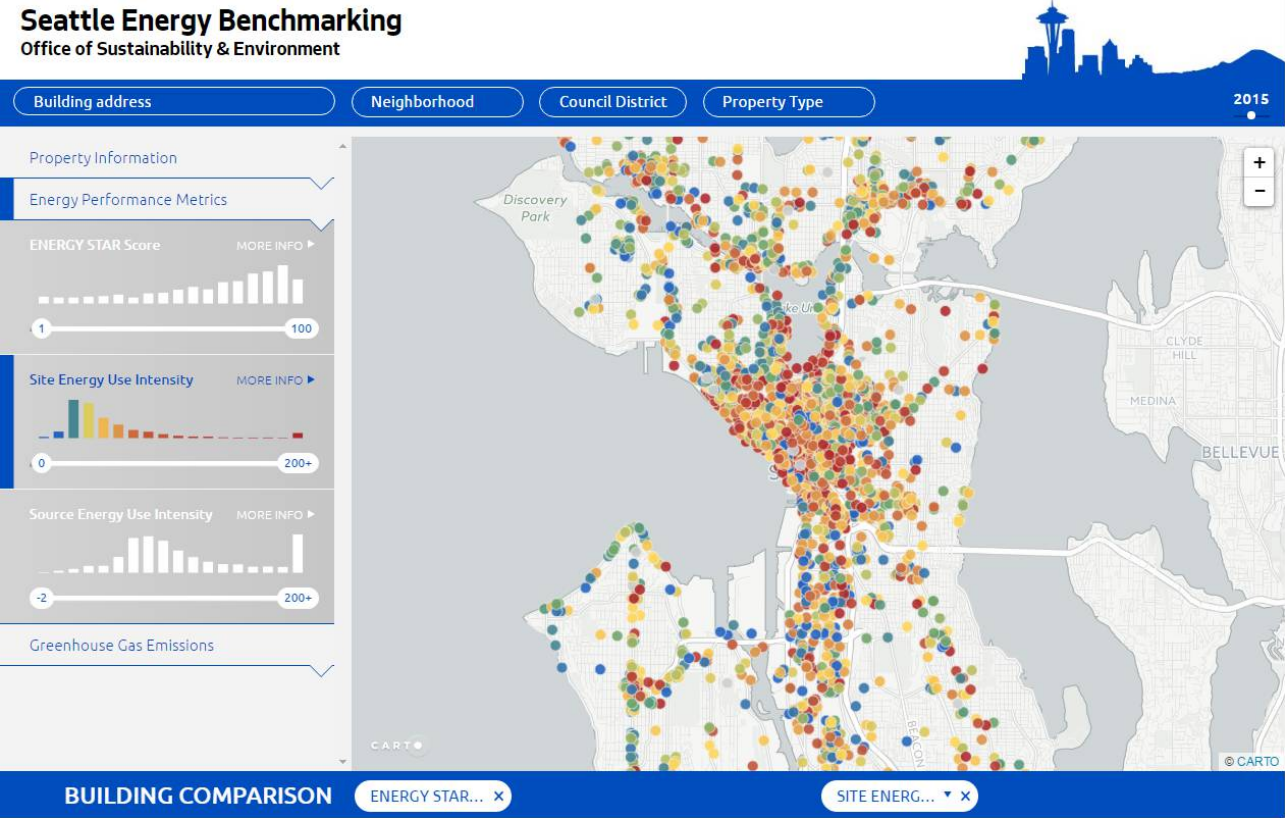


Emission de CO2 et consommation d'énergie

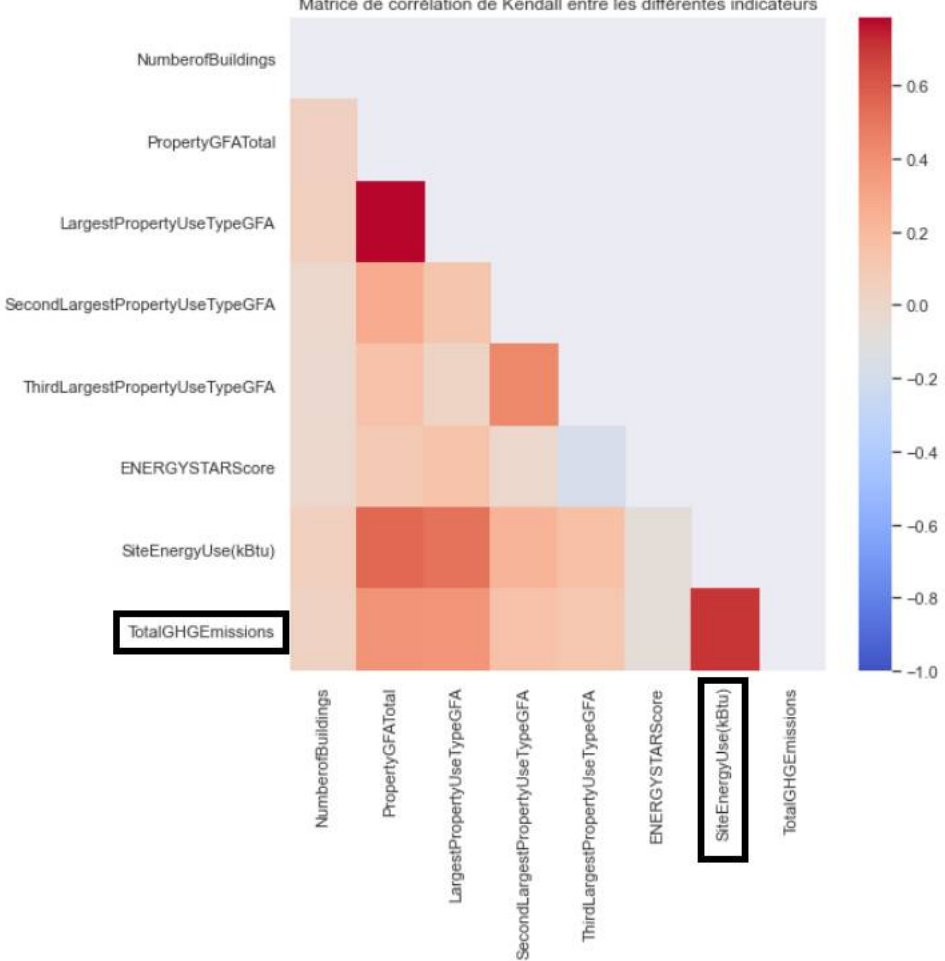
Prévoir les émissions de CO2 des futurs bâtiments

Réalisation d'une application de prévision de consommation d'énergie et d'émission de CO2. L'application se base sur le Dataset "Seattle Energy Benchmarking" qui recense la consommation d'énergie et les émissions de CO2 de nombreux bâtiments de la région de Seattle. Ce Dataset recense de nombreuses données concernant ces bâtiments :

