

Projet 4 : Anticipez les besoins en consommation électrique de bâtiments

Lancelot LECLERCQ

15 décembre 2021

Sommaire

1. Introduction
2. Nettoyage du jeu de données
3. Étapes des modélisations
4. Modélisation des émissions de carbone
5. Modélisation de la consommation énergétique
6. Conclusion

Introduction

Problématique

- Objectif de la ville de Seattle : atteindre la neutralité en émissions de carbone
- La ville s'intéresse aux émissions des bâtiments non destinés à l'habitation
- Pour cela des relevés de consommation ont été réalisés mais ils sont coûteux à obtenir
- Est-il possible de prédire les émissions et de la consommation d'énergie pour des bâtiments pour lesquels les relevés n'ont pas été réalisés à partir des relevés déjà obtenus



Seattle

Jeu de données

- Base de données issue de l'initiative de la ville de Seattle de proposer ses données en accès libre (Open Data)
- Données concernant les batiments de la ville, caractérise :
 - le type,
 - la surface,
 - le nombre d'étages,
 - la consommation énergétique,
 - les émissions de carbone,
 - :
- Données des années 2015 et 2016

Nettoyage du jeu de données

Nettoyage du jeu de données

Étapes des modélisations

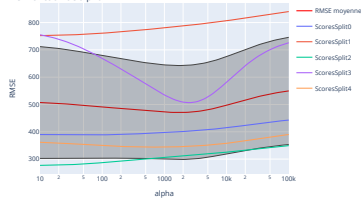
Étapes des modélisations

Modélisation émissions

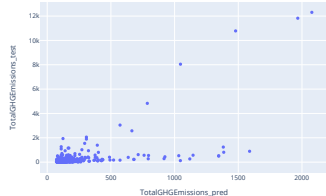
Modèle Ridge

Variable non modifiée

RMSE du modèle Ridge
pour la variable TotalGHGEmissions
en fonction de alpha



Visualisation des données de TotalGHGEmissions
prédites par le modèle Ridge()
vs les données test



←

R ²	RMSE	MAE	MAE%	FitTime(s)
0.25	784.53	192.35	6.72	0.01

paramètre	Ridge()
ridge_alpha	1668.10

←

R ²	RMSE	MAE	MAE%	FitTime(s)
0.17	888.31	182.94	2.81	0.02

paramètre Ridge()

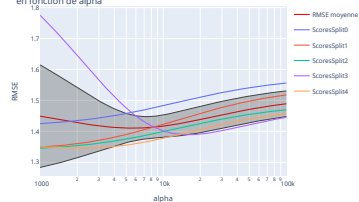
ridge_alpha 5590.81

⇒

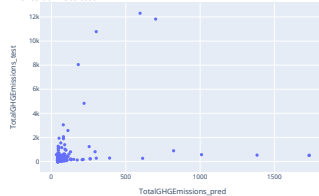
⇒

Variable au log

RMSE du modèle Ridge
pour la variable TotalGHGEmissions_log
en fonction de alpha



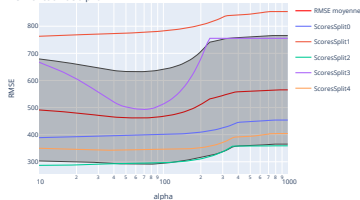
Visualisation des données de TotalGHGEmissions_log
prédites par le modèle Ridge()
vs les données test



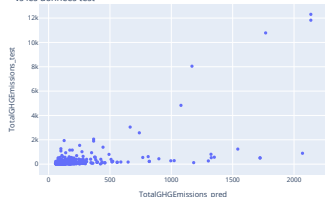
Modèle Lasso

Variable non modifiée

RMSE du modèle Lasso
pour la variable TotalGHGEmissions
en fonction de alpha



Visualisation des données de TotalGHGEmissions
prédites par le modèle Lasso()
vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.27	774.57	193.79	6.03	0.01

paramètre	Lasso()
lasso_alpha	64.28



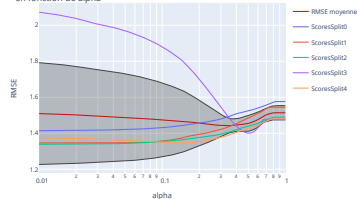
R ²	RMSE	MAE	MAE%	FitTime(s)
0.13	900.77	182.60	2.99	0.02

