

# RAPPORT FINAL

Projet de Smart System

---

## Estimation de la consommation électrique d'une maison en fonction de la température extérieure

---

### Étudiants :

**BARDEY Matthieu**

**SATGE Valentin**

### Summary

#### I. Introduction

- 1) Contexte
- 2) Objectif
- 3) Méthode choisie

#### II. Acquisition et traitement des données

- 1) Acquisition des données
- 2) Traitement des données

#### III. Résultats du modèle

- 1) Résultats obtenus
- 2) Analyse des résultats

#### IV. Conclusions et perspectives

## **I) INTRODUCTION**

### **1) Contexte :**

En France, le secteur du bâtiment (résidentiel et tertiaire) représente 49% de la consommation de l'énergie primaire du pays, et 18% des émissions de gaz à effet de serre (GES). Pour les maisons construites en 2000 ou plus tard, il est important de noter que 47% de l'énergie consommée est utilisée pour le chauffage des surfaces. Ce pourcentage est même de 62% pour les maisons construites avant 1975. On comprend bien que le chauffage est un sujet brûlant pour la transition écologique. Cette dernière doit d'abord s'accompagner d'une forte rénovation des logements pour empêcher les pertes thermiques importantes et donc le gaspillage énergétique. Le terme sobriété est également entré dans le langage courant pour réaliser des économies énergétiques et budgétaires pour les ménages. Un autre point non négligeable de la réduction des consommations d'énergie des bâtiments est la programmation du chauffage, qui permet au consommateur de décider d'une température cible pour toutes les heures, et tous les jours de la semaine en fonction de ses besoins.

Ces enjeux énergétiques liés au chauffage sont particulièrement importants pour optimiser notre consommation d'énergie mais aussi, dans un contexte d'électrification accrue des usages, déterminants pour garantir la sécurité et le bon fonctionnement du système électrique.

### **2) Objectif :**

On peut alors supposer que l'optimisation de la consommation électrique d'une habitation dépend de plusieurs paramètres liés aux besoins en chauffage : la saison, la météo, l'heure de la journée par exemple. Ces paramètres ont un point commun : ils font intervenir des températures extérieures variées qui vont beaucoup impacter la consommation énergétique des habitations qui dépend beaucoup du chauffage.

Ainsi, l'objectif de notre étude est de comprendre à quel point la consommation d'énergie d'une maison est corrélée à la température extérieure, et s'il est possible et cohérent de construire des modèles fiables destinés à déterminer la consommation des habitations à un moment donné, un jour donné ou une semaine, en fonction du paramètre déterminant : la température extérieure.

L'élaboration de tels modèles est vouée à permettre une meilleure prédiction des besoins énergétiques des ménages, et donc une consommation plus raisonnée de l'énergie et une optimisation des systèmes de production et de distribution de l'électricité. Les résultats de cette étude permettront, si les résultats sont assez précis, de passer à une nouvelle phase : réussir à anticiper la production des panneaux photovoltaïques de la maison. Ainsi, avec les données de consommation de la maison et de la production de ses panneaux photovoltaïques, il sera possible d'optimiser l'auto-consommation du bâtiment en ajoutant une batterie.

### **3) Méthode choisie :**

Pour construire notre modèle, nous allons utiliser une méthode de classification des données. Les données qui nous intéressent sont donc la température extérieure et la consommation électrique d'un bâtiment, d'un appartement ou d'une maison individuelle. Il nous faudra d'abord récupérer ces données et les traiter pour les mettre en forme et pouvoir les utiliser correctement.

Lors de la manipulation des données, la méthode des clusters sera utilisée. L'objectif est de séparer les données de consommation en trois groupes : un groupe pour la consommation lors des températures extérieures hautes, un autre groupe pour des températures extérieures moyennes et un dernier pour la consommation lors des températures basses. Cette étape permet à la fois d'observer si trois groupes de consommation électrique se détachent, validant notre hypothèse que celle-ci dépend fortement de la température extérieure, mais aussi de rendre notre modèle plus précis en créant des groupes de données plutôt qu'en lui imposant de traiter toutes les données d'un coup sans les classer.

