

Neutralité carbone 2050



Seattle

Prédictions Émissions de CO₂ Consommation totale d'énergie

Bâtiments non résidentiels

Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entrainement

Évaluation

Prédiction finale

Piste de progression



Seattle

Exposition extérieure



Localisation

- Adresse
- Longitude/latitude
- ZipCode
- Neighborhood

Bâtiment



Année de construction



Matériaux, isolation



Surfaces

- Surfaces construites
- Nombre de bâtiments
- Nombre d'étages
- Surfaces des 3 premiers types d'utilisation

Utilisation



Type d'utilisation

- Type d'utilisation principale
- 3 premiers types d'utilisations



Températures de consigne



Heures d'ouverture

Énergies



Scoring énergétique

- EnergyStar



Types d'énergie

- Répartition

Consommation totale
GHG Emissions
WN

Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entrainement

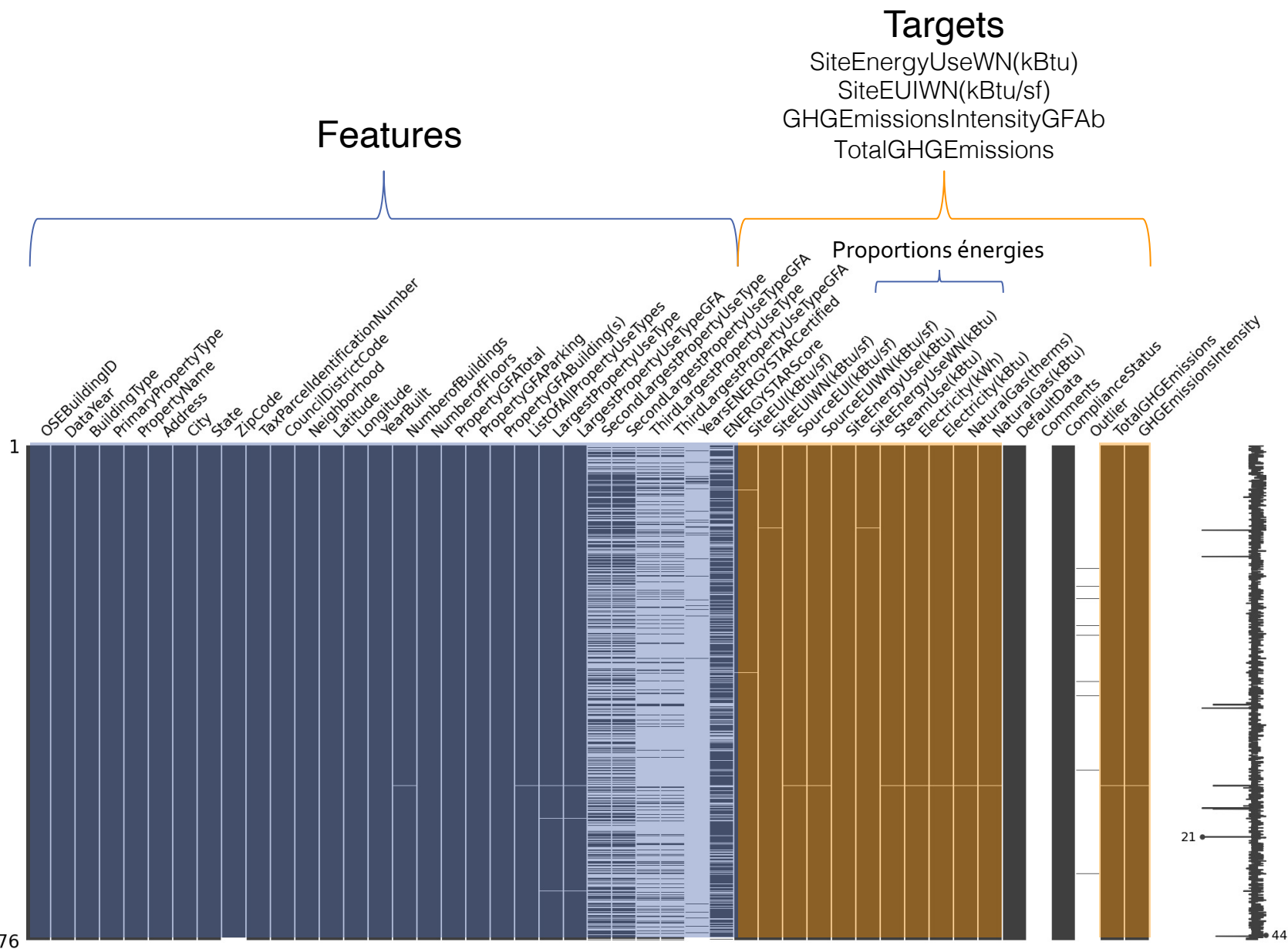
Évaluation

Prédiction finale

Piste de progression



Seattle





Seattle

Sommaire

Objectifs

Jeu de données

Process

Feature engineering 1/2

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

Prédiction finale

Piste de progression



Seattle

Données structurées

- Nettoyage :
 - Suppression des bâtiments résidentiels
 - Suppression des bâtiments non 'compliant' de la variable ComplianceStatus
 - Suppression des lignes où les consommations Electricité+Gaz+Vapeur = 0
- Transformation :
 - $\log(x+1)$ si skewness > 3
 - Agrégation de certaines valeurs de PrimaryPropertyType
- Encodage pour n'avoir que des variables numériques :
 - PrimaryPropertyType : OneHotEncoder
 - Neighborhood : OneHotEncoder
 - xxLargestPropertyUseType : PropertyUseType en colonnes avec le pourcentage de surface en donnée