paramètre	Lasso()
lasso_alpha	0.34

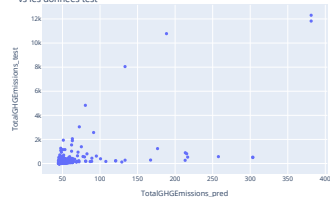


Variable au log

RMSE du modèle Lasso
pour la variable TotalGHGEmissions_log
en fonction de alpha



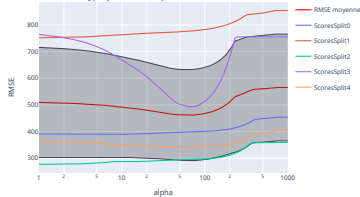
Visualisation des données de TotalGHGEmissions_log
prédites par le modèle Lasso()
vs les données test



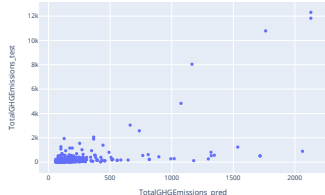
Modèle ElasticNet

Variable non modifiée

RMSE du modèle ElasticNet pour la variable TotalGHGEmissions avec le paramètre l1_ratio=1.0 en fonction de l'hyperparamètre alpha



Visualisation des données de TotalGHGEmissions prédites par le modèle ElasticNet() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.27	774.96	193.83	6.05	0.02

paramètre	ElasticNet()
elasticnet_alpha	65.79
elasticnet_l1_ratio	1.00



R ²	RMSE	MAE	MAE%	FitTime(s)
0.17	893.72	180.94	2.84	0.02

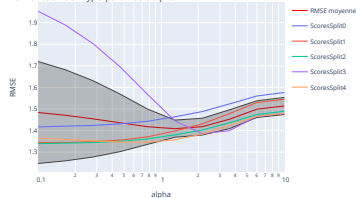
paramètre ElasticNet()

elasticnet_alpha	1.29
elasticnet_l1_ratio	0.10

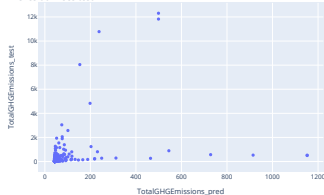


Variable au log

RMSE du modèle ElasticNet pour la variable TotalGHGEmissions_log avec le paramètre l1_ratio=0.1 en fonction de l'hyperparamètre alpha



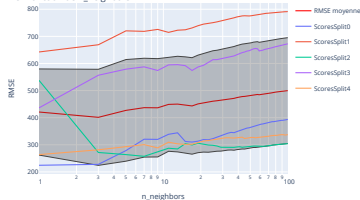
Visualisation des données de TotalGHGEmissions_log prédites par le modèle ElasticNet() vs les données test



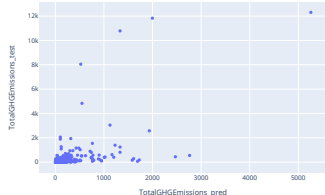
Modèle kNeighborsRegressor

Variable non modifiée

RMSE du modèle KNeighborsRegressor pour la variable TotalGHGEmissions en fonction de n_neighbors



Visualisation des données de TotalGHGEmissions prédites par le modèle KNeighborsRegressor() vs les données test



←

R ²	RMSE	MAE	MAE%	FitTime(s)
0.32	749.39	175.54	2.97	0.02

paramètre KNeighborsRegressor()

kneighborsregressor__n_neighbors 3

←

R ²	RMSE	MAE	MAE%	FitTime(s)
0.59	662.80	118.44	1.26	0.02

paramètre KNeighborsRegressor()

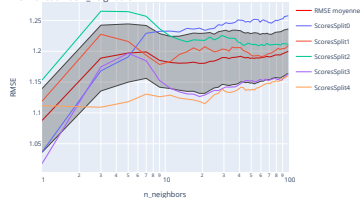
kneighborsregressor__n_neighbors 1

⇒

⇒

Variable au log

RMSE du modèle KNeighborsRegressor pour la variable TotalGHGEmissions_log en fonction de n_neighbors



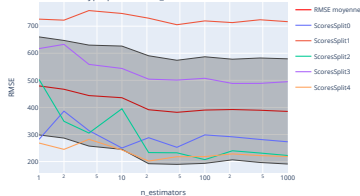
Visualisation des données de TotalGHGEmissions_log prédites par le modèle KNeighborsRegressor() vs les données test