Une fois cette étape réalisée, le modèle pourra être utilisé et évalué. Pour se faire, il sera amené à prédire des consommations électriques avec comme données d'entrées uniquement la température extérieure - l'utilisation de moyenne de la température extérieure sur la dernière semaine et sur les dernières 24 heures seront utilisées. Le résultat donné sera comparé aux données de consommation réelles. De ces résultats nous pourrons tirer des conclusions sur le modèle créé et son efficacité, ainsi qu'envisager d'éventuelles améliorations.

Pour créer le modèle, la méthode de "decision tree" sera utilisée. L'un des facteurs importants est de trouver la bonne profondeur (la bonne précision) du modèle qui peut influencer l'efficacité mais aussi le temps de calcul pour obtenir des résultats. Ici, il faut compter entre 10 et 40 minutes pour obtenir chacun des graphes ci-dessous.

## **II) ACQUISITION ET TRAITEMENT DES DONNÉES**

### **1) Acquisition des données**

Pour collecter les données nécessaires à la construction de notre modèle, nous récupérons les données de consommation électrique et de température extérieure de Expe-Smarthouse. Il s'agit d'une maison intelligente donnant accès à environ 340 points de mesure pour les scientifiques mais pas que, puisque toutes les données recueillies sont accessibles en temps réel via un portail Grafana alimenté sur une base de données Influxdb. Ce projet, développé sur la base de logiciels Open Source, nous permet d'avoir une base de données expérimentales réelles composées de données sur un pas de temps de la demi-heure depuis l'année 2018.

## 2) Traitement des données

Les fichiers obtenus sont en format .csv, que nous utilisons sous Python grâce à l'environnement panda. Un travail de resample est alors nécessaire pour s'assurer que le croisement des données de consommation électrique et de température se fera correctement : concordances des dates et heures de données initiales et finales, même pas de temps, etc. Il faudra être particulièrement précautionneux avec l'existence de données erronées (les capteurs ne fonctionnent pas certains jours ou donnent des valeurs aberrantes). Une fois cette remise en forme réalisée, la méthode K-mean des clusters peut être implémentée pour séparer les températures en trois groupes comme expliqué dans la partie I.3. Cette méthode consiste à diviser toutes les données que obtenu en un nombre défini de sous-groupes - 3 ici - en fonction du niveau de la consommation (forte consommation, consommation moyenne et faible consommation).

## III) RESULTATS DU MODELE

### 1) Résultats obtenus

Comme expliqué plus haut, la première étape de construction du modèle était de réaliser trois clusters pour les données de température correspondants à une tranche de températures basses, une moyennes et une élevées. Ensuite, il est possible d'observer le comportement de la consommation électrique de la maison en fonction de ces trois clusters :

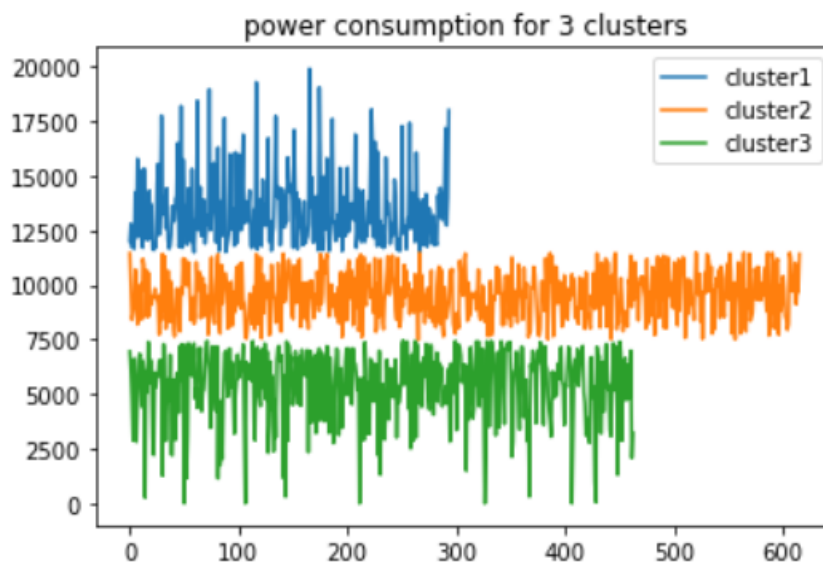


Figure 1 : consommation électrique pour les trois clusters

Cette figure fait bien apparaître des plages de consommation électrique différentes pour chacun des trois clusters : pour le cluster 1 (températures basses), la consommation est la plus élevée. Pour le cluster 2 (températures moyennes), la consommation est plus faible. Enfin pour le cluster 3 (températures hautes), la consommation est la plus basse. Ces

observations mettent en évidence la cohérence de réaliser des clusters sur la température extérieure pour estimer la température de la maison.