Sommaire

Objectifs

Jeu de données

Process

Feature engineering 2/2

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

Prédiction finale

Piste de progression

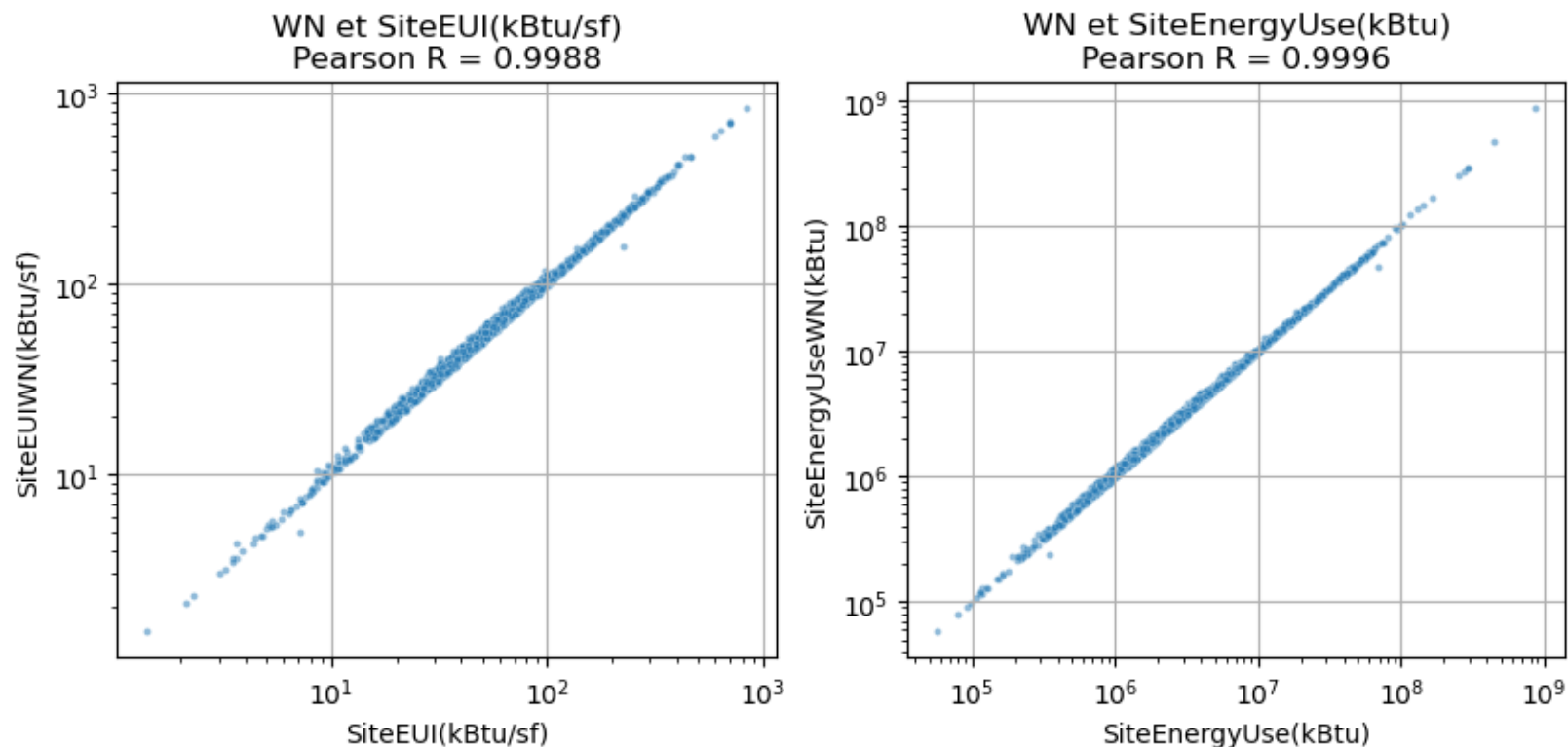


Seattle

Targets

- Nettoyage :
 - Imputation de SiteEnergyUseWN(kBtu) nul par SiteEnergyUse(kBtu)
 - Imputation de SiteEUIWN(kBtu/sf) nul par SiteEUI(kBtu/sf)

