

Projet Open Data : *Consommations d'électricité*

GARCIA Lisa - ROBERT Ambre

Université de Bordeaux

February 2, 2023

Sommaire

- 1 Introduction
- 2 Les données
- 3 Analyse descriptive
- 4 Prédictions
- 5 Conclusion

Introduction

Introduction

- Objectif : étudier les différents facteurs qui font varier la consommation d'électricité
- Plateformes d'Open Data utilisées : Odré et le site de l'INSEE
- Zone géographique étudiée : les différentes régions de France métropolitaine, hors Corse
- Domaine temporel : par heure pour des données allant de 2016 à 2022

Les données

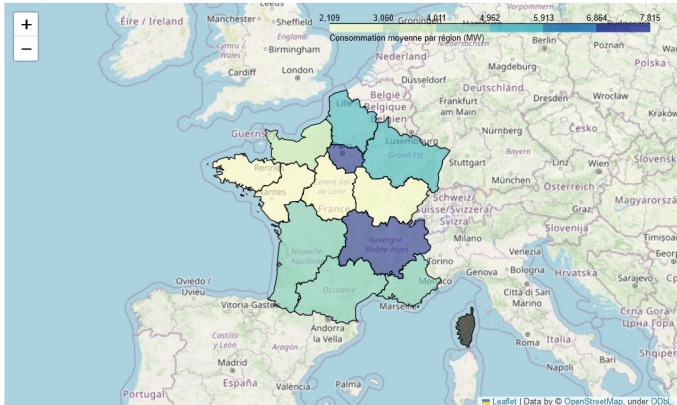
Jeu de données

Variables	Description
Région	Région de France métropolitaine (hors Corse)
Année	De 2016 à 2022
Mois	Numéro du mois de l'année
Jour	Numéro associé au jour de la semaine
Heure	Point horaire
Consommation (MW)	Consommation en Mégawatt
Date	Date du jour
TMin (°C)	Température minimale en degrés celsius
TMax (°C)	Température maximale en degrés celsius
TMoy (°C)	Température moyenne en degrés celsius
Population	Nombre d'habitant pour une région et une année données
Bioénergies (MW)	Production Bioénergies en Mégawatt
Hydraulique (MW)	Production hydraulique en Mégawatt
Eolien (MW)	Production éolienne en Mégawatt
Solaire (MW)	Production solaire en Mégawatt
Nucléaire (MW)	Production nucléaire en Mégawatt
Thermique (MW)	Production thermique en Mégawatt
Ech. physiques (MW)	Solde imports/exports (flux physiques) en Mégawatt

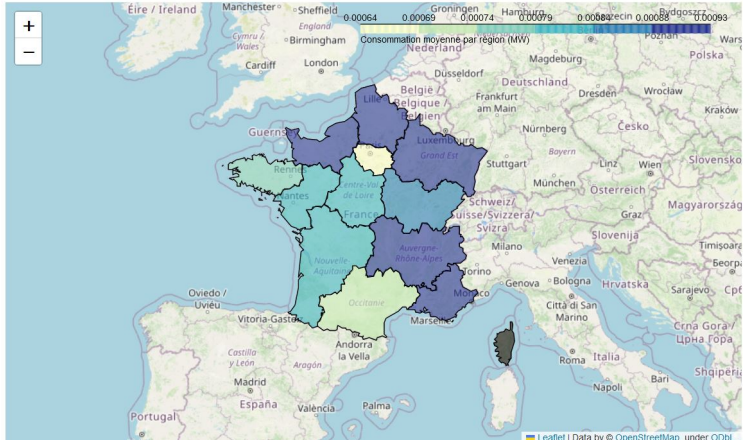
Table: Tableau des variables du jeu de données.

Analyse descriptive

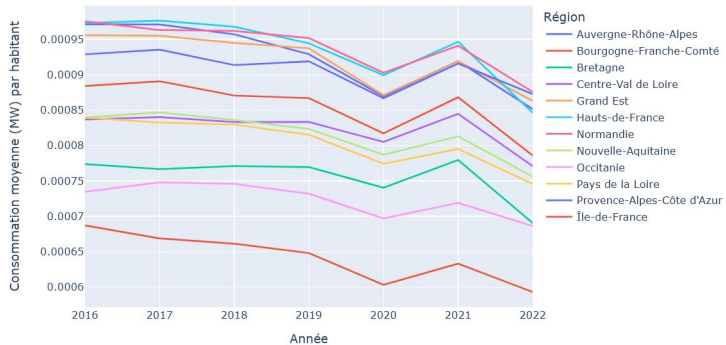
Carte de la consommation moyenne d'électricité (en MW) entre 2016 et 2022 par région



