

Rendu Machine learning : Analyse de la Consommation Énergétique

Introduction :

Dans le cadre de ce projet, nous avons entrepris une analyse de la consommation énergétique en utilisant des techniques de machine learning (régression linéaire). Notre objectif est de prédire la consommation énergétique à partir de diverses variables

1. Les Données :

- Nous avons utilisé un dataset comprenant plusieurs caractéristiques : superficie(m²), nombre de chambres, année de construction, type de chauffage, type de climatisation, isolation, orientation consommation énergétique(kWh).
- Le jeu de données contenait 100 entrées avec des informations détaillées sur chaque propriété.

2. Traitement des Données :

Avant d'appliquer des techniques de machine learning, nous avons effectué plusieurs étapes de prétraitement des données :

- Conversion des variables catégorielles en variables binaires pour les rendre compatibles avec les algorithmes de machine learning.
- Suppression des valeurs manquantes ou imputation de ces valeurs si possible.
- Normalisation des caractéristiques pour mettre à l'échelle les données et faciliter la convergence des algorithmes.

3. Problématique et Hypothèses :

Problématique :

Comment les différentes caractéristiques des bâtiments influent-elles sur leur consommation énergétique ?

Objectif d'analyse : Identifier les variables qui ont le plus d'impact sur la consommation énergétique des bâtiments afin de fournir des recommandations pour une meilleure efficacité énergétique.

