

## 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



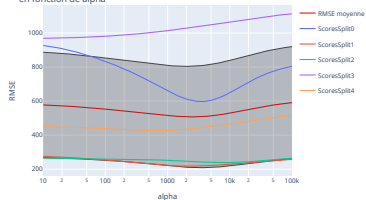
# Étapes des modélisations

# Modélisation émissions

# Modèle Ridge

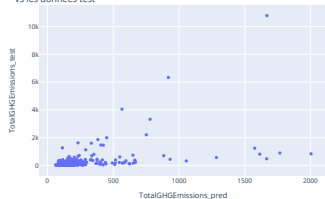
## Variable non modifiée

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



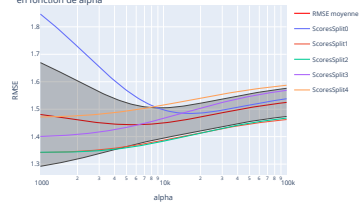
Loading [MathJax]extensions/MathMenu.js

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

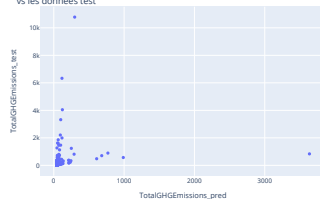


## Variable au log

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



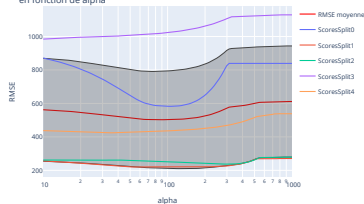
Visualisation des données de TotalGHGEmissions\_log  
prédites par le modèle Ridge()  
vs les données test



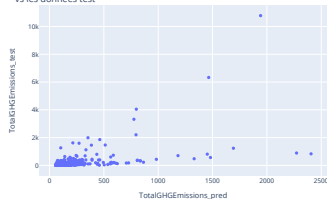
# Modèle Lasso

## Variable non modifiée

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

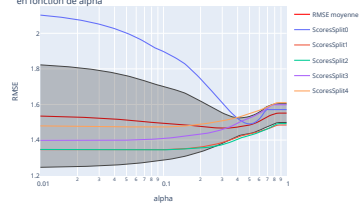


Visualisation des données de TotalGHGEmissions  
prédites par le modèle Lasso()  
vs les données test

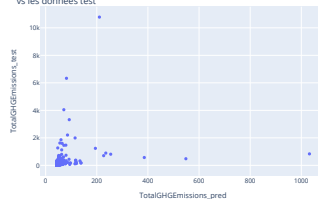


## Variable au log

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



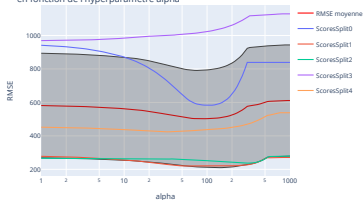
Visualisation des données de TotalGHGEmissions\_log  
prédites par le modèle Lasso()  
vs les données test



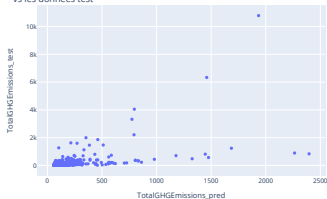
# Modèle ElasticNet

## Variable non modifiée

RMSE du modèle ElasticNet pour la variable TotalGHGEmissions avec le paramètre l1\_ratio=1.0 en fonction de l'hyperparamètre alpha

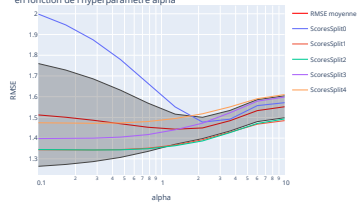


Visualisation des données de TotalGHGEmissions prédites par le modèle ElasticNet() vs les données test

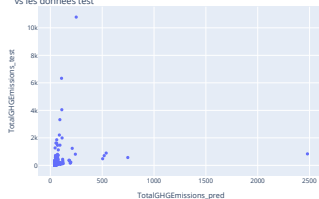


## Variable au log

RMSE du modèle ElasticNet pour la variable TotalGHGEmissions\_log avec le paramètre l1\_ratio=0.1 en fonction de l'hyperparamètre alpha



