



CentraleSupélec

Anticipation de la consommation d'énergie et des émissions de CO_2

LO Ousmane, Ingénieur Machine Learning



01

INTRODUCTION

02

ANALYSES
DESCRIPTIVES

03

MACHINE
LEARNING ET
MODELISATION

04

CONCLUSION

01

01 01 01

INTRODUCTION

Problématique, hypothèses et pistes
d'exploration



Contexte de l'étude

Objectif : Être une ville neutre en émissions de gaz à effets de serre pour 2050

Données : Deux jeux de données sur les bâtiments non-résidentiels (2016)



- Prédire la consommation d'énergie
- Prédire les émissions de CO2
- Intérêt de l'utilisation du score ENERGY STAR



02

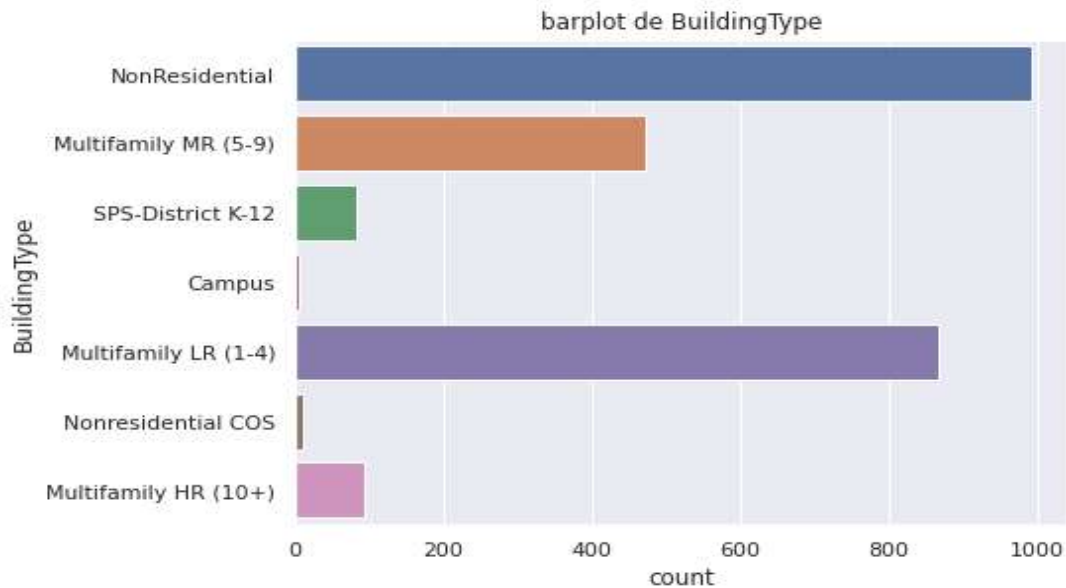
ANALYSES EXPLORATOIRES

Analyses statistiques descriptives
univariées et bivariées



Analyses univariées

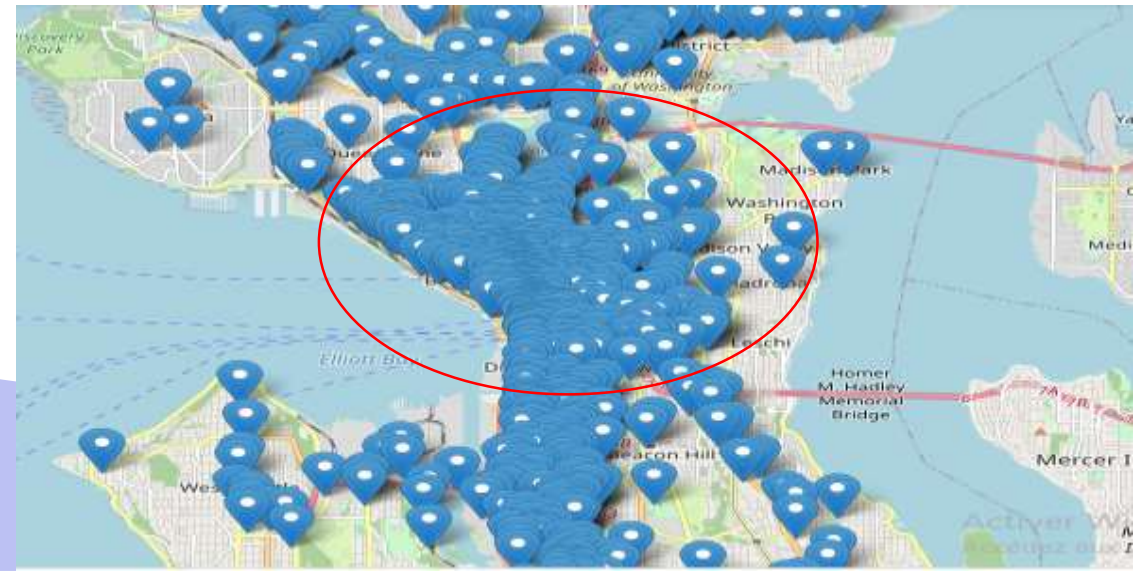
- Répartition des types bâtiments



- Concentration des bâtiments non résidentiels dans le centre-ville.