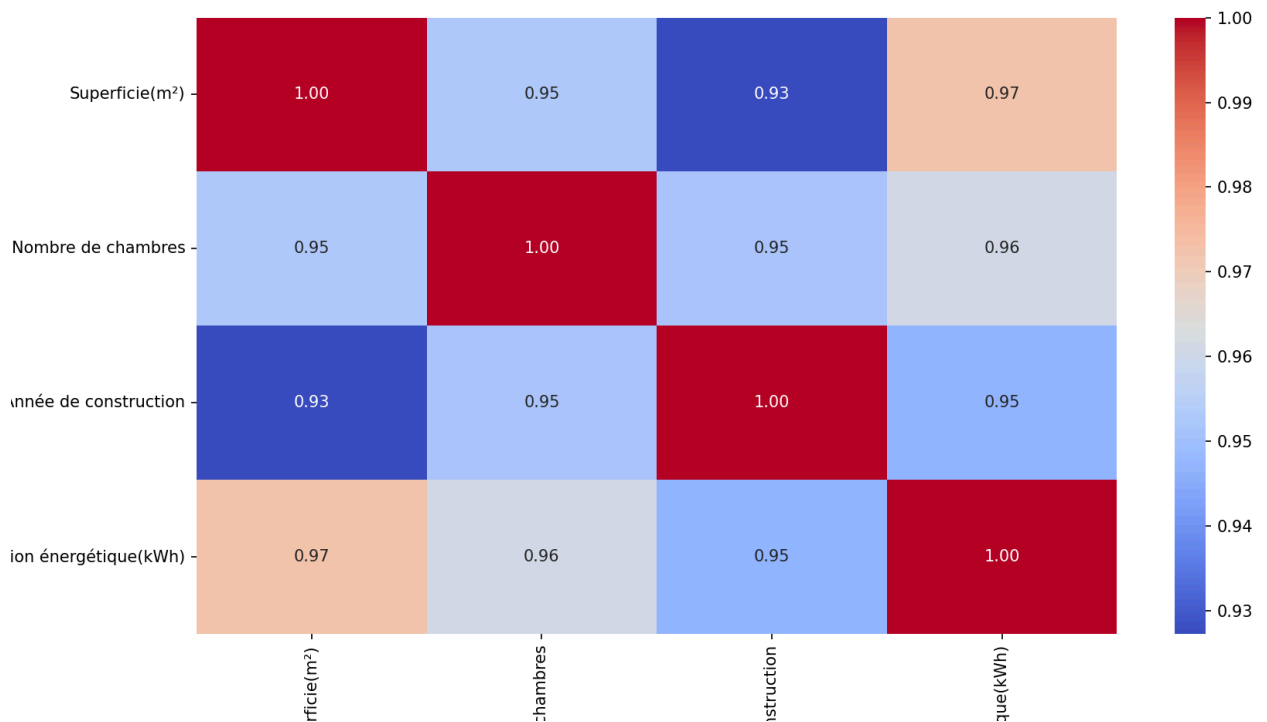
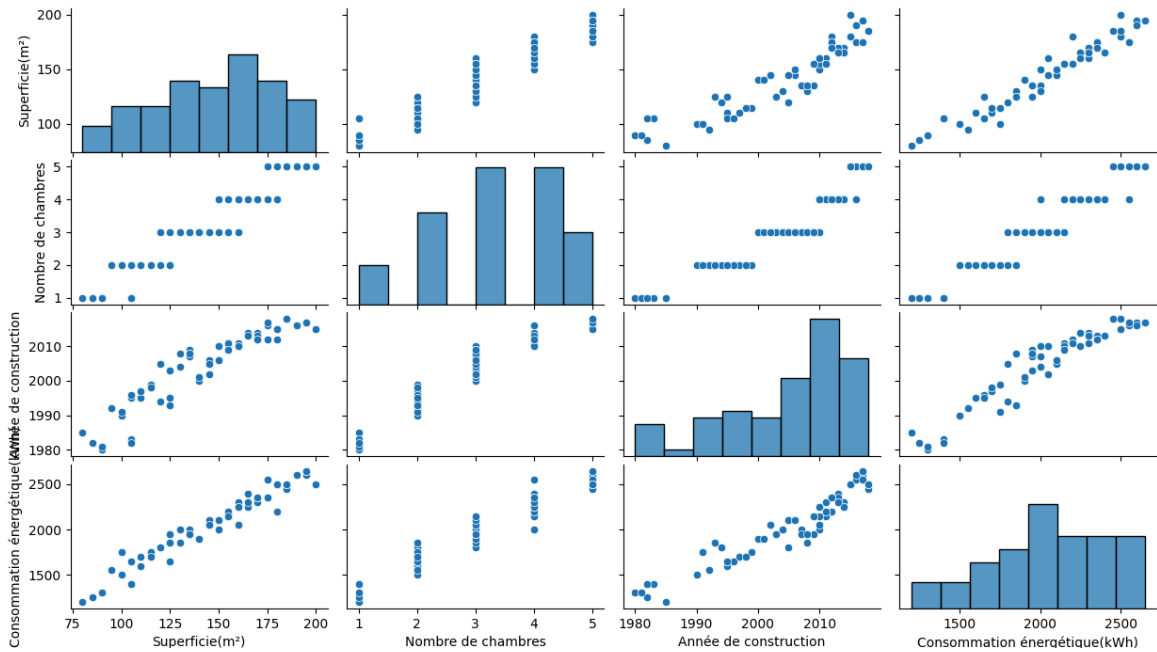
Construction d'Hypothèses :

- Plus la superficie d'un bâtiment est grande, plus sa consommation énergétique est élevée.
- Les bâtiments construits récemment ont tendance à avoir une meilleure isolation et des systèmes de chauffage/climatisation plus efficaces, ce qui entraîne une consommation énergétique plus faible.
- Le type de chauffage (par exemple, gaz, électrique) peut avoir un impact significatif sur la consommation énergétique.
- L'orientation du bâtiment peut également jouer un rôle dans sa consommation énergétique, en affectant l'exposition au soleil et aux vents.

4. Analyse exploratoire :

À l'aide d'un script python utilisant les modules panda, seaborn & matplotlib, notre premier code nous permet d'explorer et de visualiser les relations entre les variables dans les données, ainsi que de comprendre la corrélation entre elles. Cela aide à identifier les tendances, les patterns et les relations potentielles qui peuvent être importantes pour la modélisation et l'analyse ultérieures.