Corrélations entre les consommations et les consommations normalisées sur 30 ans



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

Prédiction finale

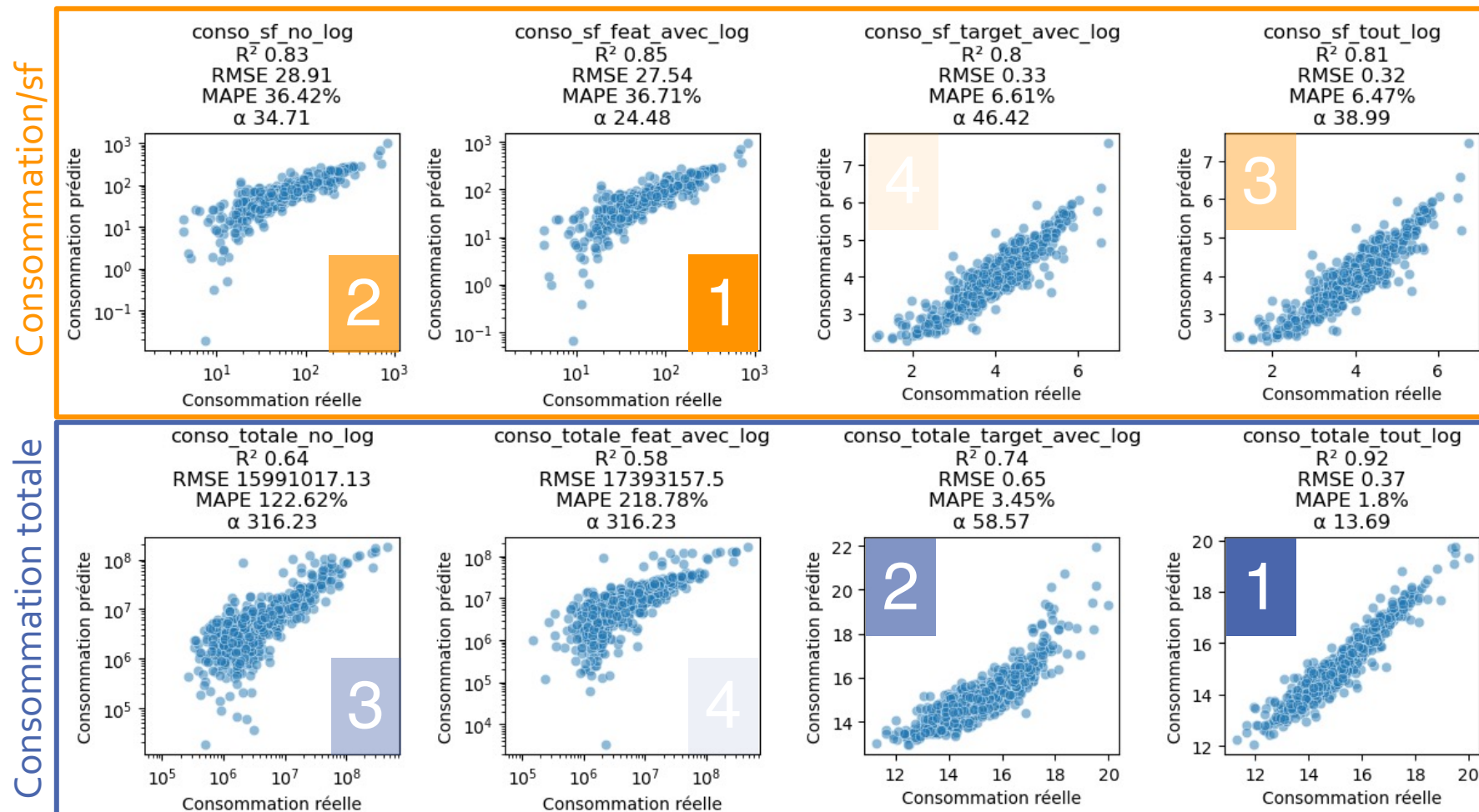
Piste de progression



Seattle

Transformation log de la variable si skewness > 3

Influence de la transformation log



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Transformation log

Jeux de features

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

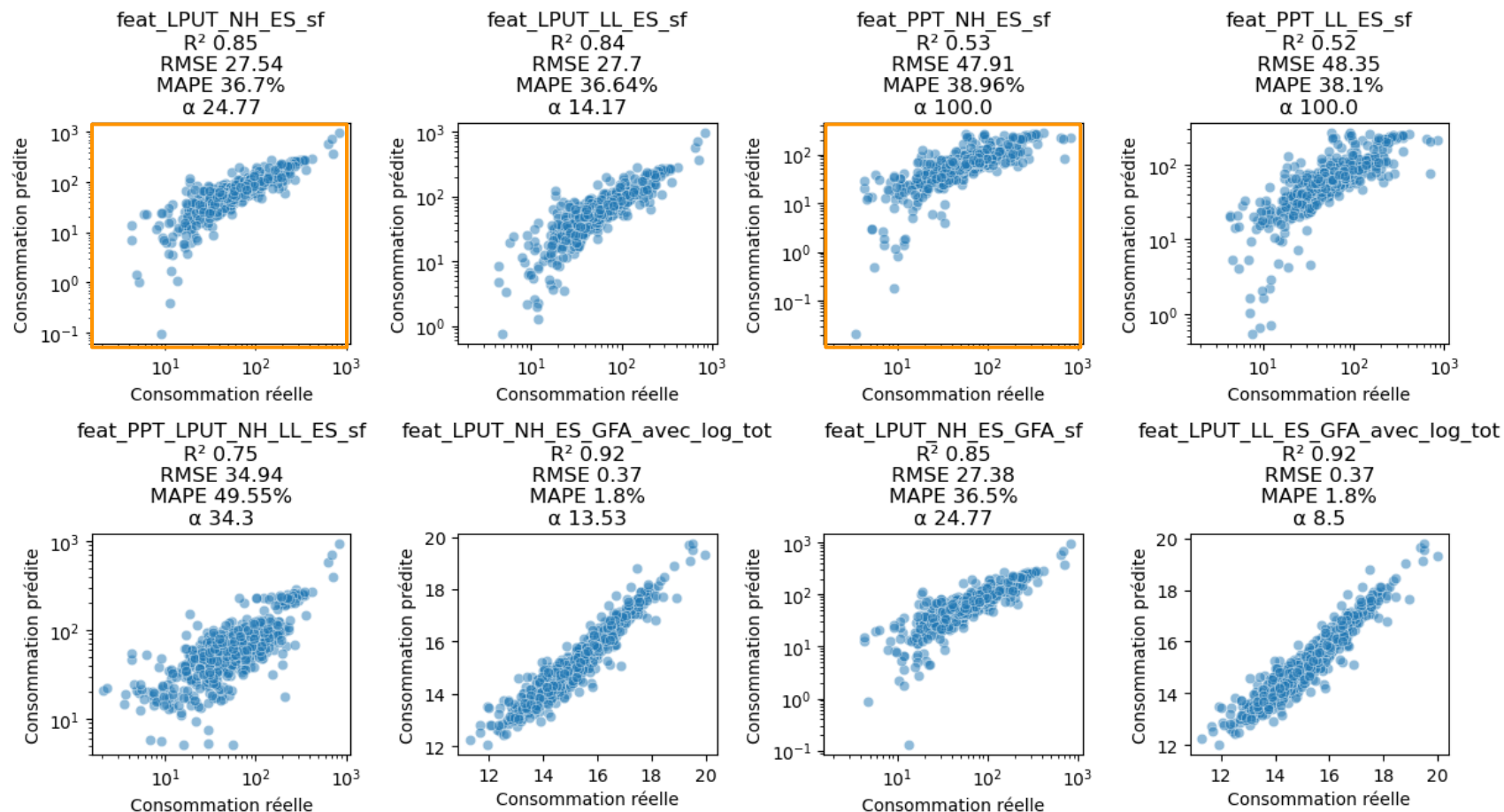
Prédiction finale

Piste de progression



Seattle

Influence de la sélection des features



LPUT PPT

Conso_tot

Conso_sf

NH

LL

GFA

Ø

Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

Prédiction finale

Piste de progression



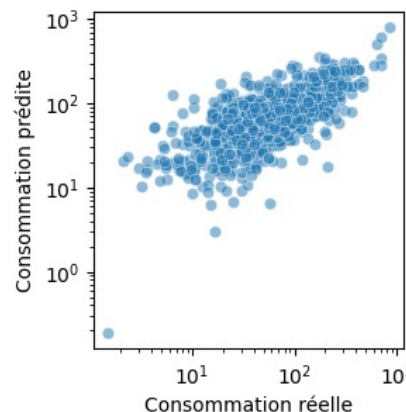
Seattle

06/06/2023

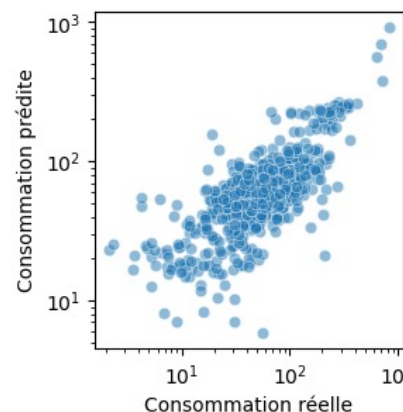
Suppression des individus sans EnergyStarScore