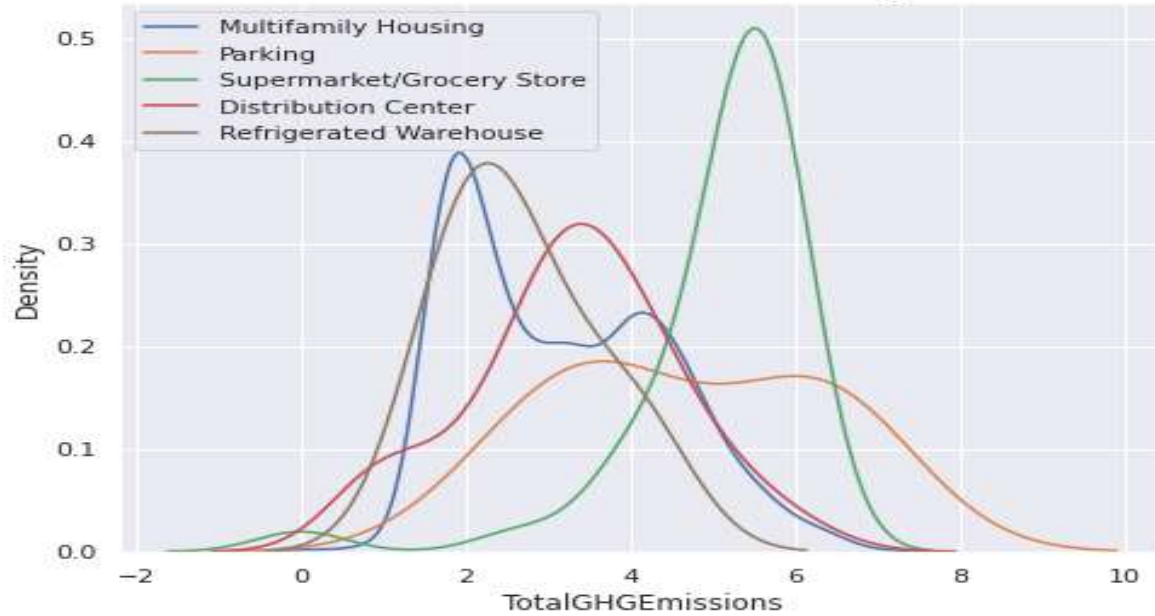


cartographie des bâtiments

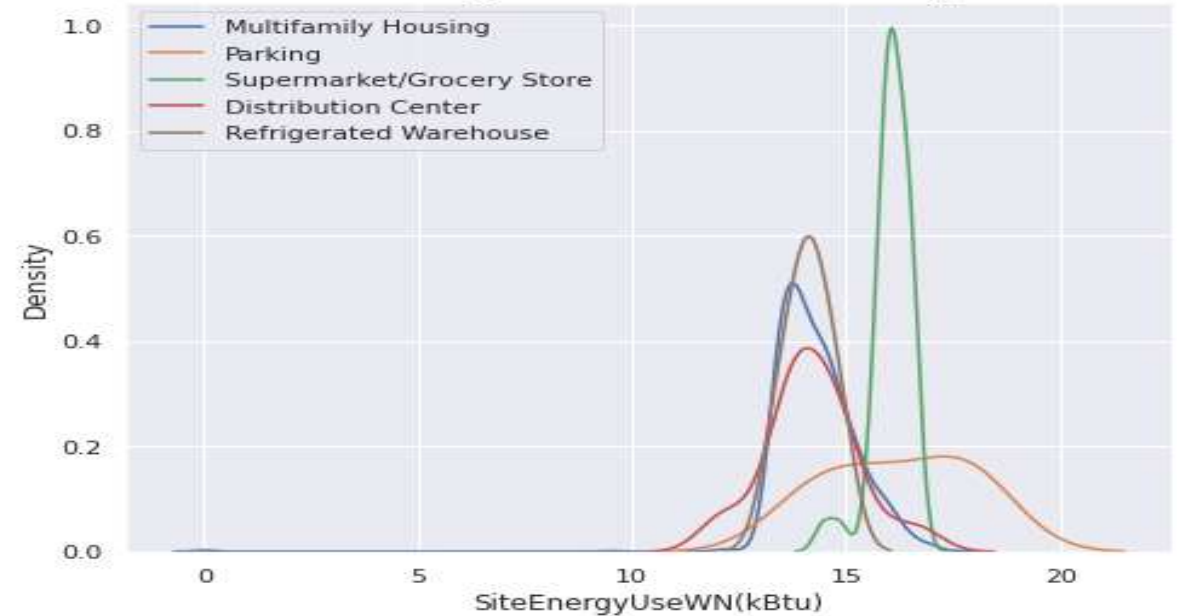


Analyses bivariées

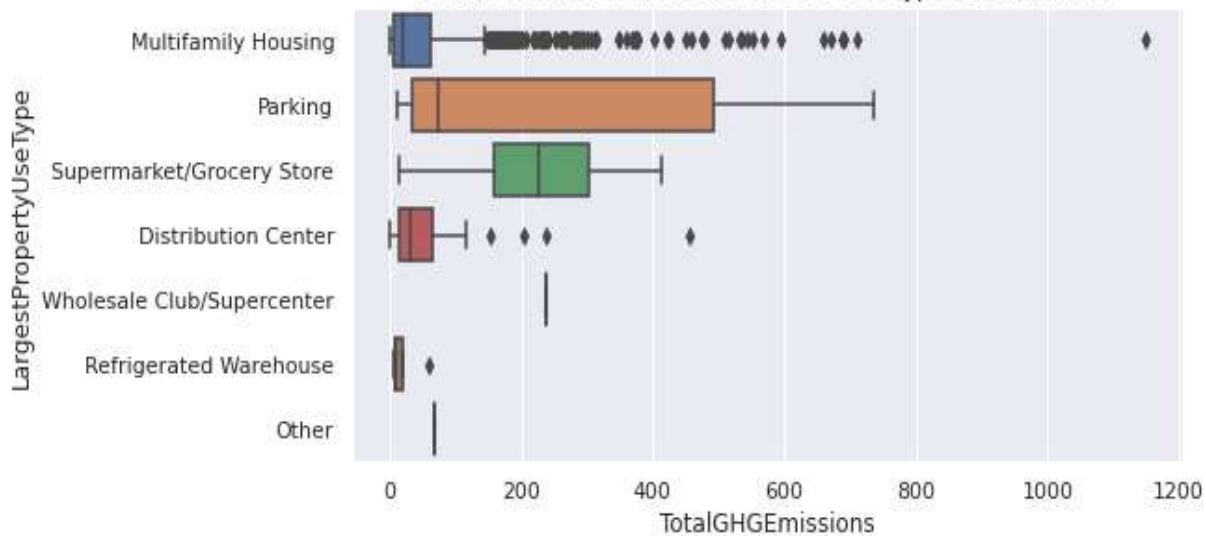
