

TP : Smart System

Estimation du taux d'occupation de l'EXPE-SMARTHOUSE afin d'appliquer un plan de sobriété énergétique en conséquence.



Enseignant : Jérôme FERRARI

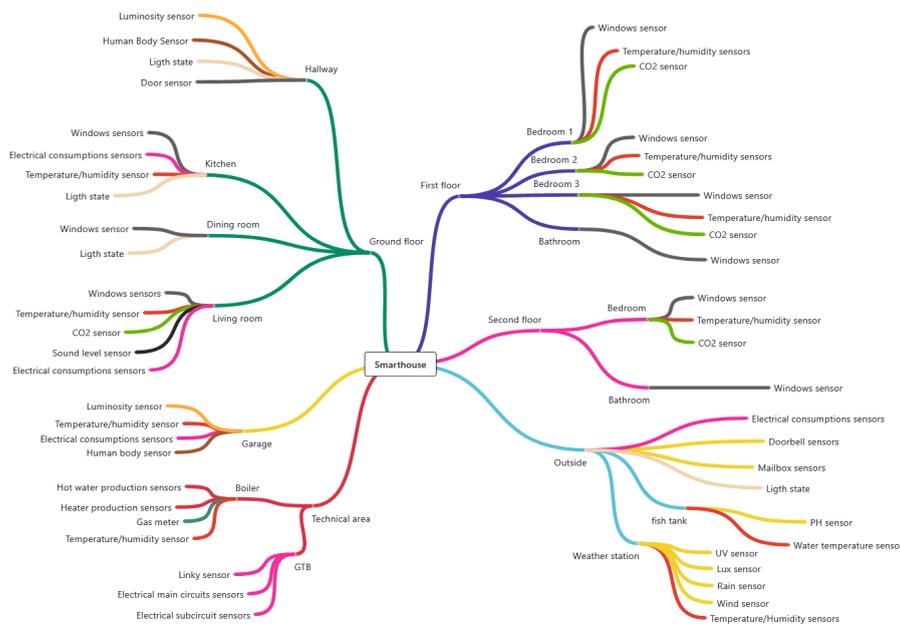
Étudiants : Antonin ANDRE, Pauline BERNARD

Sujet : Estimer le nombre de personnes dans une pièce à l'aide des taux de CO2 dans la pièce, des capteurs de mouvement dans les couloirs, ...
 Objectif : calculer les économies possibles d'énergie

Introduction :

L'EXPE-SMARTHOUSE est une maison test ultra connectée fournissant de nombreuses données en temps réel. Dans cette étude nous cherchons à estimer le taux de présence dans la maison en fonction de différentes données. Pour réaliser cette étude nous avons choisi d'utiliser les données des capteurs de CO2, de mouvement et l'état (allumé ou éteint) des différentes lumières. Nous avons décidé d'étudier le taux de présence pour chaque étage de la maison indépendamment.

L'objectif n'est pas de calculer le nombre exact de personnes présentes mais d'être le plus précis possible sur la présence ou non de personnes dans les différents étages. Cela permettra dans un second temps de mettre en place des méthodes afin de réduire la consommation en énergie et donc de participer à la sobriété énergétique de la maison.



1. Méthode

1.1 Traitement des données et création des features

Les données acquises via différents capteurs sont les suivantes :

- les taux de CO2 dans le rez-de-chaussée, dans les chambres du premier étage et dans le 2ème étage;

- l'état de fonctionnement des lumières, luminaires et LED dans le salon, la bibliothèque, la buanderie, le couloir, les escaliers, la cuisine, le garage, la salle à manger et la salle de bain du 2ème étage (1=allumé, 0=éteint);
- le mouvement dans la buanderie, l'entrée et l'escalier (1=présence détectée, 0 sinon).