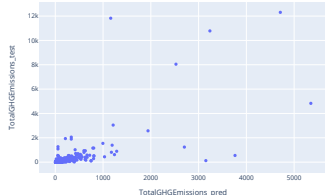
Modèle RandomForestRegressor

Variable non modifiée

RMSE du modèle RandomForestRegressor pour la variable TotalGHGEmissions avec le paramètre max_features=log2 en fonction de l'hyperparamètre n_estimators



Visualisation des données de TotalGHGEmissions prédites par le modèle RandomForestRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.42	687.32	135.01	2.24	0.25

paramètre	RandomForestRegres
randomforestregressor__n_estimators	46
randomforestregressor__max_features	log2



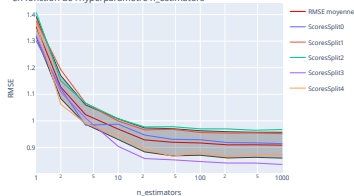
R ²	RMSE	MAE	MAE%	FitTime(s)
0.66	738.02	132.49	1.04	4.42

paramètre	RandomForestRegres
randomforestregressor__n_estimators	1000
randomforestregressor__max_features	log2

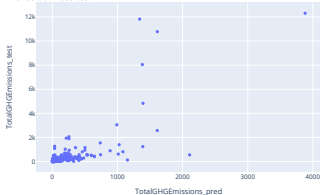


Variable au log

RMSE du modèle RandomForestRegressor pour la variable TotalGHGEmissions_log avec le paramètre max_features=log2 en fonction de l'hyperparamètre n_estimators



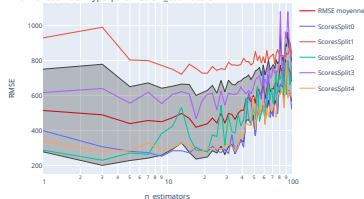
Visualisation des données de TotalGHGEmissions_log prédites par le modèle RandomForestRegressor() vs les données test



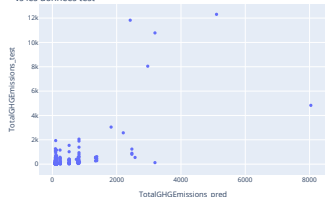
Modèle AdaBoostRegressor

Variable non modifiée

RMSE du modèle AdaBoostRegressor pour la variable TotalGHGEmissions avec le paramètre loss=exponential en fonction de l'hyperparamètre n_estimators



Visualisation des données de TotalGHGEmissions prédites par le modèle AdaBoostRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.44	679.78	210.35	6.90	0.11

paramètre	AdaBoostRegressor()
adaboostregressor__n_estimators	17
adaboostregressor__loss	exponential



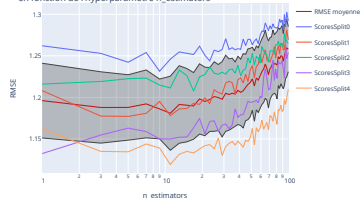
R ²	RMSE	MAE	MAE%	FitTime(s)
0.34	793.85	168.49	1.64	0.09

paramètre	AdaBoostRegressor()
adaboostregressor__n_estimators	11
adaboostregressor__loss	exponential

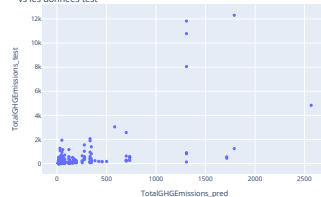


Variable au log

RMSE du modèle AdaBoostRegressor pour la variable TotalGHGEmissions_log avec le paramètre loss=exponential en fonction de l'hyperparamètre n_estimators



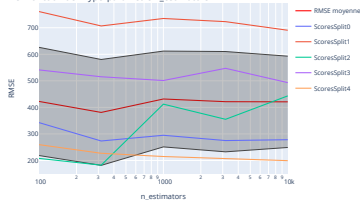
Visualisation des données de TotalGHGEmissions_log prédites par le modèle AdaBoostRegressor() vs les données test



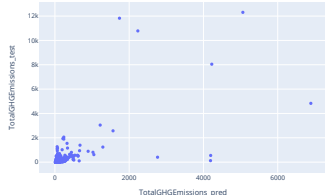
Modèle GradientBoostingRegressor

Variable non modifiée

RMSE du modèle GradientBoostingRegressor pour la variable TotalGHGEmissions avec le paramètre loss=huber en fonction de l'hyperparamètre n_estimators