Distribution de TotalGHGEmissions en fonction du type de bâtiment



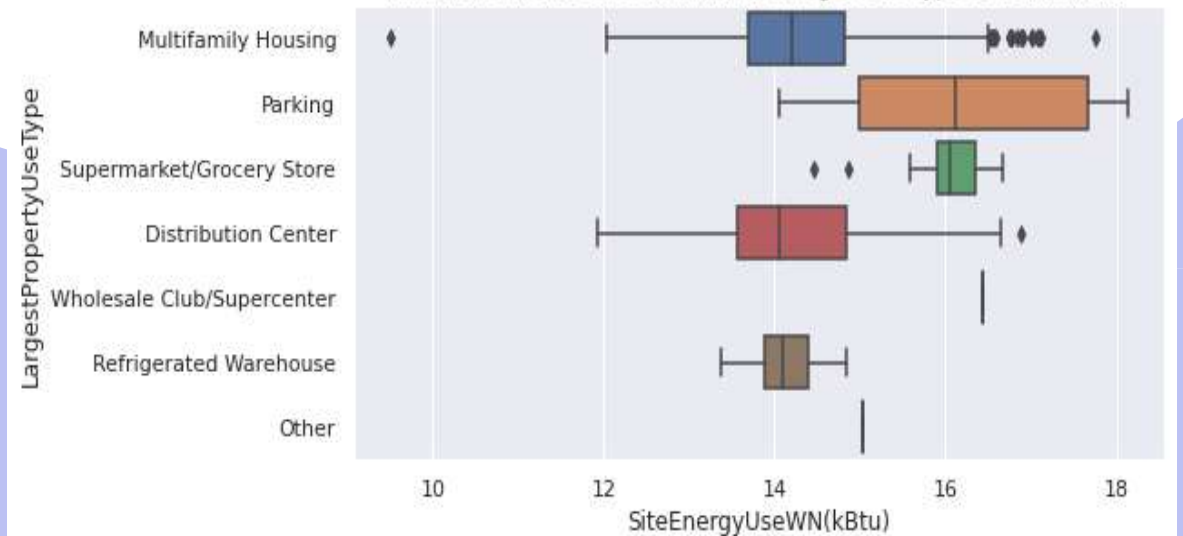
Distribution de SiteEnergyUseWN(kBtu) en fonction du type de bâtiment