conso_sf_1503_indiv
 R^2 0.59

Consommation / sf : + 0.16

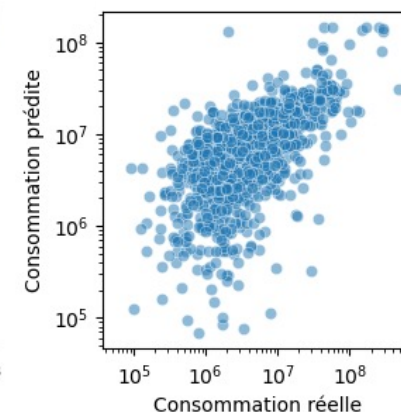


conso_sf_968_indiv
 R^2 0.75

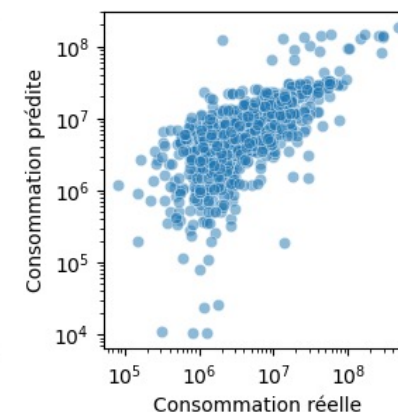


conso_totale_1503_indiv
 R^2 0.39

Consommation totale : + 0.15



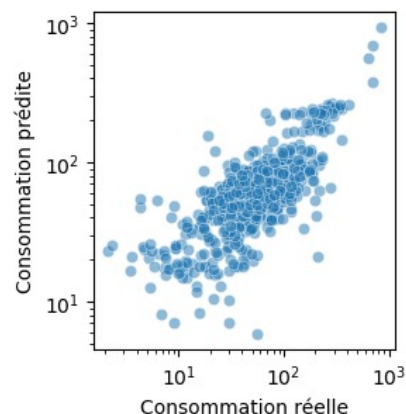
conso_totale_968_indiv
 R^2 0.54



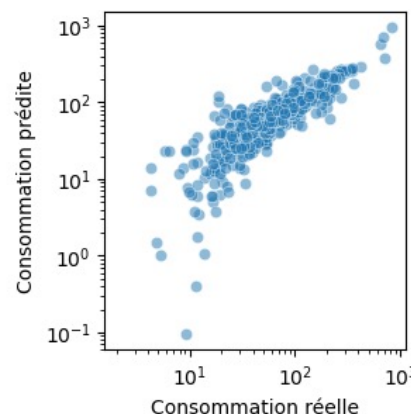
Influence de EnergyStarScore

conso_sf_sans_es
 R^2 0.75

Consommation / sf : + 0.10

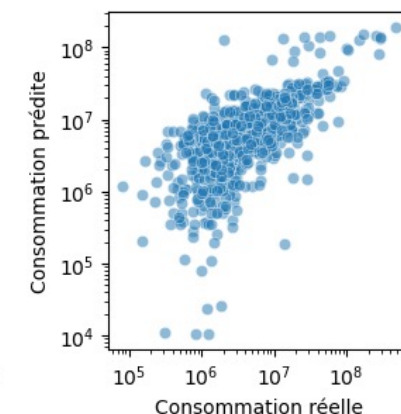


conso_sf_avec_es
 R^2 0.85

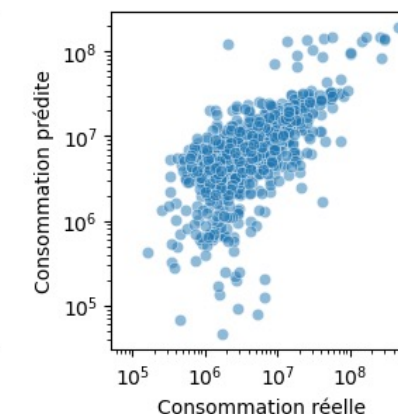


conso_totale_sans_es
 R^2 0.54

Consommation totale : + 0.01



conso_totale_avec_es
 R^2 0.55



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