Visualisation des données de TotalGHGEmissions prédites par le modèle GradientBoostingRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.41	694.86	153.06	3.02	2.81

paramètre	GradientBoostingRe
gradientboostingregressor__n_estimators	316
gradientboostingregressor__loss	huber



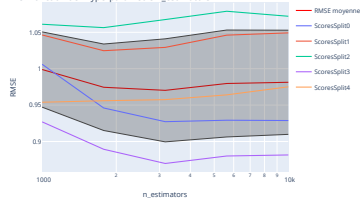
R ²	RMSE	MAE	MAE%	FitTime(s)
0.68	650.17	117.39	0.95	11.09

paramètre	GradientBoostingRe
gradientboostingregressor__n_estimators	3162
gradientboostingregressor__loss	squared_error

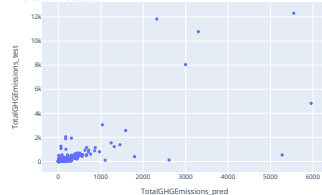


Variable au log

RMSE du modèle GradientBoostingRegressor pour la variable TotalGHGEmissions_log avec le paramètre loss=squared_error en fonction de l'hyperparamètre n_estimators

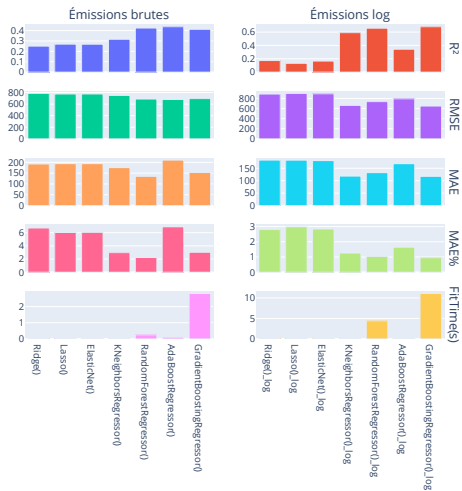


Visualisation des données de TotalGHGEmissions_log prédites par le modèle GradientBoostingRegressor() vs les données test



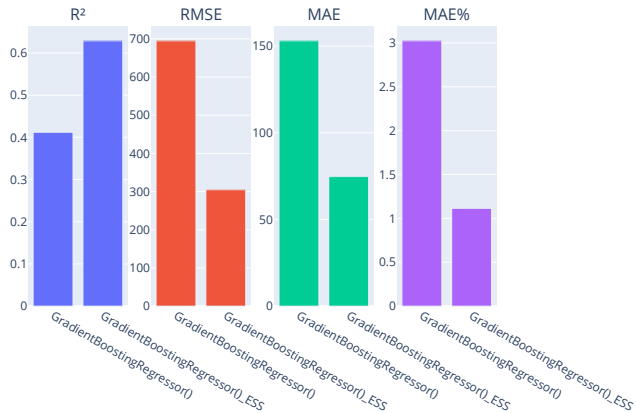
Comparaison des résultats selon que la variable est au log ou non

Comparaison des scores des modèles d'émissions



Influence de l'EnergyStar score sur la prédiction des Émissions

Comparaison avec et sans ajout de l'energy score stars

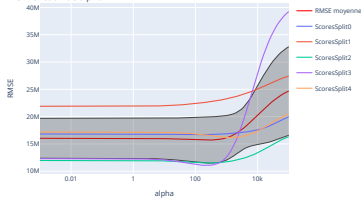


Modélisation consommation

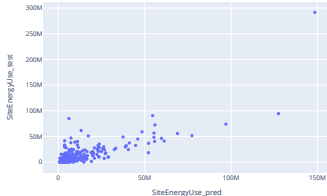
Modèle Ridge

Variable non modifiée

RMSE du modèle Ridge
pour la variable SiteEnergyUse
en fonction de alpha

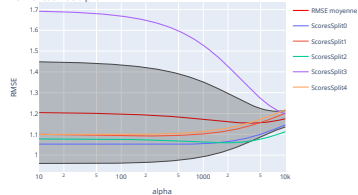


Visualisation des données de SiteEnergyUse
prédites par le modèle Ridge()
vs les données test