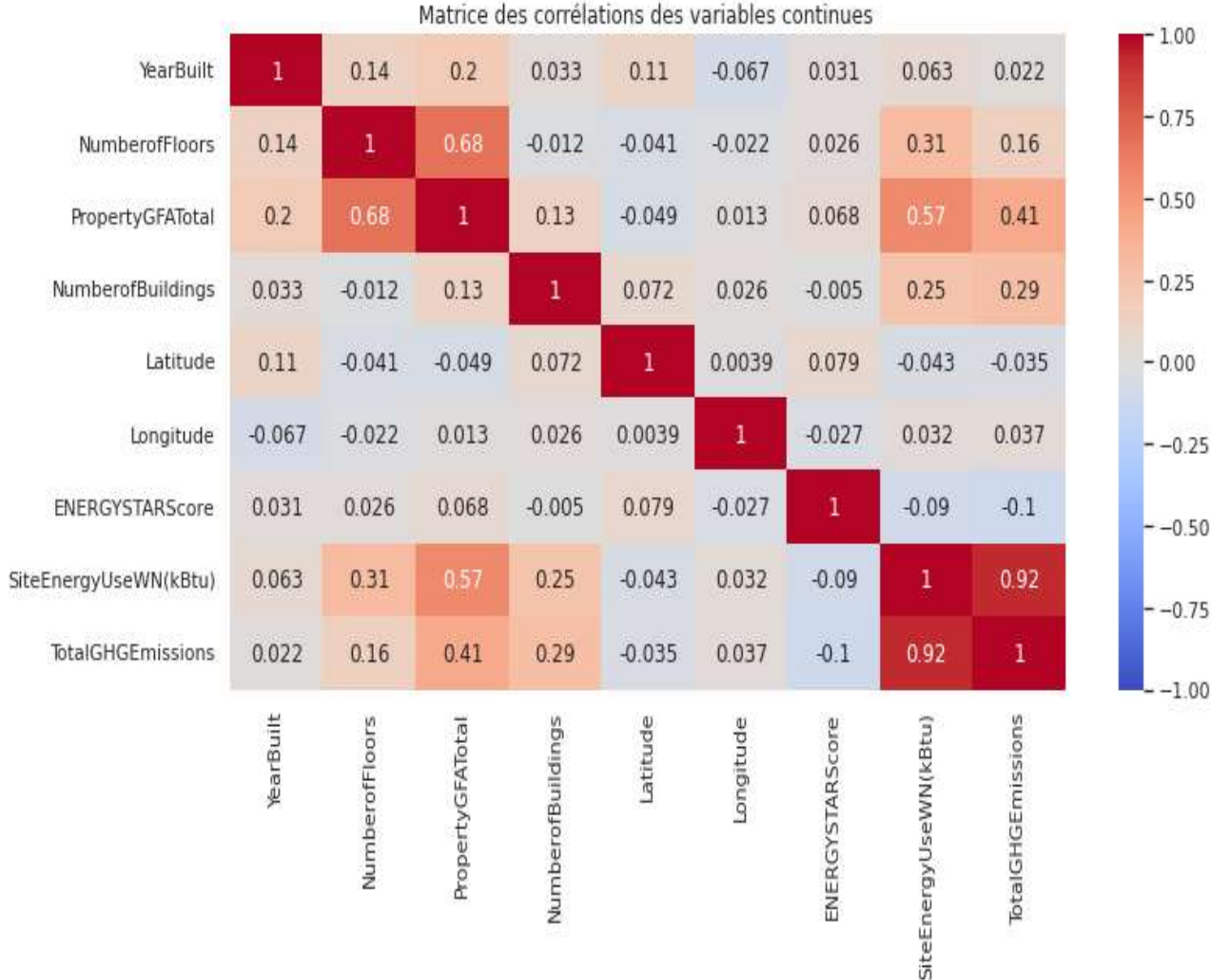
Relation entre les émissions de CO2 et le type de bâtiment



Relation entre la consommation d'énergie et le type de bâtiment



Analyses bivariées



- Consommation d'énergie

Var.	Coef. de corrélation
PropertyGFATotal	0.57
NumberofFloors	0.37
NumberofBuildings	0.25

- Emissions de CO₂

Var.	Coef. de corrélation
PropertyGFATotal	0.41
NumberofFloors	0.16
NumberofBuildings	0.29

03

03/03/20

MODELISATION

Recherche d'un algorithme de
régression



Méthodologie

- Pré-traitement des données :
 - Création de deux jeux de données
 - Séparation des variables catégorielles et numériques
 - Variables catégorielles : *Imputer* → Mode - *OneHotEncoding*
 - Variables numériques : *Imputer* → Médiane - *StandardScaler*
 - Création d'un pipeline de pré-traitement
- Tests de plusieurs modèles de machine learning :
 - Linéaire (ElasticNet, SVR) et ensembliste (Random Forest, XGBoost, Adaboost, etc.)
- Recherche des hyperparamètres les plus optimisés :
 - Grille de recherche + Validation croisée
- Evaluation du modèle :
 - MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), RMPSE (% de RMSE), MAPE (Mean Absolute Percentage Error) & R^2 (Coefficient de détermination)
- Enregistrement dans un pipeline :
 - Pré-traitement + modèle avec hyperparamètres

Prédiction de la consommation d'énergie



Test de plusieurs modèles

- Test de plusieurs modèles
- Avec des hyperparametres
- Utilisation de **Train_Score**, **CV_Score** et le **Time_train** pour choisir un modèle

Modèle	Train_Score	cv_scores_mean_	Time_train
Dummyregressor	0.000000	-0.005198	0.001828
ElasticNet	0.586888	0.493729	62.300573
SVR	0.823467	0.629986	104.970582
Randomforest	0.955960	0.667582	165.141325
Xgboost	0.814295	0.679769	61.886445

