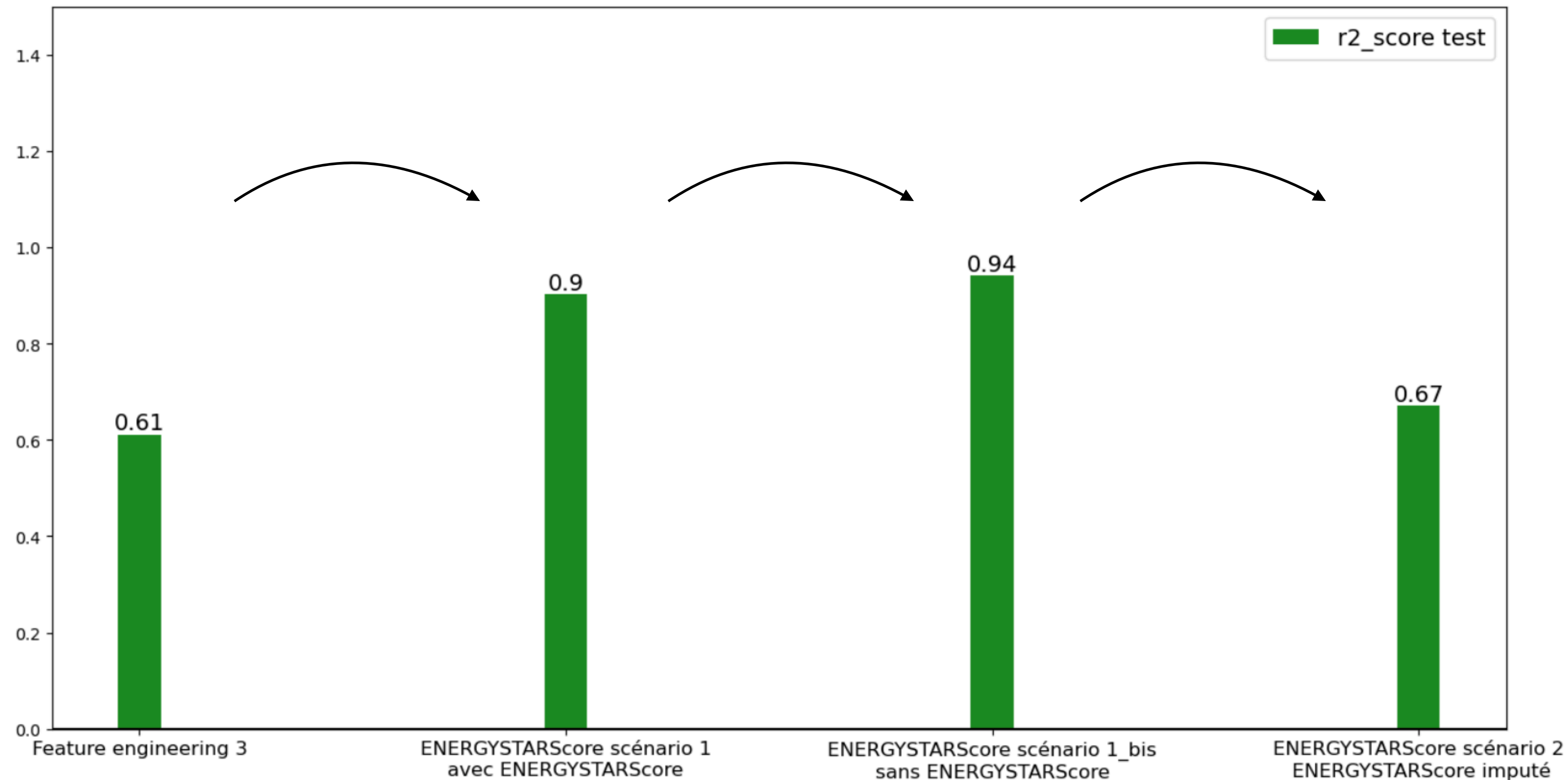
Optimisation du modèle

Sur l'ensemble des sites  
avec une imputation l'ENERGYSTARScore  
si la valeur est absente.

# 6 - Etude de l'ENERGYSTARScore

Identifier l'apport de l'ENERGYSTARScore pour la prédiction des émissions de CO2

Prédictions des émissions de CO2  
Comparaison des performances du modèle XGBoosRegressor  
Intérêt de l'ENERGYSTARScore



Augmentation des performances

Augmentation légèrement plus importante

Légère augmentation des performances

Influence minime de l'ENERGYSTARScore

# 6 - Conclusion

## Prédictions des émissions de CO<sub>2</sub>

### Meilleure combinaison

- feature engineering 3 (OneHotEncoding « personnalisé »), avec stratification
- modèle XGBoostRegressor

## Prédictions des consommations énergétiques

### Meilleure combinaison

- feature engineering 3 (OneHotEncoding « personnalisé »), avec stratification
- modèle Lasso

### Réserve

- Optimiation pour les sites les plus courants  
(Sites avec grand nombre de bâtiments non pris en compte)

### Abandon du calcul de l'ENERGYSTARScore

- Imputation des valeurs manquantes est préférable.

**Question ?**