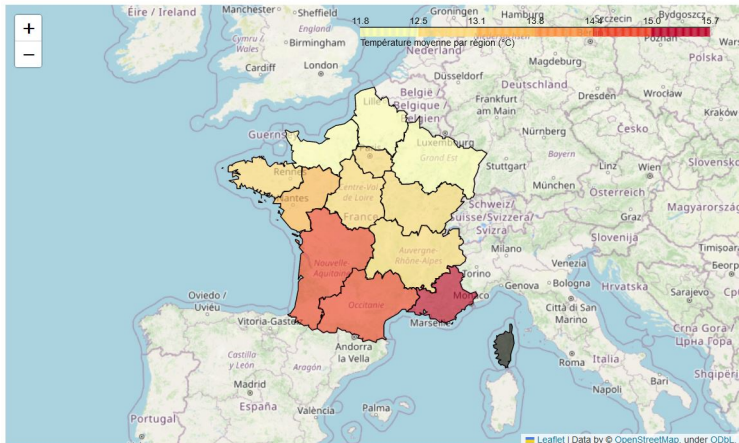
Carte de la consommation moyenne d'électricité (en MW) par habitant par région entre 2016 et 2022



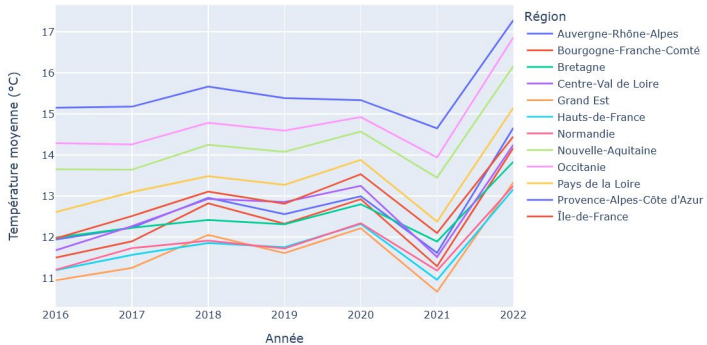
Consommation moyenne par habitant par région en fonction de l'année



Carte de chaleur des températures moyennes par région entre 2016 et 2022



Température moyenne par région en fonction de l'année



Etude sur les mois

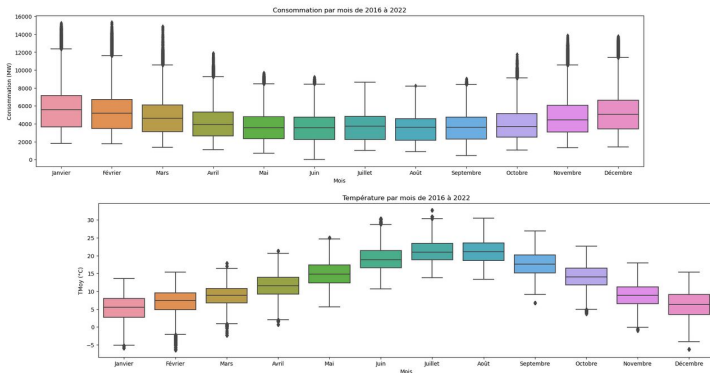
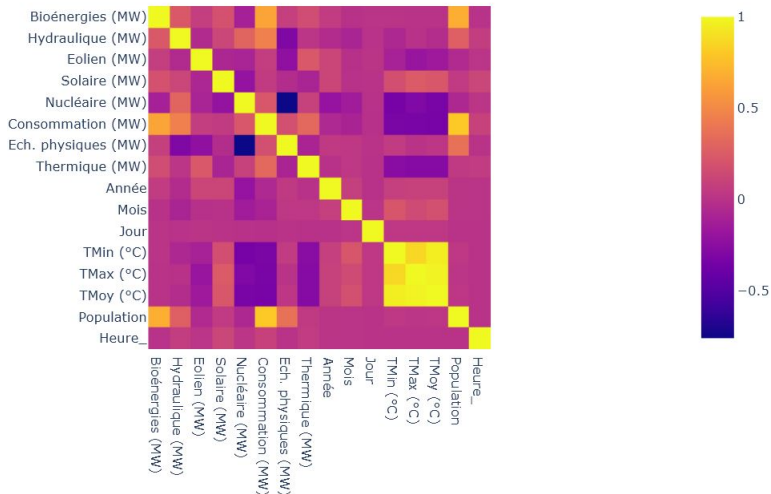


Figure: Consommations d'électricité (haut) et températures (bas) par mois

Matrice des corrélations



Prédictions

Prédictions

Objectif

Nous souhaitons prédire les consommations d'électricité pour le dernier trimestre de 2022, soit pour des dates entre le 1er octobre 2022 et le 31 décembre 2022.