- Problème de régression non linéaire
- Modèle Choisit
- **XGBOOST**

Algorithme final : Sans score Energy Star

Hyperparamètres de l'algorithme

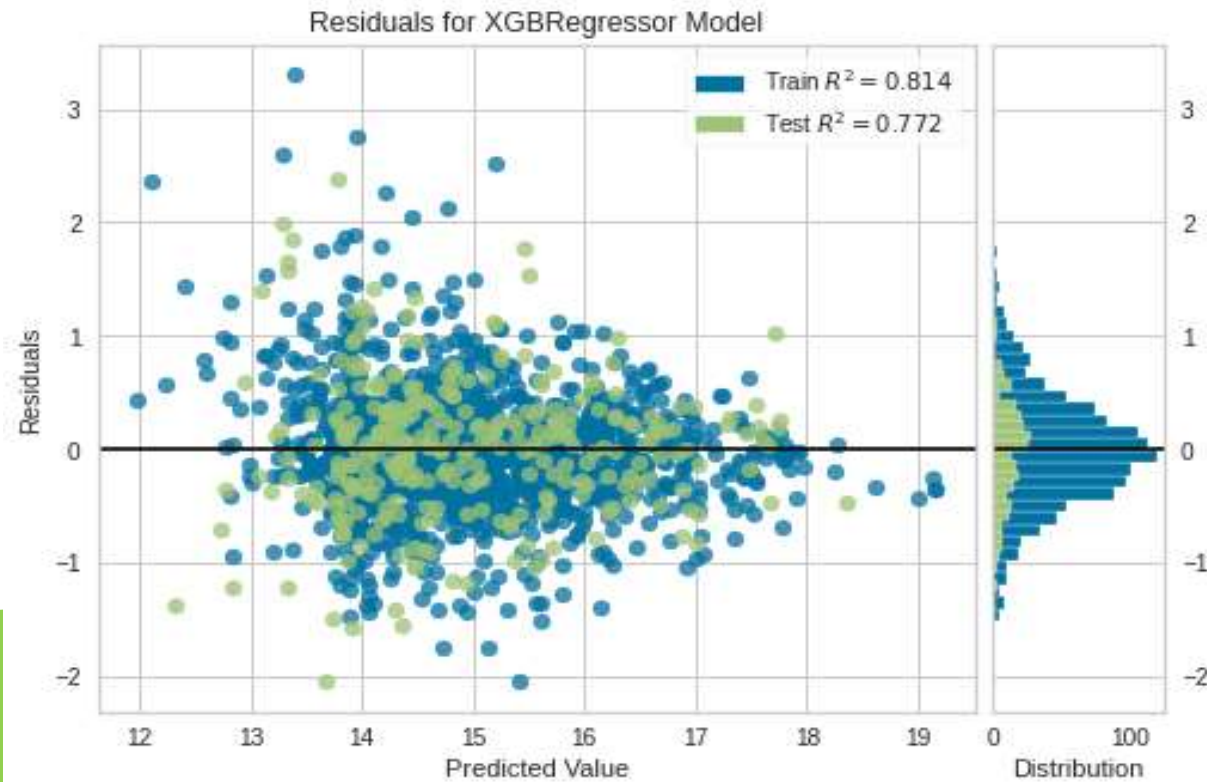
Hyperparamètres	Valeurs
max_features	sqrt
min_samples_leaf	1
n_estimators	500

Feature importances

Features	Score
Social/Meeting Hall	0.17
Supermaker/Grocery Store	0.056
Residential Care Facility	0.056
Library	0.047
Prison/incarceration	0.041