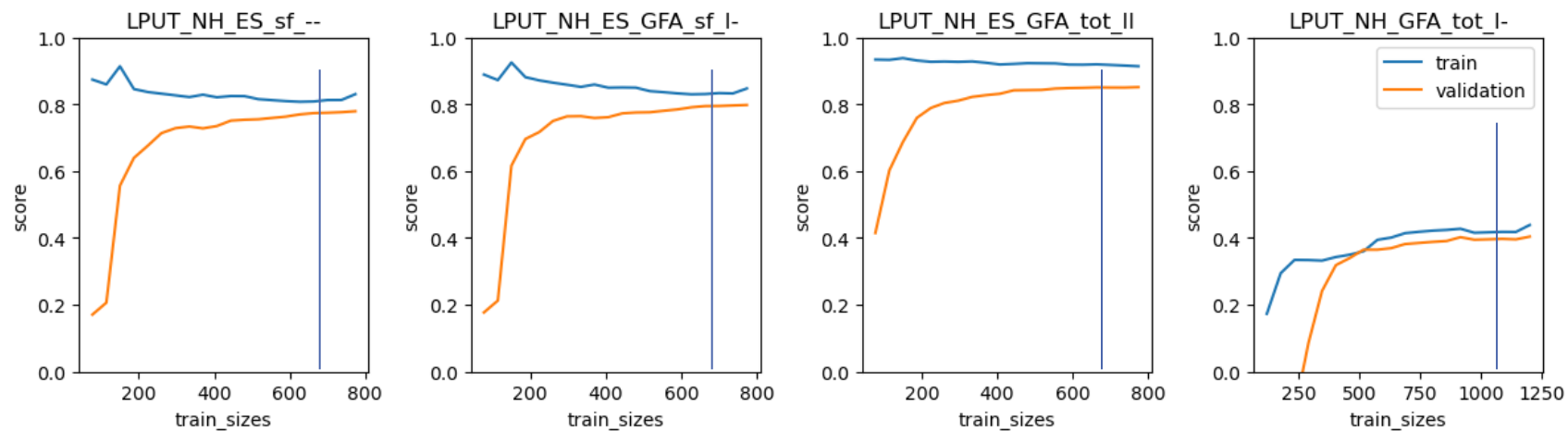
Prédiction finale

Piste de progression

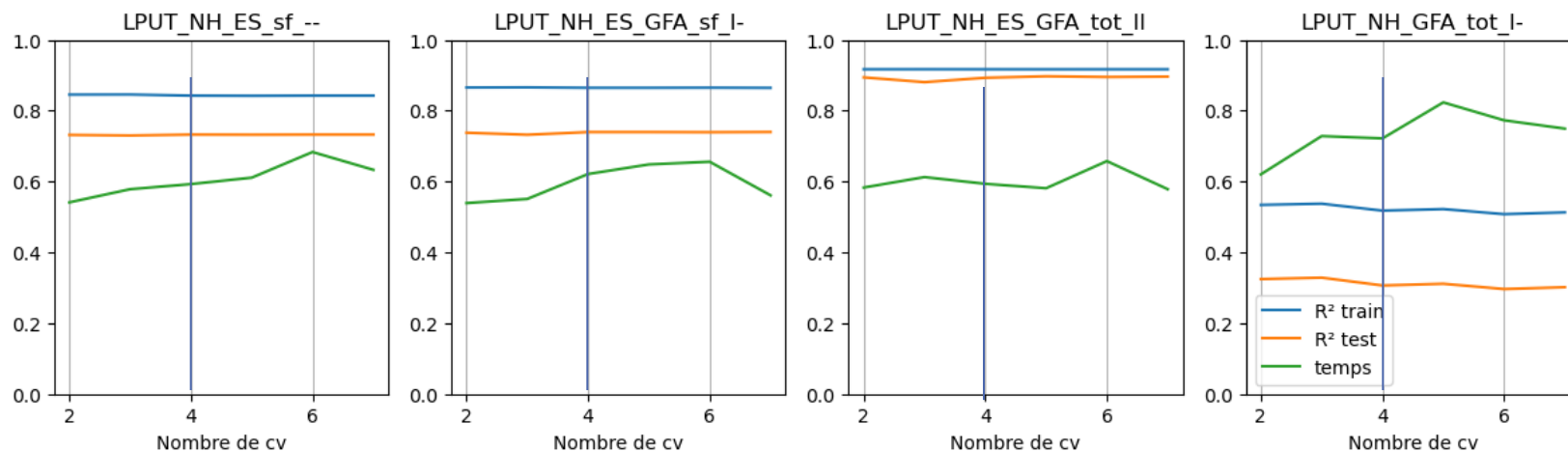


Seattle

Choix de la taille du train_test_split optimale **test_size = 0,30**



Choix du nombre de séparations de la CrossValidation **cv = 4**



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement 1/3

Évaluation

Prédiction finale

Piste de progression



Seattle

Régressions

Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement 2/3

Évaluation

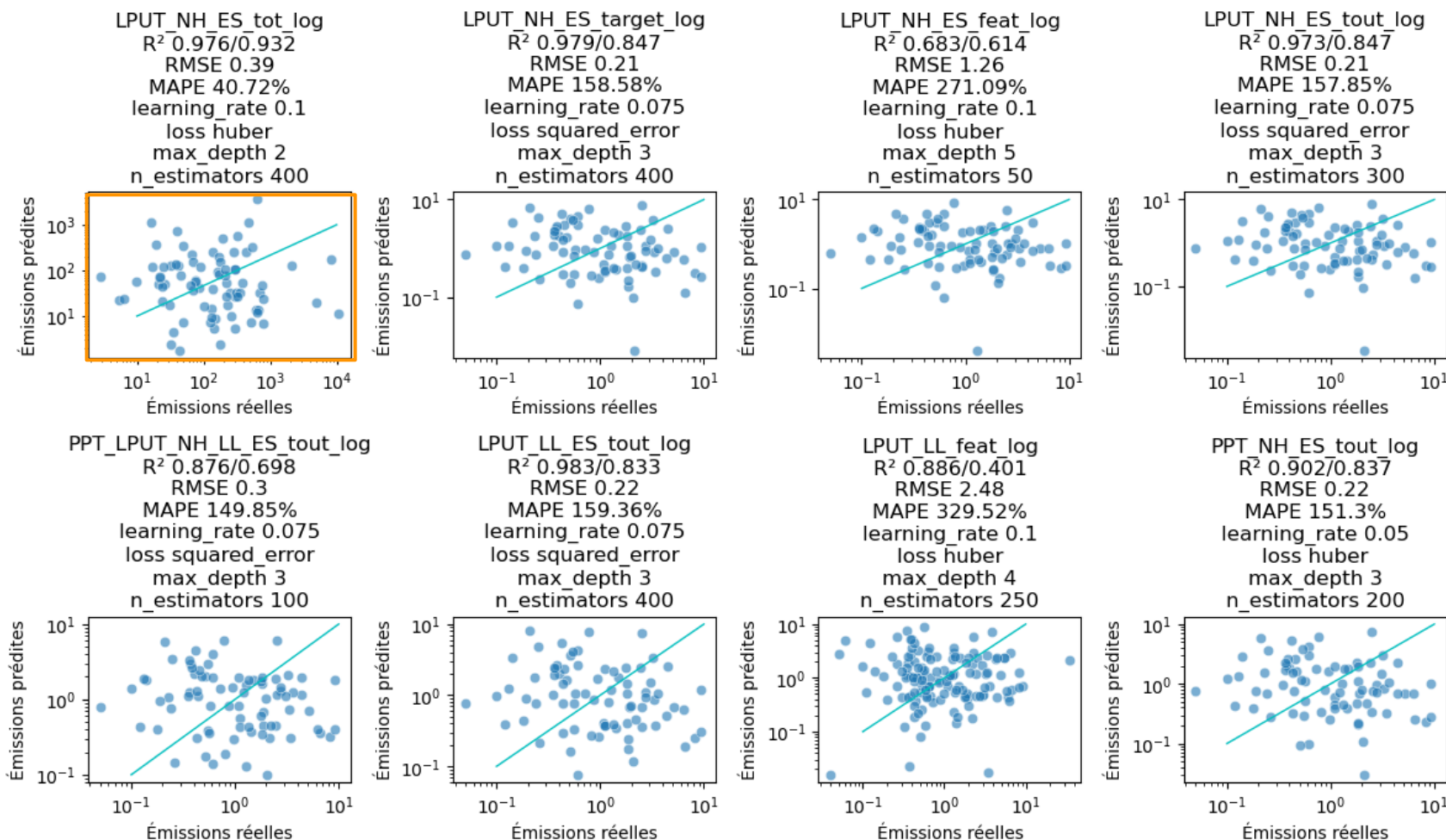
Prédiction finale

Piste de progression



Seattle

Pour le meilleur R^2 sur le jeu test, récupération des valeurs des hyperparamètres – ici GradientBoosting pour target GHG