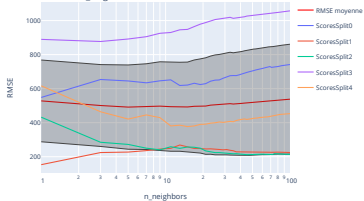
Visualisation des données de TotalGHGEmissions\_log prédites par le modèle ElasticNet() vs les données test



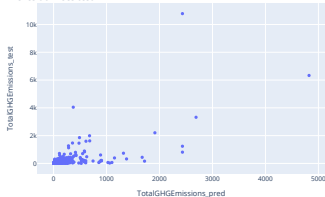
# Modèle kNeighborsRegressor

## Variable non modifiée

RMSE du modèle KNeighborsRegressor  
pour la variable TotalGHGEmissions  
en fonction de n\_neighbors

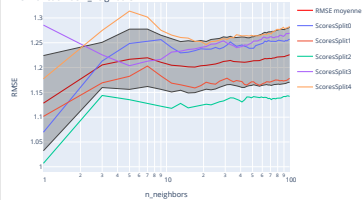


Visualisation des données de TotalGHGEmissions  
prédites par le modèle KNeighborsRegressor()  
vs les données test

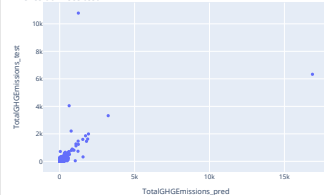


## Variable au log

RMSE du modèle KNeighborsRegressor  
pour la variable TotalGHGEmissions\_log  
en fonction de n\_neighbors



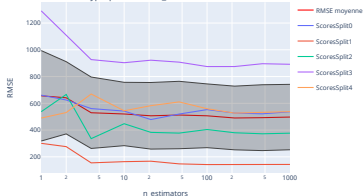
Visualisation des données de TotalGHGEmissions\_log  
prédites par le modèle KNeighborsRegressor()  
vs les données test



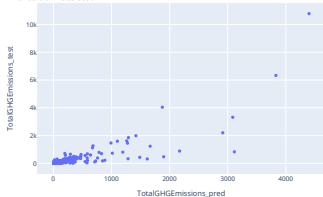
# Modèle RandomForestRegressor

## Variable non modifiée

RMSE du modèle RandomForestRegressor pour la variable TotalGHGEmissions avec le paramètre max\_features=sqrt en fonction de l'hyperparamètre n\_estimators

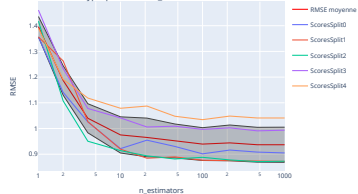


Visualisation des données de TotalGHGEmissions prédites par le modèle RandomForestRegressor() vs les données test

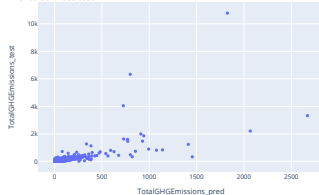


## Variable au log

RMSE du modèle RandomForestRegressor pour la variable TotalGHGEmissions\_log avec le paramètre max\_features=log2 en fonction de l'hyperparamètre n\_estimators



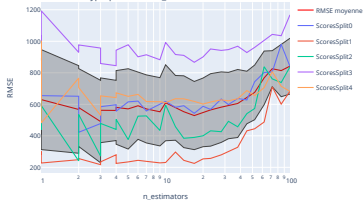
Visualisation des données de TotalGHGEmissions\_log prédites par le modèle RandomForestRegressor() vs les données test



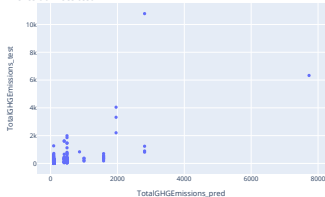
# Modèle AdaBoostRegressor

## Variable non modifiée

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

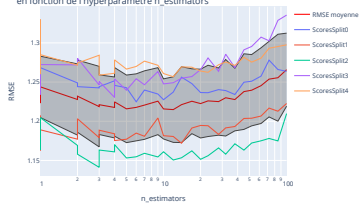


Visualisation des données de TotalGHGEmissions prédites par le modèle AdaBoostRegressor() vs les données test