Evaluation du modèle

Analyse des résidus

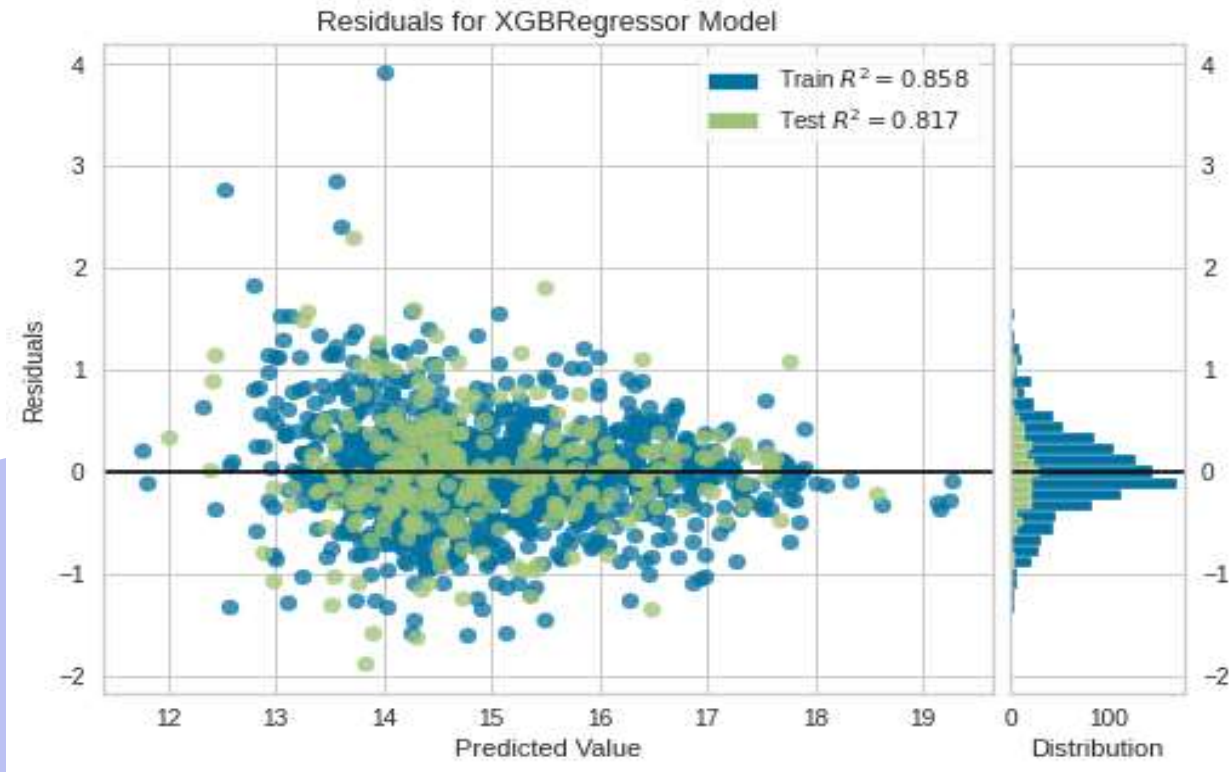


Evaluation des performances

- **MAE** : 0.29
- **MSE** : 0.18 (\log^2)
- **RMSE** : 0.42 (log)
- **RMPSE** : 0.10%
- **MAPE** : 0.02%
- **R^2** : 0.77

Evaluation du modèle : Avec score Energy Star

Analyse des résidus



Evaluation des performances

Metrics	Sans_Energystar_Score	Avec_EnergyStar_Score
MAE	0.457619	0.396760
MSE	0.377220	0.302177
RMSE	0.614183	0.549706
RMPSE	0.117217	0.117600
MAPE	0.031766	0.027462
R2	0.771503	0.816960

- Ajout du score Energy Star :

Amélioration de certaines performances

Prédiction des émissions de CO₂ :

- 1) Sans score Energy Star
- 2) Avec score Energy Star



Sans score Energy Star

- Test de plusieurs modèles
- Avec des hyperparametres
- Utilisation de **Train_Score**, **CV_Score** et le **Time_train** pour choisir un modèle

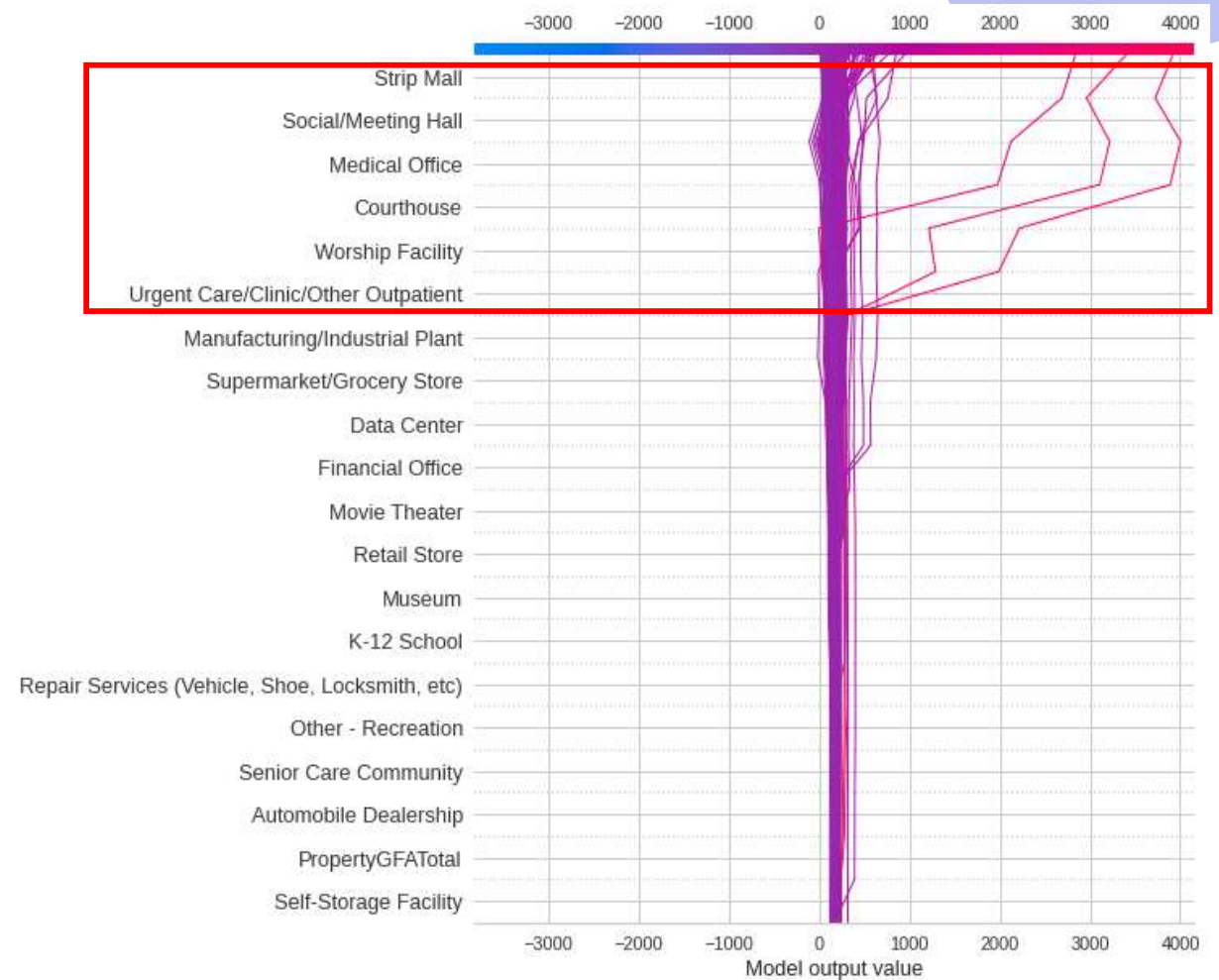
Modèle	Train_Score	cv_scores_mean_	Time_train
Dummyregressor	0.000000	-0.015829	0.004520
ElasticNet	0.369774	-0.012258	105.306370
SVR	0.033236	0.045810	128.915853
Randomforest	0.947423	0.654253	180.560164
Xgboost	0.965571	0.396291	58.278413