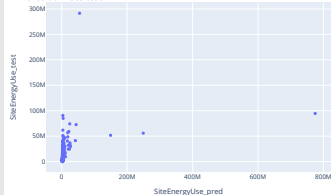


Variable au log

RMSE du modèle Ridge
pour la variable SiteEnergyUse_log
en fonction de alpha



Visualisation des données de SiteEnergyUse_log
prédites par le modèle Ridge()
vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.66	9498871.94	4216742.65	1.66	0.01
paramètre		Ridge()		
ridge_alpha	215.44			



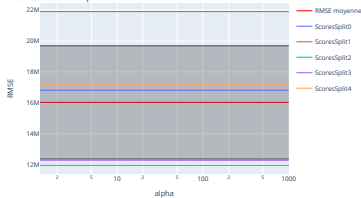
R ²	RMSE	MAE	MAE%	FitTime(s)
0.30	31314930.87	6430474.76	1.28	0.0
paramètre		Ridge()		
ridge_alpha	3764.94			



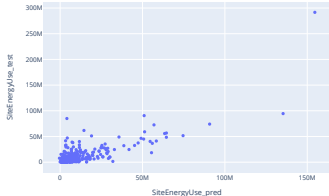
Modèle Lasso

Variable non modifiée

RMSE du modèle Lasso
pour la variable SiteEnergyUse
en fonction de alpha

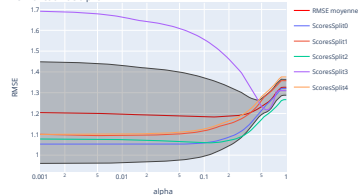


Visualisation des données de SiteEnergyUse
prédites par le modèle Lasso()
vs les données test

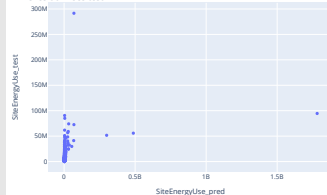


Variable au log

RMSE du modèle Lasso
pour la variable SiteEnergyUse_log
en fonction de alpha



Visualisation des données de SiteEnergyUse_log
prédites par le modèle Lasso()
vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.64	9740467.35	4523527.68	1.74	0.03
paramètre Lasso()				
lasso_alpha 1000.00				



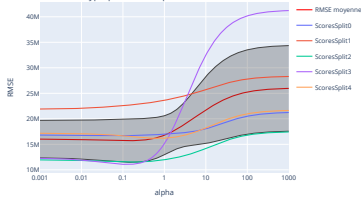
R ²	RMSE	MAE	MAE%	FitTime(s)
0.32	70636736.23	8470585.69	1.26	0.03
paramètre Lasso()				
lasso_alpha 0.11				



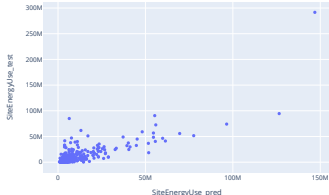
Modèle ElasticNet

Variable non modifiée

RMSE du modèle ElasticNet pour la variable SiteEnergyUse avec le paramètre l1_ratio=0.1 en fonction de l'hyperparamètre alpha

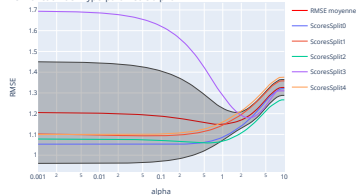


Visualisation des données de SiteEnergyUse prédites par le modèle ElasticNet() vs les données test

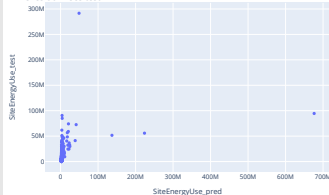


Variable au log

RMSE du modèle ElasticNet pour la variable SiteEnergyUse_log avec le paramètre l1_ratio=0.1 en fonction de l'hyperparamètre alpha



Visualisation des données de SiteEnergyUse_log prédites par le modèle ElasticNet() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.66	9502168.63	4200468.92	1.66	0.00

paramètre	ElasticNet()
elasticnet__alpha	0.12
elasticnet__l1_ratio	0.10



R ²	RMSE	MAE	MAE%	FitTime(s)
0.30	27831759.90	6264806.68	1.28	0.00

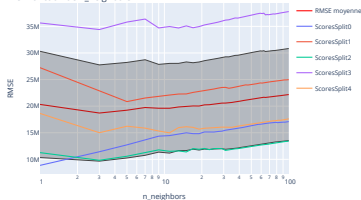
paramètre	ElasticNet()
elasticnet__alpha	0.89
elasticnet__l1_ratio	0.10