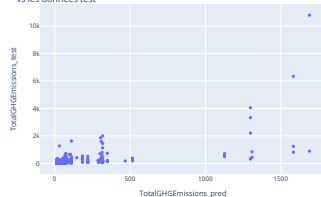


## Variable au log

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

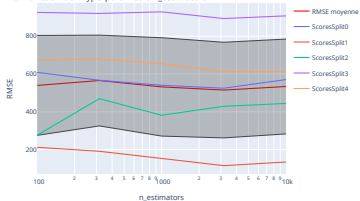




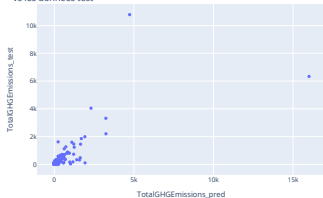
# Modèle GradientBoostingRegressor

## Variable non modifiée

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

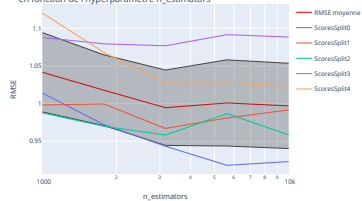


Visualisation des données de TotalGHGEmissions prédites par le modèle GradientBoostingRegressor() vs les données test

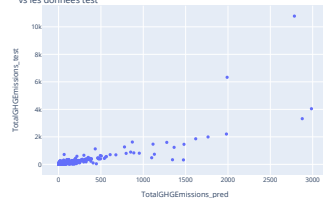


## Variable au log

RMSE du modèle GradientBoostingRegressor pour la variable TotalGHGEmissions\_log avec le paramètre loss=huber 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

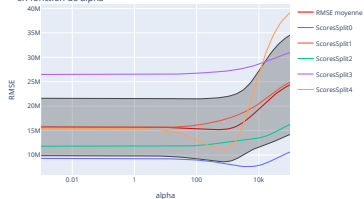
# Influence de l'EnergyStar score sur la prédiction des Émissions

# Modélisation consommation

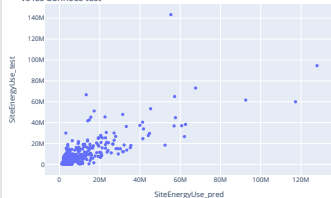
# Modèle Ridge

## Variable non modifiée

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

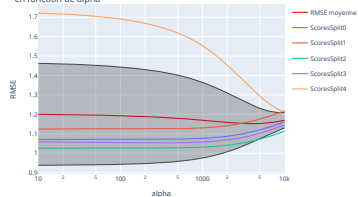


Visualisation des données de SiteEnergyUse  
prédites par le modèle Ridge()  
vs les données test

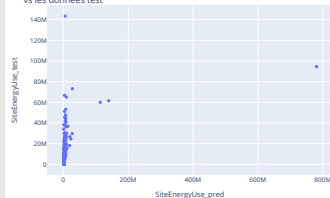


## Variable au log

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



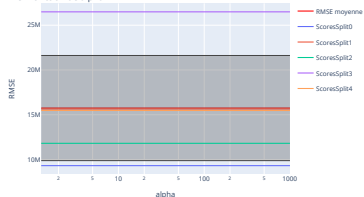
Visualisation des données de SiteEnergyUse\_log  
prédites par le modèle Ridge()  
vs les données test



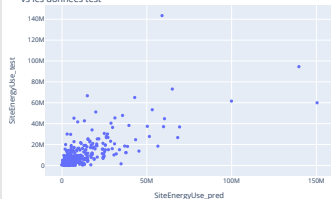
# Modèle Lasso

## Variable non modifiée

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

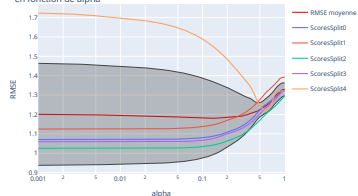


Visualisation des données de SiteEnergyUse  
prédites par le modèle Lasso()  
vs les données test