- Problème de régression non linéaire
- Choix d'un modèle ensembliste
- **RandomForest**

Algorithme final : Sans score Energy Star

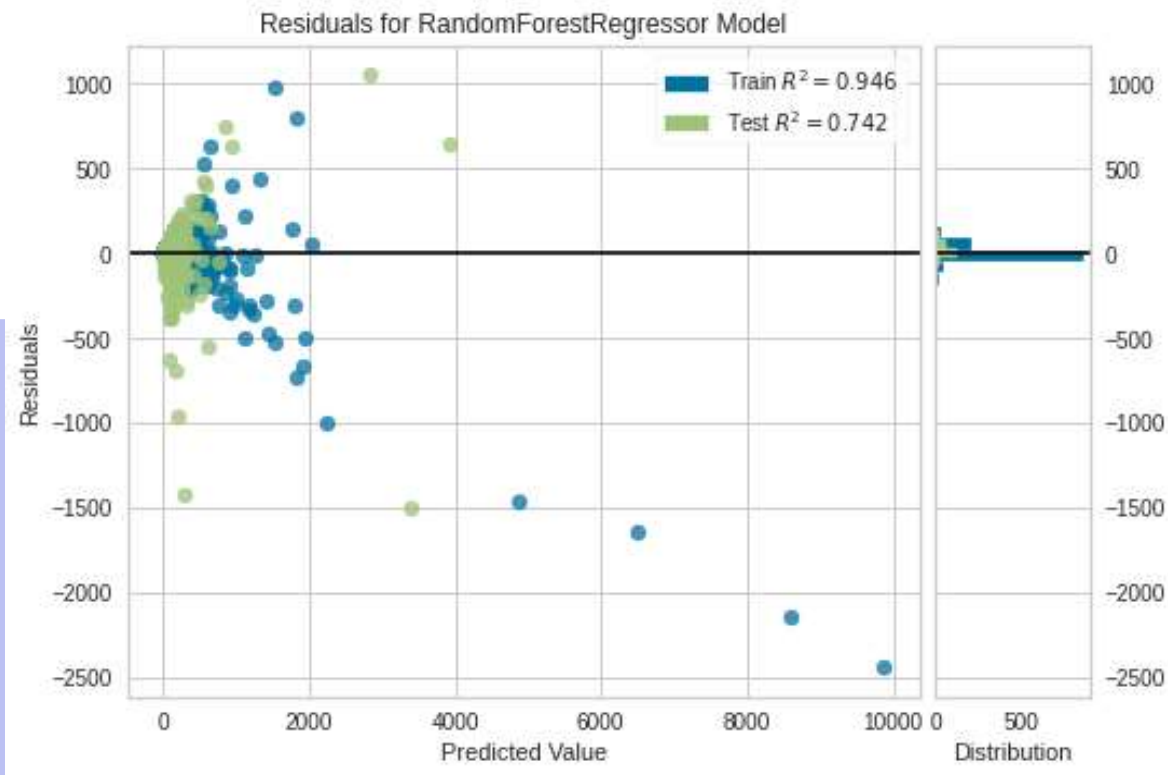
- Feature importances

Features	Score
Courthouse	0.207
Urgence Care/Clinic/Other outpatient	0.200
Strip Mall	0.190
Worship Facility	0.141
Social/Meeting Hall	0.089
Supermarket/Grocery store	0.073