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entraînement 3/3

Évaluation

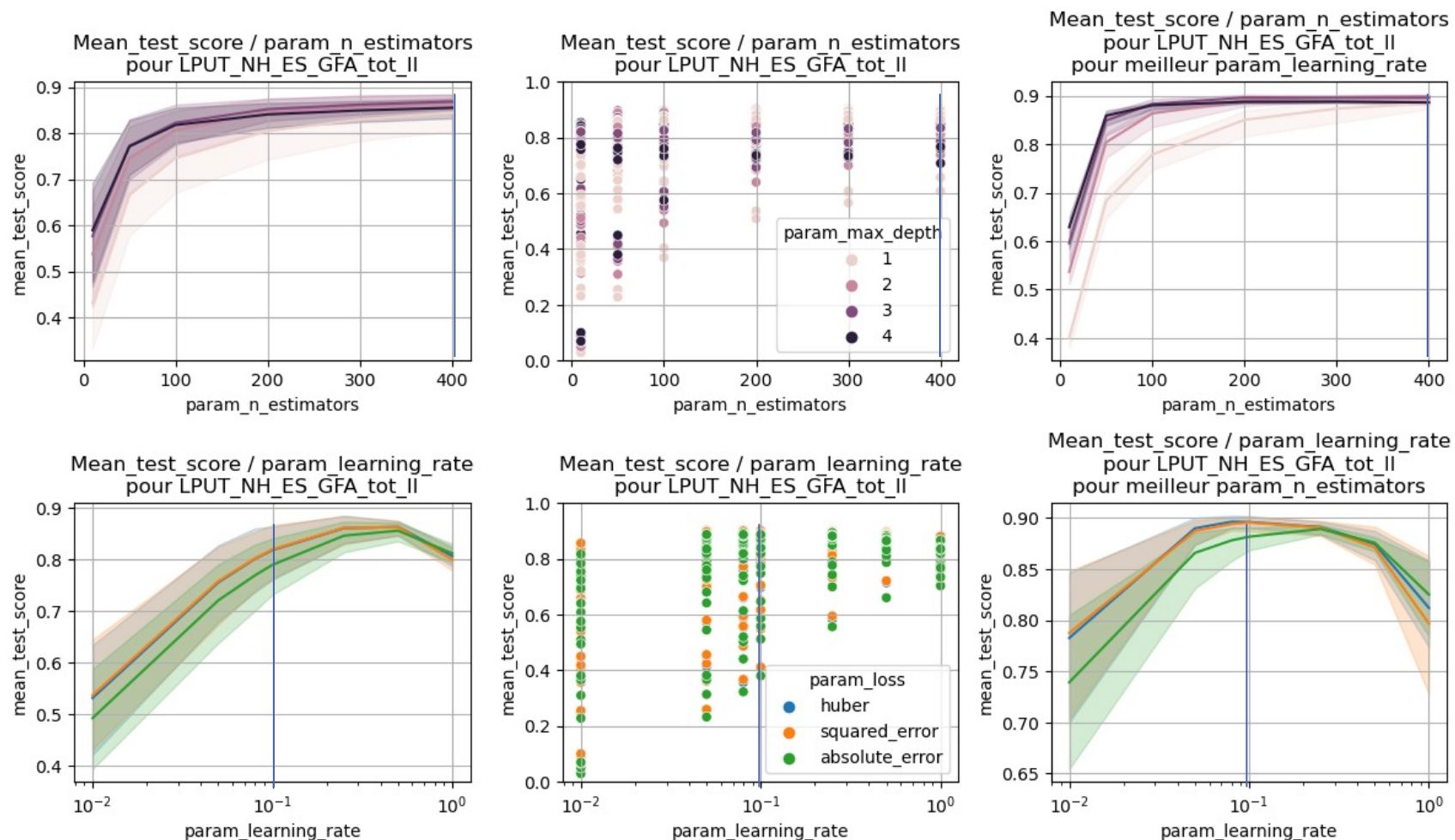
Prédiction finale

Piste de progression



Seattle

Affichage des courbes de score en fonction de ces hyperparamètres –
ici pour GradientBoosting pour la consommation d'énergie



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement

Évaluation 1/2

Prédiction finale

Piste de progression



Seattle

Comparaison des scores R^2_{test} selon les méthodes utilisées



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement

Évaluation 2/2

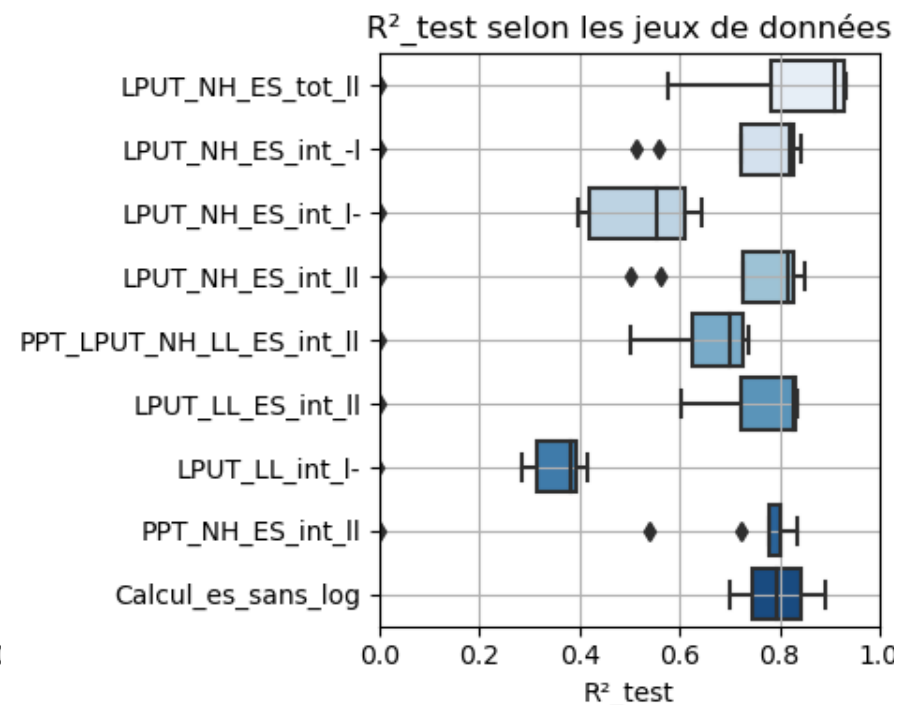
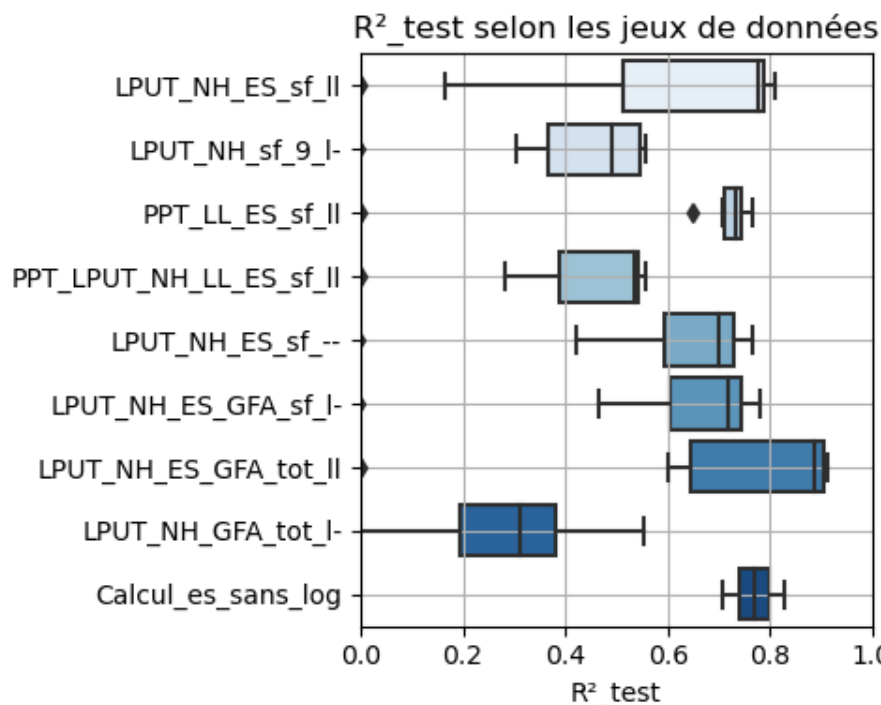
Prédiction finale

Piste de progression



Seattle

Comparaison des scores R^2_{test} selon les jeux de données utilisés



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement

Évaluation

Prédiction finale 1/2

Piste de progression



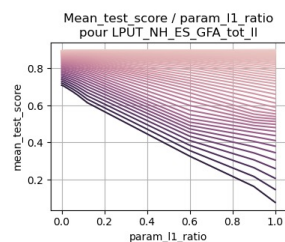
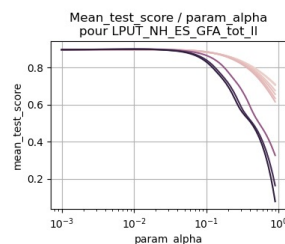
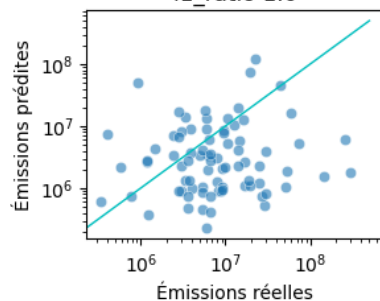
Seattle