Désormais nous sommes en mesure d'exploiter le modèle pour estimer sa précision en termes de prédiction des consommations électriques lorsqu'on lui fournit les données de températures extérieures.

Les prédictions de consommations électriques concernant les températures du cluster 1 sont les suivantes, comparées aux consommations réelles :

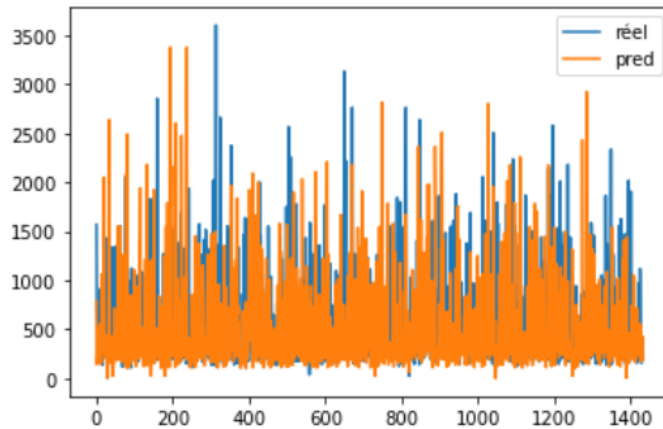


Figure 2 : comparaison des consommations réelles et estimées par le modèle pour le cluster 1

Pour les consommations correspondant au cluster 2, les résultats suivants sont obtenus :

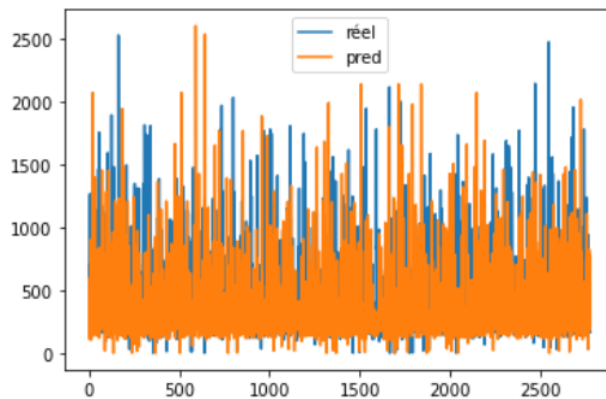


Figure 3 : comparaison des consommations réelles et estimées par le modèle pour le cluster 2

Enfin pour les températures du cluster 3, nous obtenons les résultats suivants :

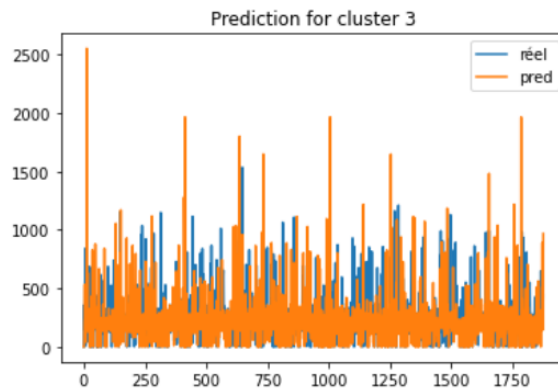


Figure 4 : comparaison des consommations réelles et estimées par le modèle pour le cluster 3

## 2) Analyse des résultats

Sur les trois résultats obtenus, on constate d'abord que l'ordre de grandeur des consommations estimées et des consommations réelles sont les mêmes. De plus, le modèle la méthode de clustering semble efficace car les estimations du modèle restent fidèles aux variations de consommations sur des pas de temps relativement longs, et on voit bien la consommation estimée réduire considérablement entre le cluster 1 (températures extérieures basses), et le cluster 3 (températures extérieures hautes).

En revanche, on remarque que les estimations s'écartent des consommation réelles enregistrées pour des variations importantes sur des pas de temps courts. Ici, la fiabilité du modèle peuvent atteindre leurs limites car nous n'avons pris en compte qu'un seul paramètre : la température extérieure. Or, de nombreux autres paramètres peuvent faire varier la consommation électrique comme le nombre de personnes dans chaque pièce, l'ensoleillement, la date de l'année, le nombre d'appareils allumés...