- latitude, longitude
- année de construction
- surface totale
- nombre de building
- nombre d'étage
- surface parking
- ...



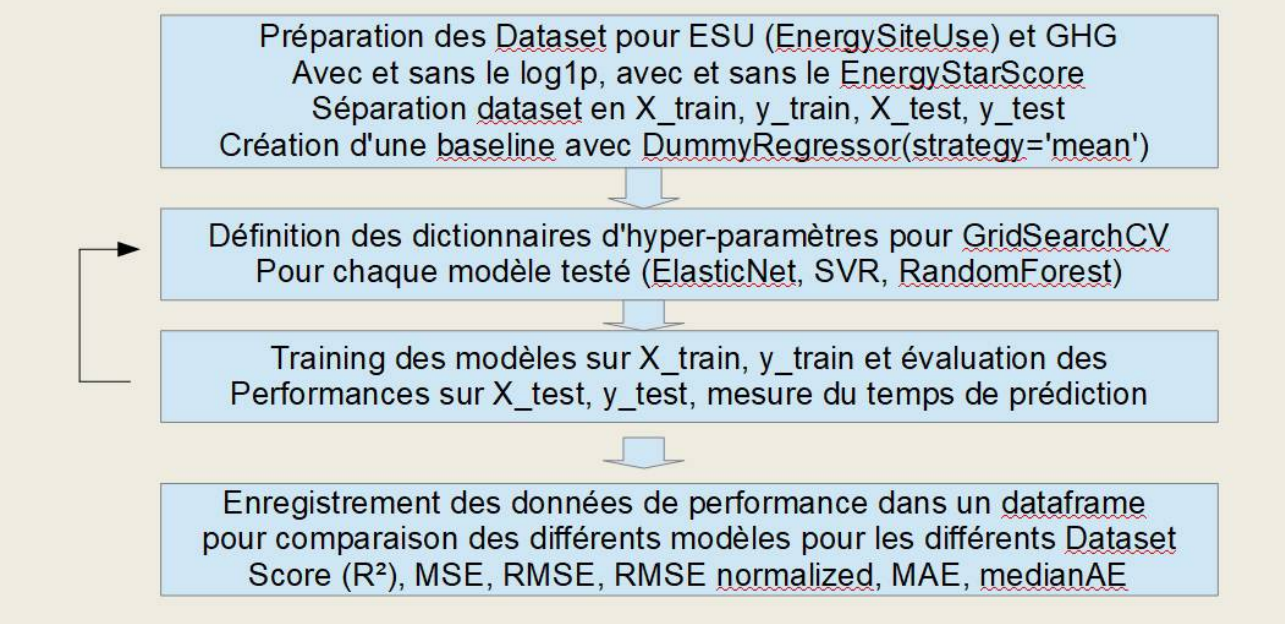
L'objectif du projet est de créer un modèle permettant la prévision de consommation d'énergie totale et d'émission de CO2 des futures constructions. Ces deux grandeurs sont fortement corrélées comme le montre la matrice de Kendal