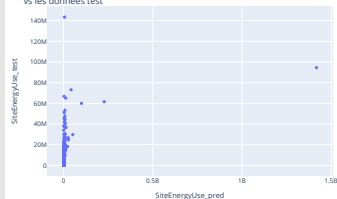


## Variable au log

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



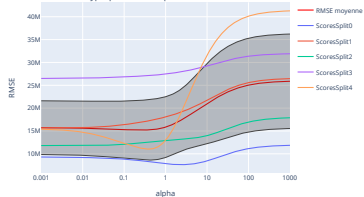
Visualisation des données de SiteEnergyUse\_log  
prédites par le modèle Lasso()  
vs les données test



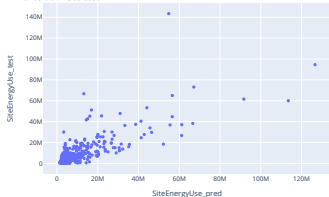
# Modèle ElasticNet

## Variable non modifiée

RMSE du modèle ElasticNet pour la variable SiteEnergyUse avec le paramètre l1\_ratio=0.1 en fonction de l'hyperparamètre alpha

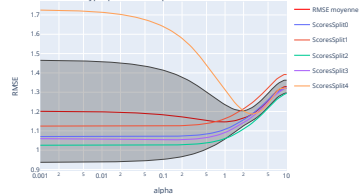


Visualisation des données de SiteEnergyUse prédites par le modèle ElasticNet() vs les données test

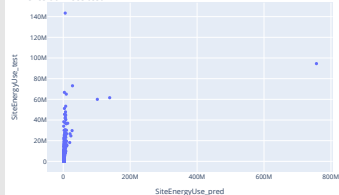


## Variable au log

RMSE du modèle ElasticNet pour la variable SiteEnergyUse\_log avec le paramètre l1\_ratio=0.1 en fonction de l'hyperparamètre alpha



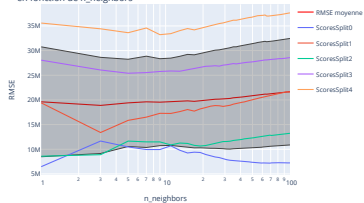
Visualisation des données de SiteEnergyUse\_log prédites par le modèle ElasticNet() vs les données test



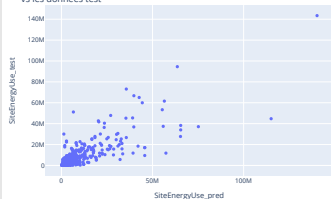
# Modèle kNeighborsRegressor

## Variable non modifiée

RMSE du modèle KNeighborsRegressor  
pour la variable SiteEnergyUse  
en fonction de n\_neighbors

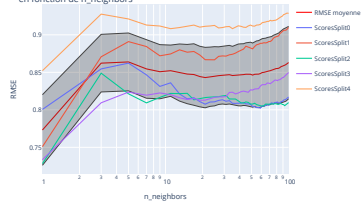


Visualisation des données de SiteEnergyUse  
prédites par le modèle KNeighborsRegressor()  
vs les données test

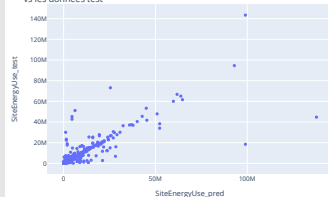


## Variable au log

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

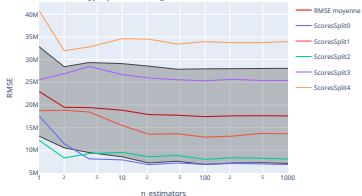




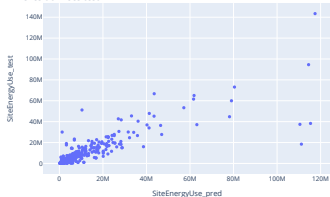
# Modèle RandomForestRegressor

## Variable non modifiée

RMSE du modèle RandomForestRegressor pour la variable SiteEnergyUse avec le paramètre max\_features=log2 en fonction de l'hyperparamètre n\_estimators