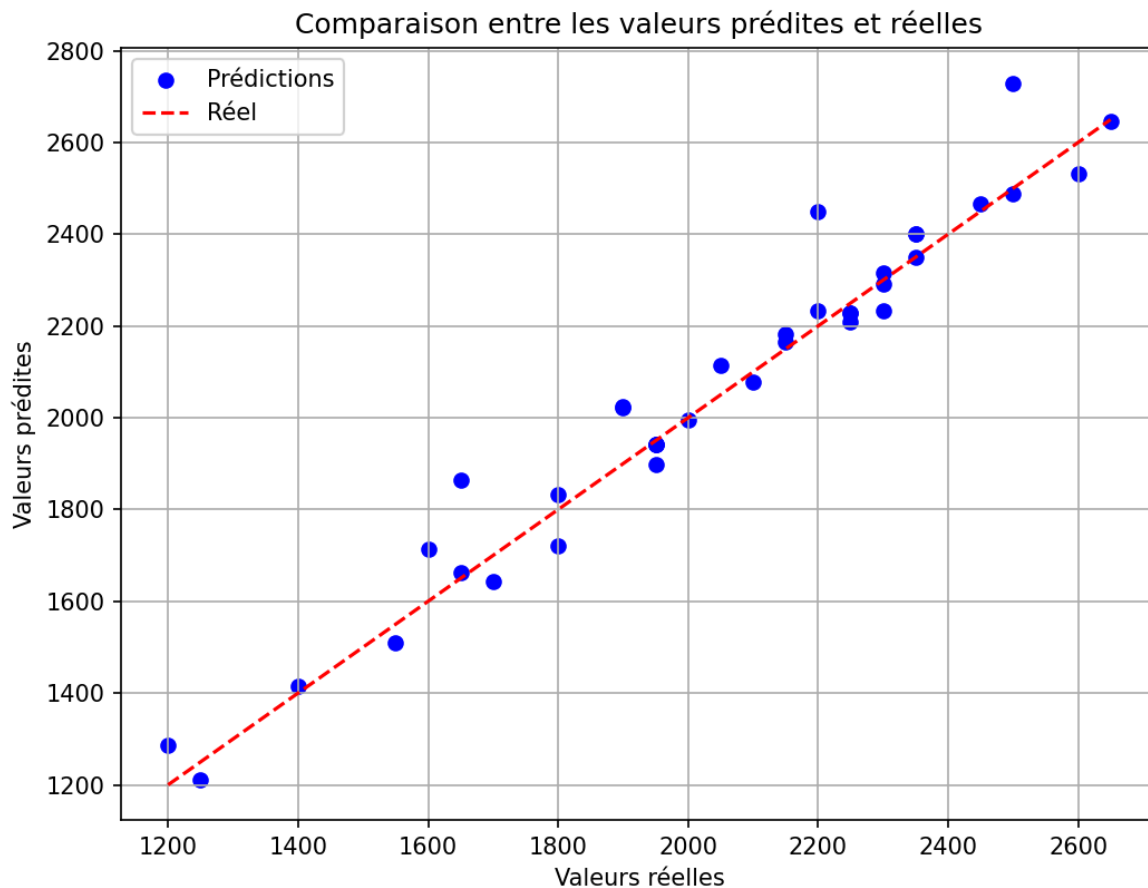
	Superficie(m²)	Nombre de chambres	Année de construction	Consommation énergétique(kWh)
count	73.000000	73.000000	73.000000	73.000000
mean	145.205479	3.232877	2005.150685	2047.945205
std	30.395979	1.172847	10.211202	364.428620
min	80.000000	1.000000	1980.000000	1200.000000
25%	125.000000	2.000000	1998.000000	1800.000000
50%	150.000000	3.000000	2008.000000	2100.000000
75%	170.000000	4.000000	2013.000000	2300.000000
max	200.000000	5.000000	2018.000000	2650.000000