Evaluation du modèle : Sans score Energy Star

Analyse des résidus



Evaluation des performances

- **MAE** : 94.32
- **MSE** : 39523.70
- **RMSE** : 198.81
- **RMPSE** : 85.28%
- **MAPE** : 3.01%
- **R^2** : 0.74

Avec score Energy Star

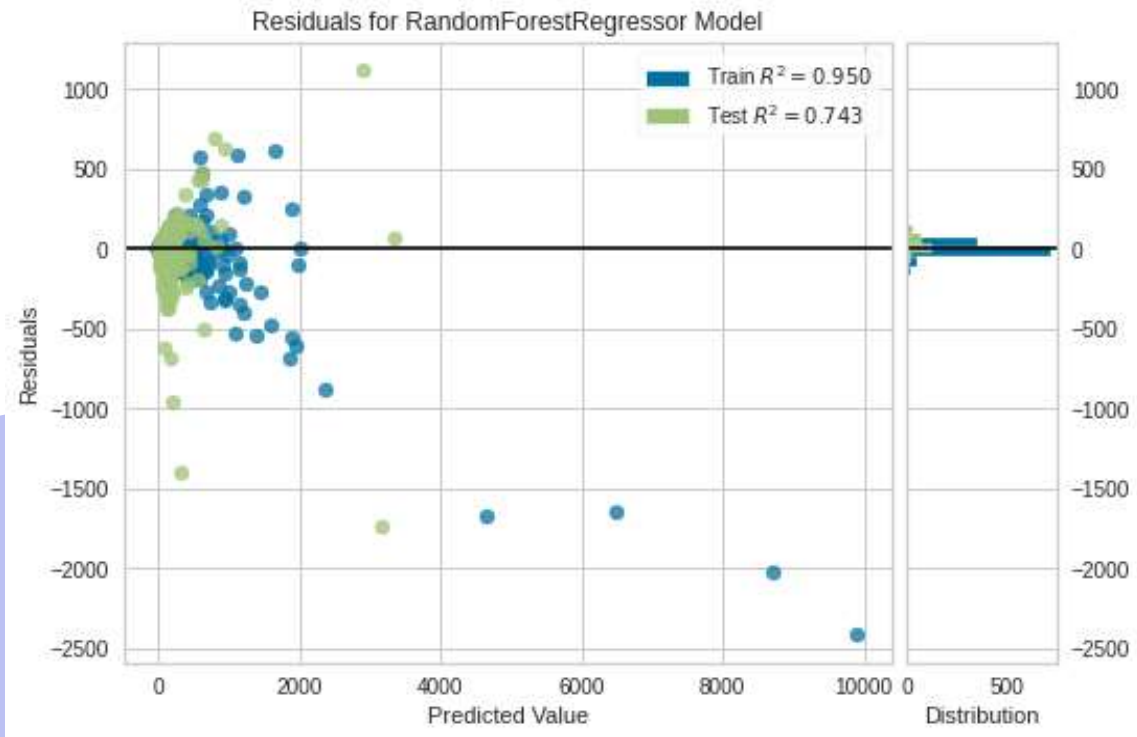
- Test de plusieurs modèles
- Avec des hyperparametres
- Utilisation de **Train_Score**, **CV_Score**, **RMSE** et le **Time_train** pour choisir un modèle

	Modèle	Train_Score	cv_scores_mean_	Time_train
0	Dummyregressor	0.000000	-0.015829	0.001609
1	ElasticNet	0.687900	0.096731	102.289907
2	SVR	0.034863	0.051037	119.195480
3	Randomforest	0.958341	0.644010	195.568701
4	Xgboost	0.969908	0.415196	63.506404

- Problème de régression non linéaire
- Choix d'un modèle ensembliste
- **RandomForest**

Evaluation du modèle : Avec score Energy Star

Analyse des résidus



Evaluation des performances

Metrics	Sans_Energystar_Score	Avec_EnergyStar_Score
MAE	94.320177	90.822672
MSE	39523.696470	39396.408882
RMSE	198.805675	198.485286
RMPSE	85.282324	81.072727
MAPE	3.015336	2.826937
R2	0.742367	0.743197

- Ajout du score Energy Star :

Amélioration de certaines performances

04

CONCLUSION



Choix du modèle à déployer

Concernant la consommation d'énergie :

- Informations suffisantes pour la mise en place d'une prédiction
- Modèle avec de bonnes performances (77% de variance expliquée, 0.1% d'erreur dans les prédictions)

Concernant les émissions de CO2 :

- Ajout du score Energy Star : Amélioration de certaines performances
- Beaucoup de valeurs manquantes → Imputation par la médiane
- Peu d'intérêt d'utiliser le score Energy Star
- Performances moins bonnes que pour la consommation d'énergie
- Manque d'informations pour construire ce modèle ?





CentraleSupélec

Merci pour votre attention !