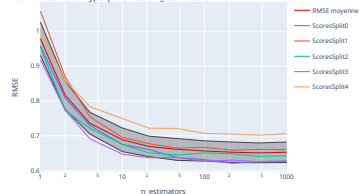


Visualisation des données de SiteEnergyUse prédites par le modèle RandomForestRegressor() vs les données test

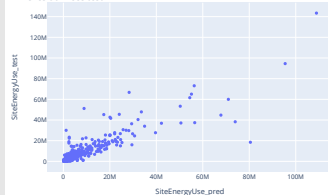


## Variable au log

RMSE du modèle RandomForestRegressor pour la variable SiteEnergyUse\_log avec le paramètre max\_features=log2 en fonction de l'hyperparamètre n\_estimators



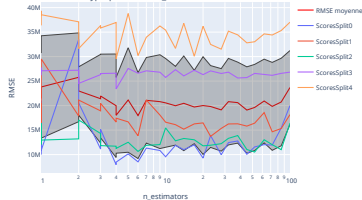
Visualisation des données de SiteEnergyUse\_log prédites par le modèle RandomForestRegressor() vs les données test



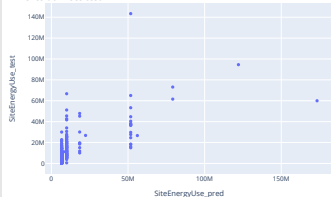
# Modèle AdaBoostRegressor

## Variable non modifiée

RMSE du modèle AdaBoostRegressor pour la variable SiteEnergyUse avec le paramètre loss=square en fonction de l'hyperparamètre n\_estimators

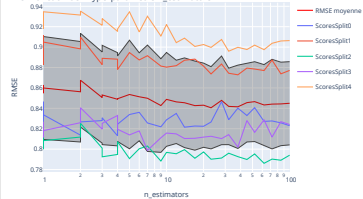


Visualisation des données de SiteEnergyUse prédites par le modèle AdaBoostRegressor() vs les données test

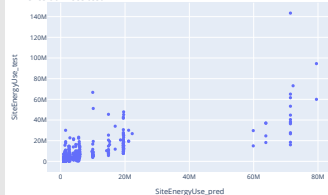


## Variable au log

RMSE du modèle AdaBoostRegressor pour la variable SiteEnergyUse\_log avec le paramètre loss=linear en fonction de l'hyperparamètre n\_estimators



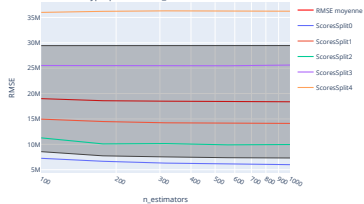
Visualisation des données de SiteEnergyUse\_log prédites par le modèle AdaBoostRegressor() vs les données test



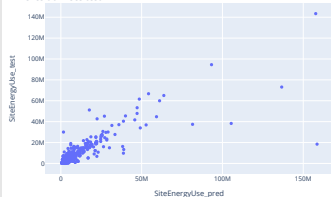
# Modèle GradientBoostingRegressor

## Variable non modifiée

RMSE du modèle GradientBoostingRegressor pour la variable SiteEnergyUse avec le paramètre loss=squared\_error en fonction de l'hyperparamètre n\_estimators

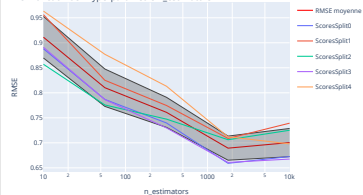


Visualisation des données de SiteEnergyUse prédites par le modèle GradientBoostingRegressor() vs les données test

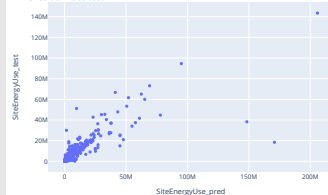


## Variable au log

RMSE du modèle GradientBoostingRegressor pour la variable SiteEnergyUse\_log avec le paramètre loss=squared\_error en fonction de l'hyperparamètre n\_estimators



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



## Conclusion

# Conclusion