Modèle de prédiction :

Nous avons utilisé la méthode de **régression linéaire** pour prédire la consommation énergétique des bâtiments. Une technique de modélisation statistique qui cherche à établir une relation linéaire entre une variable dépendante (la cible à prédire, dans ce

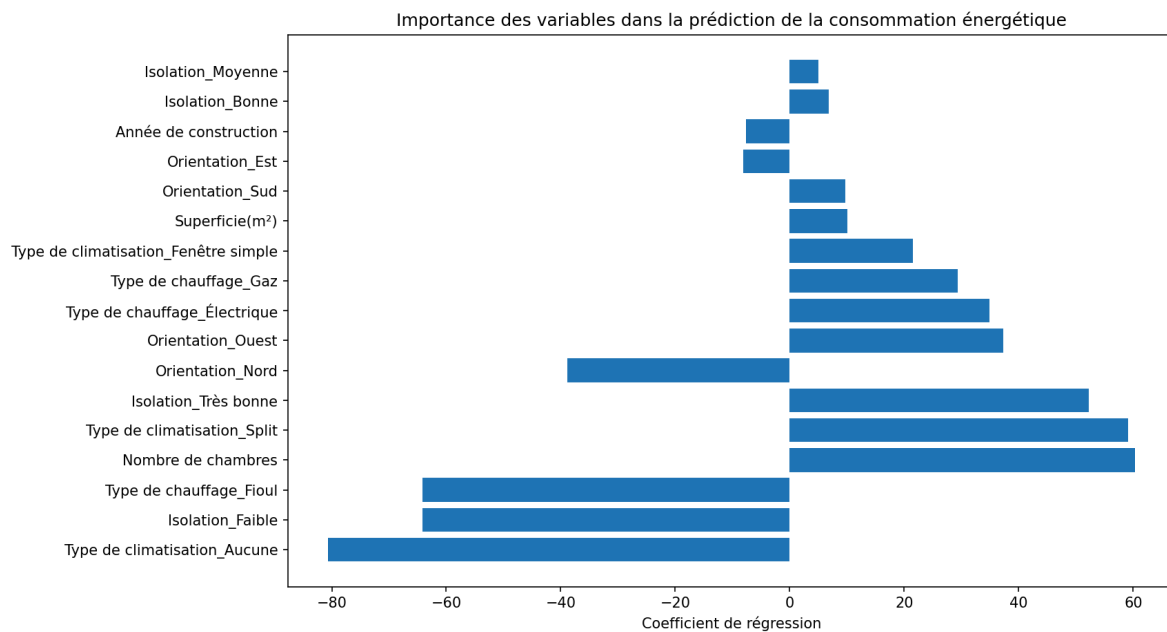
cas la consommation énergétique) et une ou plusieurs variables indépendantes (les caractéristiques qui influencent la cible, dans ce cas les différentes données de notre dataset)



Calcul du **coefficient de détermination (R^2)** : Ce coefficient mesure la proportion de variance des valeurs cibles expliquée par le modèle. Un R^2 proche de 1 indique que le modèle explique bien la variance des données, tandis qu'un R^2 proche de 0 indique un modèle qui n'explique pas bien la variance.

Dans notre cas, il est de **0,94**, ce qui peut être un peu élevé, idéalement un R^2 doit se situer aux alentours de **0,8**.

Pour identifier les caractéristiques les plus importantes qui influencent la consommation énergétique, nous allons examiner les coefficients de régression estimés par le modèle de régression linéaire. Les coefficients indiquent l'importance de chaque variable explicative dans la prédiction de la variable cible, dans ce cas la consommation énergétique.