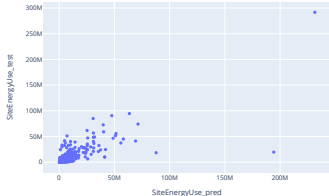
Modèle kNeighborsRegressor

Variable non modifiée

RMSE du modèle KNeighborsRegressor pour la variable SiteEnergyUse en fonction de n_neighbors



Visualisation des données de SiteEnergyUse prédites par le modèle KNeighborsRegressor() vs les données test

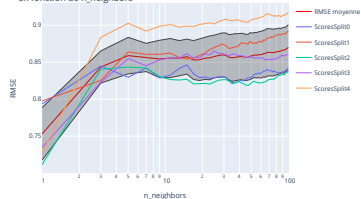


R ²	RMSE	MAE	MAE%	FitTime(s)
0.58	10545930.52	3852872.12	0.96	0.02
paramètre KNeighborsRegressor()				
kneighborsregressor__n_neighbors				
3				
paramètre KNeighborsRegressor()				
kneighborsregressor__n_neighbors				
1				
R ²	RMSE	MAE	MAE%	FitTime(s)
0.76	6523063.66	2125360.91	0.40	0.02

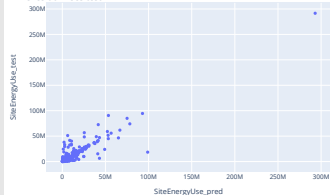


Variable au log

RMSE du modèle KNeighborsRegressor pour la variable SiteEnergyUse_log en fonction de n_neighbors



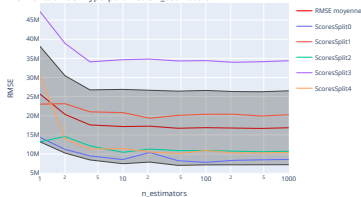
Visualisation des données de SiteEnergyUse_log prédites par le modèle KNeighborsRegressor() vs les données test



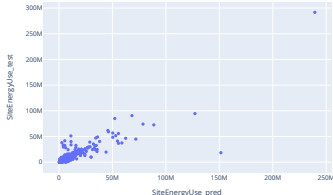
Modèle RandomForestRegressor

Variable non modifiée

RMSE du modèle RandomForestRegressor pour la variable SiteEnergyUse avec le paramètre max_features=auto en fonction de l'hyperparamètre n_estimators



Visualisation des données de SiteEnergyUse prédites par le modèle RandomForestRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.77	7897252.92	2655419.73	0.65	5.50

paramètre		RandomForestRegressor
randomforestregressor__n_estimators	464	
randomforestregressor__max_features	auto	



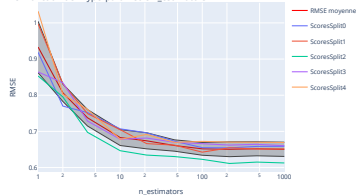
R ²	RMSE	MAE	MAE%	FitTime(s)
0.81	6699738.26	2444802.20	0.43	5.90

paramètre		RandomForestRegressor
randomforestregressor__n_estimators	1000	
randomforestregressor__max_features	sqrt	

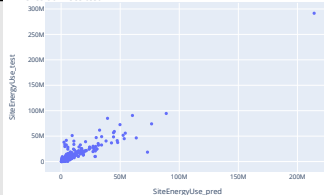


Variable au log

RMSE du modèle RandomForestRegressor pour la variable SiteEnergyUse_log avec le paramètre max_features=sqrt en fonction de l'hyperparamètre n_estimators



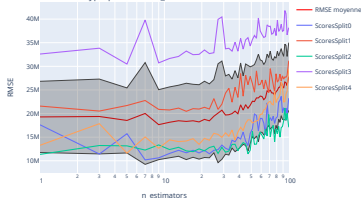
Visualisation des données de SiteEnergyUse_log prédites par le modèle RandomForestRegressor() vs les données test



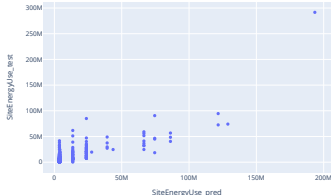
Modèle AdaBoostRegressor

Variable non modifiée

RMSE du modèle AdaBoostRegressor pour la variable SiteEnergyUse avec le paramètre loss=exponential en fonction de l'hyperparamètre n_estimators