Méthode

- Centrer réduire nos données.
- Découpage du jeu de données en apprentissage/test.

Prédictions

Objectif

Nous souhaitons prédire les consommations d'électricité pour le dernier trimestre de 2022, soit pour des dates entre le 1er octobre 2022 et le 31 décembre 2022.

Méthode

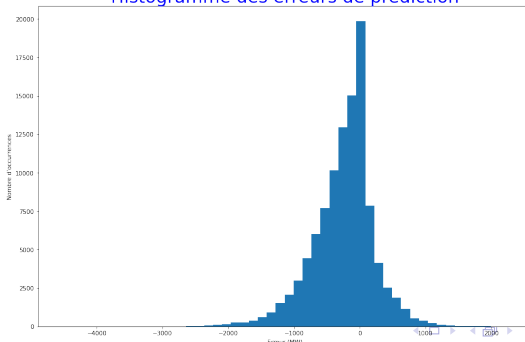
- Centrer réduire nos données.
- Découpage du jeu de données en apprentissage/test.

Forêt Aléatoire

Avec paramètres par défaut

$$\text{RMSE} = \sqrt{\mathbb{E} \left((\hat{y} - y)^2 \right)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \approx 528.957$$

Histogramme des erreurs de prédiction



Forêt Aléatoire - Variables importantes

	feature	importance
0	Population	0.7
1	TMax (°C)	0.15
2	Ech. physiques (MW)	0.04
3	Thermique (MW)	0.04
4	TMoy (°C)	0.02
5	Hydraulique (MW)	0.01
6	Solaire (MW)	0.01
7	Nucléaire (MW)	0.01

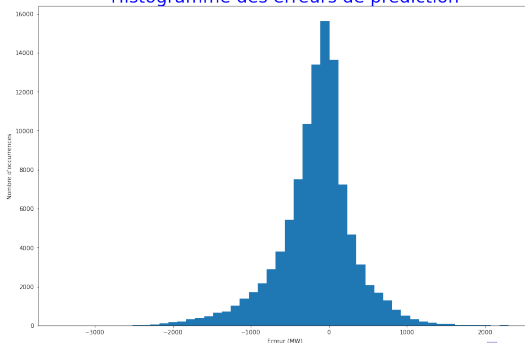
Figure: Variables importantes dans la forêt aléatoire.

Forêt Aléatoire

Sur le jeu de données réduit et avec les paramètres par défaut

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \approx 516.165$$

Histogramme des erreurs de prédiction



Forêt Aléatoire - Comparaison des résultats

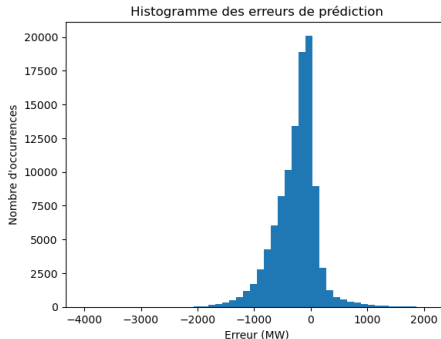
Jeu de données	Paramètres	RMSE
Complet	Par défaut	528.957
Réduit	Par défaut	516.165
Réduit	CV	515.602

Table: Comparaison des résultats.

Réseau de neurones

Avec codage disjonctif complet sur la variable Mois

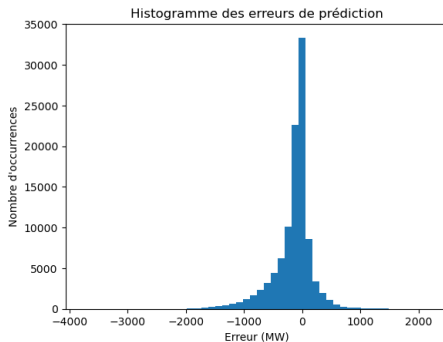
$$\text{RMSE} = \sqrt{\mathbb{E}((\hat{y} - y)^2)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \approx 461.76$$



Réseau de neurones

Avec la variable Mois comme nombre entre 1 et 12

$$\text{RMSE} = \sqrt{\mathbb{E}((\hat{y} - y)^2)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \approx 368.75$$



Conclusion

Conclusion

Pour conclure

- Différents facteurs influençant la consommation d'électricité
- Tendance du modèle à surestimer car moins de consommation en 2022

Pistes d'améliorations

- Enlever l'année 2020
- Ajouter d'autres variables exogènes
- Prédire sur la variable de consommation par habitant