Variable Coefficient

- Type de climatisation_Aucune : **-80.619669**
- Isolation_Faible : **-64.168738**
- Type de chauffage_Fioul : **-64.168738**
- Nombre de chambres : **60.257670**
- Type de climatisation_Split : **59.124489**
- Isolation_Très bonne : **52.259364**
- Orientation_Nord : **-38.901278**
- Orientation_Ouest : **37.316025**
- Type de chauffage_Électrique : **34.835914**
- Type de chauffage_Gaz : **29.332824**
- Type de climatisation_Fenêtre simple : **21.495180**
- Superficie(m²) : **10.106281**
- Orientation_Sud : **9.664672**
- Orientation_Est : **-8.079418**
- Année de construction : **-7.659966**
- Isolation_Bonne : **6.865125**
- Isolation_Moyenne : **5.044249**

Donc dans notre cas, les caractéristiques qui influent le plus la consommation énergétiques sont :

En analysant les coefficients de régression estimés, voici quelques conclusions que l'on peut tirer :

1. Impact négatif sur la consommation énergétique :

- Les caractéristiques telles que l'absence de climatisation, une isolation faible, et le chauffage au fioul ont des coefficients négatifs élevés. Cela signifie que ces facteurs sont associés à une diminution de la consommation énergétique prédite.
- En d'autres termes, les maisons sans climatisation, avec une isolation faible, et chauffées au fioul ont tendance à consommer moins d'énergie.

2. Impact positif sur la consommation énergétique :

- En revanche, le nombre de chambres, le type de climatisation (split), une isolation très bonne, et certaines orientations (Ouest, Électrique) ont des coefficients positifs significatifs. Cela suggère que ces caractéristiques sont associées à une augmentation de la consommation énergétique prédite.
- Par exemple, les maisons avec plus de chambres, une climatisation de type split, une isolation très bonne, et une orientation Ouest ont tendance à consommer plus d'énergie.

3. Facteurs moins significatifs :

- Certains facteurs ont des coefficients proches de zéro ou de faible magnitude, ce qui indique qu'ils ont moins d'influence sur la consommation énergétique prédite. Cela inclut des caractéristiques telles que la superficie, l'orientation Sud, l'année de construction, etc.

Hypothèses validées :

1. **Type de climatisation et de chauffage** : Les coefficients négatifs pour les variables telles que "Type de climatisation_Aucune", "Isolation_Faible", "Type de chauffage_Fioul", suggèrent que l'absence de climatisation ou de chauffage central, ainsi qu'une isolation faible, ont un effet **négatif** sur la consommation énergétique, ce qui est cohérent avec l'intuition.
2. **Orientation** : Les coefficients **positifs et négatifs** pour les différentes orientations suggèrent que l'orientation du bâtiment peut jouer un rôle dans la consommation énergétique.
Par exemple, une orientation sud peut augmenter la consommation énergétique, tandis qu'une orientation nord peut la réduire.

Hypothèses invalidées :

1. **Superficie** : Le coefficient positif pour la variable "Superficie(m²)" suggère que la superficie du bâtiment a un effet positif sur la consommation énergétique, ce qui est **contraire à l'hypothèse** selon laquelle une superficie plus grande entraînerait une consommation énergétique plus élevée. Cela pourrait indiquer

une complexité supplémentaire dans la relation entre la superficie et la consommation énergétique, nécessitant une analyse plus approfondie.

2. **Année de construction** : Le coefficient **négatif** pour l'année de construction suggère que des bâtiments plus récents ont une consommation énergétique plus **faible**, ce qui est contraire à l'hypothèse selon laquelle des bâtiments plus anciens pourraient être moins efficaces sur le plan énergétique. Cela nécessiterait également une exploration plus détaillée pour comprendre les raisons derrière cette relation.

Ces conclusions basées sur les coefficients de régression doivent être interprétées avec prudence et peuvent nécessiter des analyses supplémentaires pour confirmer ou infirmer définitivement les hypothèses.