Consommation d'énergie

ElasticNet $R^2 = 0.91$

0,0019 secondes

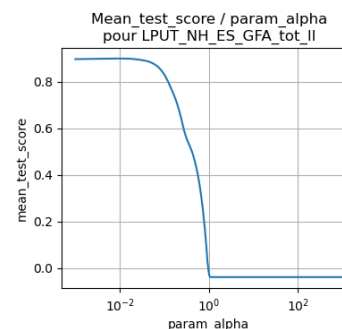
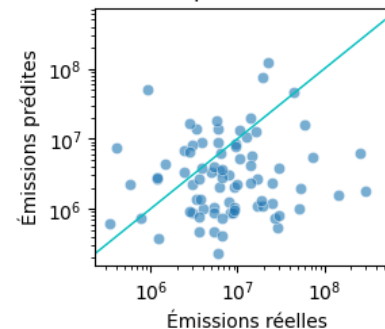
LPUT_NH_ES_GFA_tot_II
 R^2 0.914/0.91
RMSE 0.39
MAPE 9.84%
alpha 0.0103
l1_ratio 1.0



Lasso $R^2 = 0.91$

0,0020 secondes

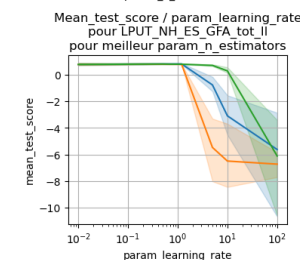
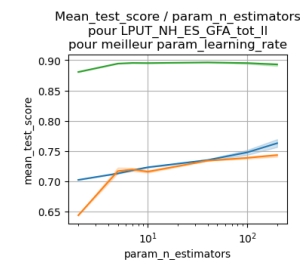
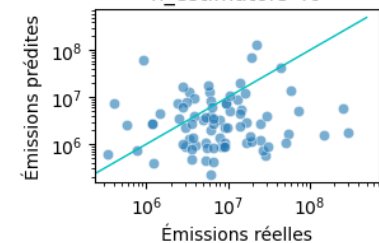
LPUT_NH_ES_GFA_tot_II
 R^2 0.914/0.91
RMSE 0.39
MAPE 9.84%
alpha 0.0106



Adaboost $R^2 = 0.911$

0,1297 secondes

LPUT_NH_ES_GFA_tot_II
 R^2 0.916/0.911
RMSE 0.39
MAPE 9.84%
estimator ElasticNet(alpha=0.0103)
learning_rate 0.01
loss exponential
n_estimators 40



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entrainement

Évaluation

Prédiction finale 2/2

Piste de progression



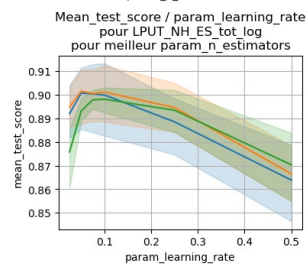
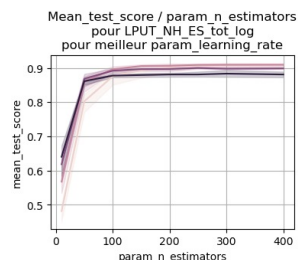
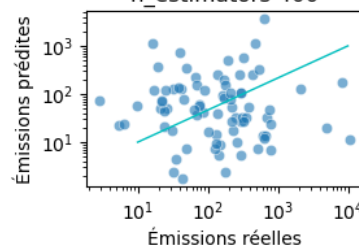
Seattle

Émissions de gaz à effet de serre

Gradient Boosting

$R^2=0.932$ 0,1412 secondes

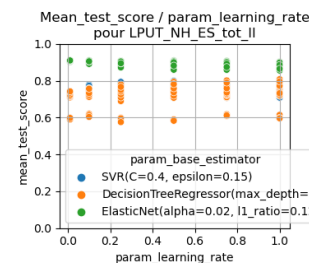
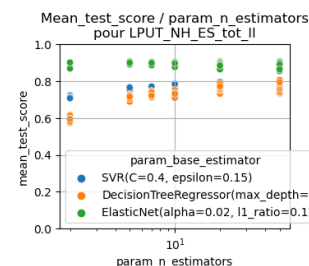
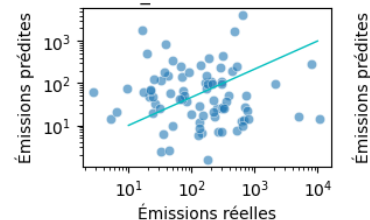
LPUT_NH_ES_tot_log
 R^2 0.976/0.932
RMSE 0.39
MAPE 40.72%
learning_rate 0.1
loss huber
max_depth 2
n_estimators 400



Adaboost $R^2 = 0.928$

0,1683 secondes

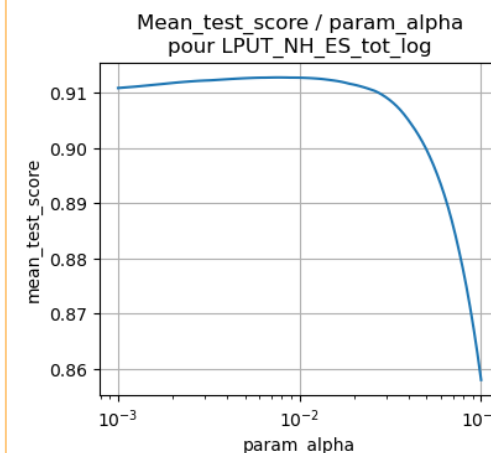
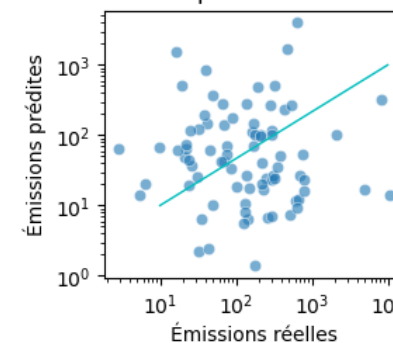
LPUT_NH_ES_tot_II
 R^2 0.929/0.928
RMSE 0.4
MAPE 40.47%
estimator ElasticNet(alpha=0.02, l1_ratio=0.12)
learning_rate 0.01
loss linear
n_estimators 50