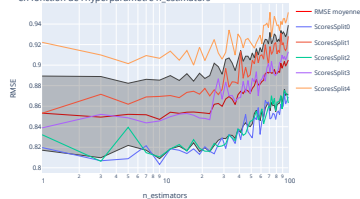


Visualisation des données de SiteEnergyUse prédites par le modèle AdaBoostRegressor() vs les données test

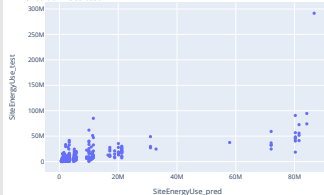


Variable au log

RMSE du modèle AdaBoostRegressor pour la variable SiteEnergyUse_log avec le paramètre loss=exponential en fonction de l'hyperparamètre n_estimators



Visualisation des données de SiteEnergyUse_log prédites par le modèle AdaBoostRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.66	9577498.02	4797313.47	2.14	0.06

paramètre	AdaBoostRegressor()
adaboostregressor__n_estimators	9
adaboostregressor__loss	exponential



R ²	RMSE	MAE	MAE%	FitTime(s)
0.54	11946475.76	4171710.66	0.80	0.06

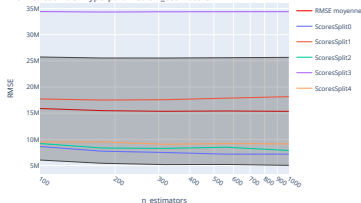
paramètre	AdaBoostRegressor()
adaboostregressor__n_estimators	9
adaboostregressor__loss	exponential



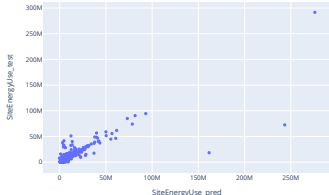
Modèle GradientBoostingRegressor

Variable non modifiée

RMSE du modèle GradientBoostingRegressor pour la variable SiteEnergyUse avec le paramètre loss=squared_error en fonction de l'hyperparamètre n_estimators



Visualisation des données de SiteEnergyUse prédites par le modèle GradientBoostingRegressor() vs les données test



R ²	RMSE	MAE	MAE%	FitTime(s)
0.62	10038705.95	2819213.14	0.89	3.2
paramètre		GradientBoostingRegressor		
gradientboostingregressor__n_estimators		1000		
gradientboostingregressor__loss		squared_error		

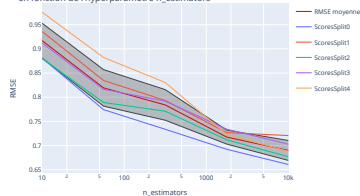


R ²	RMSE	MAE	MAE%	FitTime(s)
0.79	5284322.90	1713173.55	0.37	77.9
paramètre		GradientBoostingRegressor		
gradientboostingregressor__n_estimators		10000		
gradientboostingregressor__loss		huber		

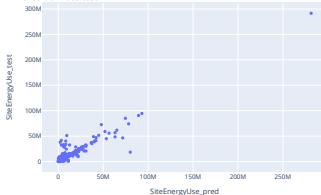


Variable au log

RMSE du modèle GradientBoostingRegressor pour la variable SiteEnergyUse_log avec le paramètre loss=huber en fonction de l'hyperparamètre n_estimators

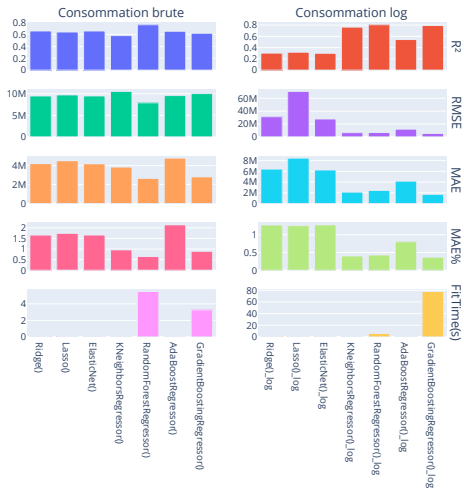


Visualisation des données de SiteEnergyUse_log prédites par le modèle GradientBoostingRegressor() vs les données test



Comparaison des résultats selon que la variable est au log ou non

Comparaison des scores des modèles de consommation



Conclusion

Conclusion