Il est également important pour notre étude de nous intéresser à la robustesse de la méthode des clusters. Pour cela on obtient ci-dessous les estimations des consommations pour toutes les températures, sans utiliser de clusters :

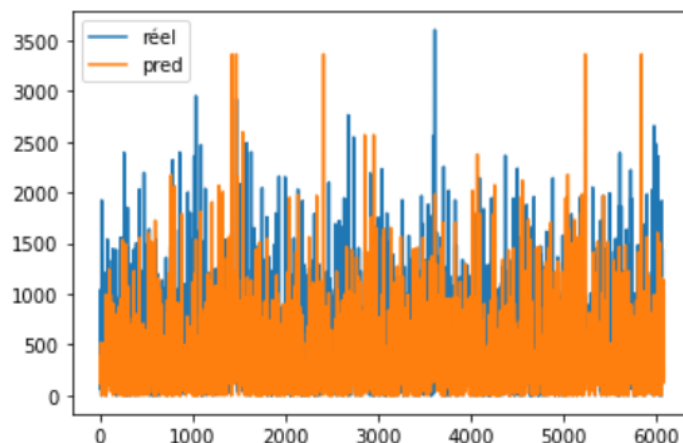


Figure 5 : comparaison des consommations réelles et estimées par le modèle sans la méthode des clusters

Voici les Rscores obtenus pour chacune des quatre simulations précédemment effectuées :

	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Pas de clusters</b>
<b>Rscore</b>	0,50	0,54	0,78	0,41

Table 1 : Rscore en fonction des features (température extérieure) fournis au modèle

Bien que ces valeurs de Rscore soient relativement faibles - plus le Rscore est proche de 1, plus la valeur sera précise - la méthode des clusters est plus efficace que lorsque toutes les données sont traitées en même temps. Utiliser cette méthode des clusters semble être une étape primordiale pour améliorer l'efficacité d'un modèle, même lorsque le nombre de types de données d'entrées est assez faible.

#### **IV) CONCLUSIONS ET PERSPECTIVES**

Certes, les résultats montrent qu'il existe une grande corrélation entre ce paramètre et la consommation électrique de la maison. Mais pour avoir des estimations plus précises sur des variations de consommation de l'ordre du jour ou de l'heure, il faudra plus de paramètres (comme le nombre de personnes dans la maison par exemple, expliqué plus haut) qui vont faire varier la consommation électrique liée au chauffage ou non. Néanmoins, les prédictions du modèle se rapprochent de la réalité, ce qui est un très bon début pour le développement d'un modèle plus performant. Effectivement, les résultats montrent bien que la température extérieure est un des facteurs qui fait le plus varier la consommation électrique des ménages.

Pour développer un outil plus performant, on peut imaginer se concentrer sur d'autres paramètres que le modèle pourrait traiter grâce à des données recueillies sur des capteurs (capteur de présence par exemple) pour gagner en précision sur les variations de consommation qui ne sont pas dues uniquement à la variation de la température extérieure. De plus, on a vu qu'avec des clusters, le modèle est plus précis. Le développement plus poussé d'un modèle similaire doit se concentrer sur le nombre optimal de cluster pour maximiser la précision des estimations.

Des travaux futurs pourraient se concentrer sur la modélisation de la production des panneaux solaires pour améliorer l'auto-consommation et utiliser les résultats de la présente étude. Cependant, il faut faire attention aux résultats obtenus lors de la mise en place de ces panneaux. En effet, la consommation du bâtiment est prise jusqu'en mai 2022, avant l'implémentation de ces panneaux. Avec des panneaux solaires sur une maison, les consommations d'un bâtiment peuvent changer : il n'est plus nécessaire de chercher à consommer majoritairement dans les heures dites "creuses" : au milieu de la nuit, mais d'augmenter sa consommation lorsque les panneaux produisent le plus : au milieu de la journée. Ainsi, la courbe de charge peut évoluer entre l'avant installation des panneaux et l'après installation. Une étude pourrait donc vérifier l'évolution de la précision de ces modèles lors de l'ajout des panneaux. Dans le cas où la précision diminue, il faudrait changer de techniques de machine learning pour réussir à utiliser les données actuelles.