Lasso $R^2 = 0.927$

0,0018 secondes

LPUT_NH_ES_tot_log
 R^2 0.928/0.927
RMSE 0.4
MAPE 40.57%
alpha 0.0077



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

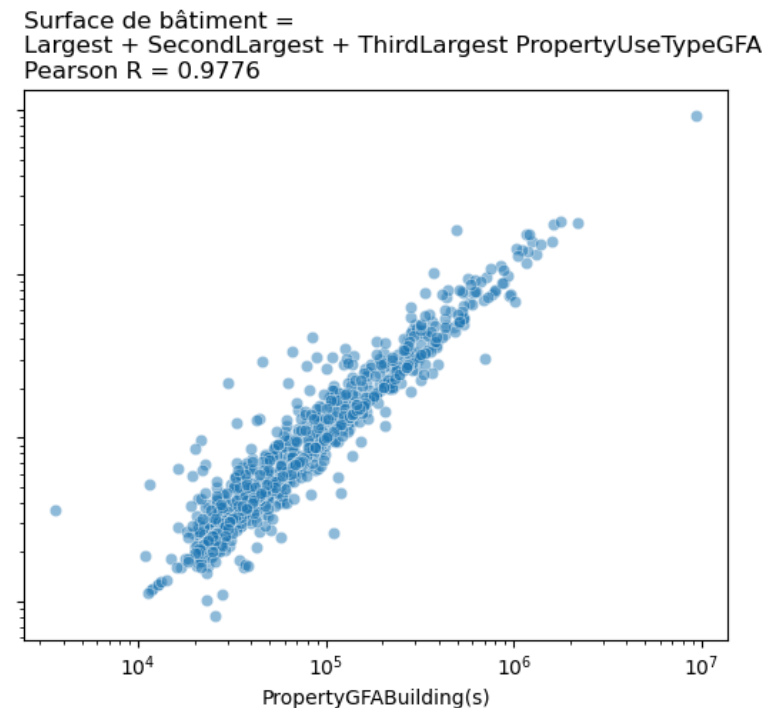
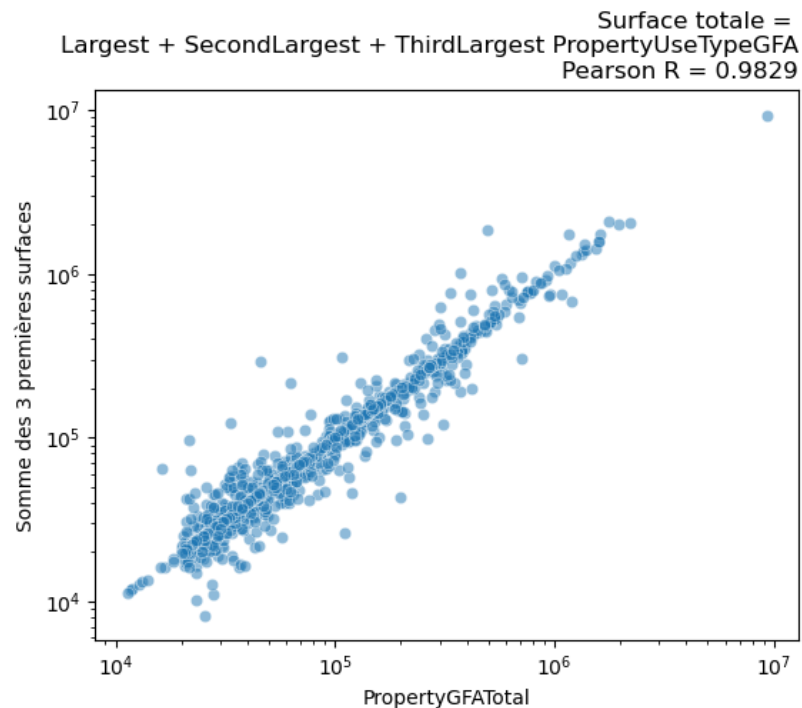
Prédiction finale

Piste de progression
1/2



Seattle

- Avoir des surfaces cohérentes dans le dataset de base



Sommaire

Objectifs

Jeu de données

Process

Feature engineering

Jeux de features

Transformation log

EnergyStarScore

Modélisation

Préparation

Entraînement

Évaluation

Prédiction finale

Piste de progression

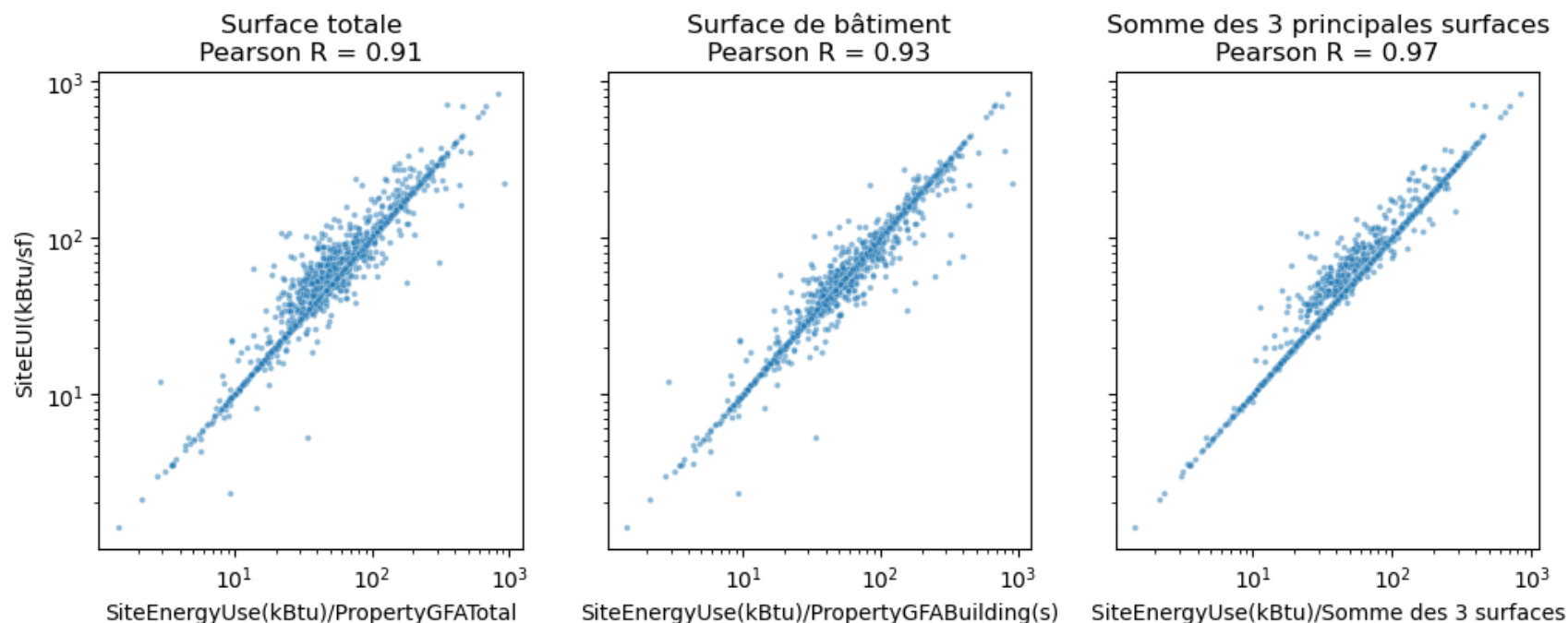
2/2



Seattle

- Avoir une cohérence entre l'énergie consommée par unité de surface et l'énergie totale

Corrélations entre surfaces et la consommation d'énergie totale



Neutralité carbone 2050



Seattle

Merci pour votre attention