Quatre Datasets nettoyés ont été tiré du Dataset original selon quatre variantes (passage par le log ou non) et plusieurs modèles testés :

- SVM : support vector machine
- Random Forest
- ElasticNet

Démarche générale



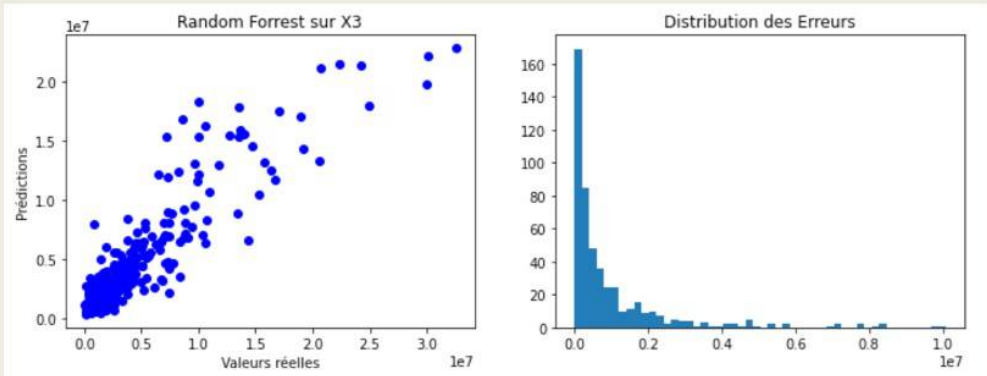
Synthèse des résultats

	Modèle	Score	nRMSE	TimePredict
0	DummyRegressor	0,00	1,4390	0,0030
1	ElasticNet sur X1	0,58	0,9303	0,0211
2	RandomForest sur X1	0,62	0,8812	0,0450
3	SVM sur X1	-0,13	1,5300	0,6639
4	ElasticNet sur X2	0,30	0,9022	0,0092
5	RandomForest sur X2	0,28	0,9606	0,1933
6	SVM sur X2	0,23	0,8785	0,2718
7	ElasticNet sur X3	0,76	0,6605	0,0199
8	RandomForest sur X3	0,84	0,5514	0,2232
9	SVM sur X3	-0,11	1,4324	0,3842
10	ElasticNet sur X4	0,74	7,6724	0,0077
11	RandomForest sur X4	0,78	0,5590	0,1702
12	SVM sur X4	0,80	1,0169	0,1416
13	SGDR sur X1	0,57	0,9453	0,0112
14	Stacking avec KNNRegressor	0,75	0,6216	0,3262

Modèle final retenu

RandomForrest sur X3  
{'max\_features': 'sqrt', 'min\_samples\_leaf': 2, 'n\_estimators': 40}

Features avec le log1p, avec le EnergyStarScore  
Target sans le log1p



Mon rôle :

- analyse
- développement
- rédactionnel

Technologies : Matplotlib, Seaborn, Pandas, Numpy, Scikit-learn, Python