L'étude porte sur une année d'acquisition de mesures, commençant le 30/01/2022 et finissant le 30/02/2023, jour de l'extraction des données. L'objectif est d'obtenir un tableau avec toutes les données commençant et finissant au même moment avec la même fréquence d'échantillonnage, ce qui est nécessaire pour appliquer la fonction `kmeans` sur Python. Les étapes nécessaires sont :

- 1) Extraire les données et supprimer les doublons grâce au code `test_connection.py` ;
- 2) Convertir les chaînes de caractères des dates en format `datetime` et ne pas garder les secondes dans le code `traitement.py` ;
- 3) Supprimer les valeurs apparaissant dans la même minute. Pour les taux de CO2 et le bruit, on prend la première apparition avec la fonction `doubleindexation` et pour les lumières et le mouvement on prend la dernière apparition avec la fonction `doubleindexationbis`;
- 4) Echantillonner les listes de valeurs: il faut les convertir en dataframe et utiliser la fonction `resample('T')` qui échantillonne toutes les minutes ainsi que la fonction `ffill` qui remplit les cases vides par la dernière valeur vérifiée. Pour le mouvement, la méthode de remplissage n'est pas la même. On remplace simplement les cases vides par des 0 avec la fonction `fillna(0)`;
- 5) Faire commencer et finir tous les tableaux à la même date à la main en supprimant directement les valeurs dans les nouveaux CSV.

1.2 Elaboration de l'arbre de décision : machine learning non supervisé

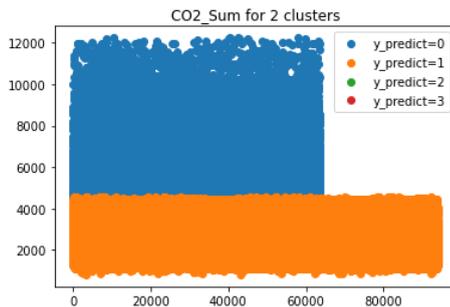
Pour répondre à notre problématique nous avons décidé d'utiliser du machine-learning en entraînant notre modèle à partir des différentes données collectées. N'ayant pas le taux de présence à chaque instant dans la maison nous avons donc dû utiliser une méthode d'apprentissage non supervisée (K-means) afin de développer un modèle de prédiction. Nous avons donc utilisé la méthode de clustering qui classe les données en différents clusters (groupes) ayant des propriétés similaires. Cette méthode permet de plus de choisir le nombre de clusters que l'on souhaite créer. Le nombre de clusters sera une variable importante dans notre étude car la précision de notre modèle y est directement lié. En effet, en modifiant le nombre de clusters nous pouvons créer un groupe qui représente les cas où il n'y a personne dans la maison avec une grande certitude.

2. Résultats

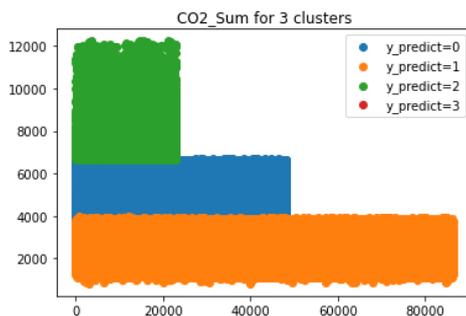
Les résultats obtenus sont présentés de la manière suivante : les valeurs prises par une propriété (par exemple le taux de CO2 dans une chambre) au cours de l'année, sont réparties en plusieurs groupes correspondant au cluster dans lequel elles ont été placées. Ainsi il est possible d'observer les ressemblances et d'interpréter la signification des clusters en les rapprochant de la notion d'occupation. La méthode K-means ressort une valeur, nommée `y_predict` qui correspond à un cluster, il est donc possible d'observer visuellement la répartition de ces clusters sur la propriété étudiée. Nous nous basons sur la moyenne du taux de CO2 afin d'identifier les différents clusters.

2.1 Résultats 1er étage

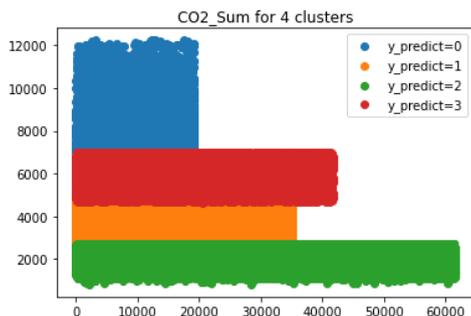
Les données prises en compte pour l'évaluation des clusters sont les taux de CO2 dans chaque chambre (chambre 1, chambre 2 et chambre 3). Ici, la propriété observée correspond à la somme des taux de CO2 dans toutes les chambres, ce qui reflète ainsi ce qui se passe à l'échelle de l'étage tout entier. La méthode K-means est testée pour 2, 3 et 4 clusters. Pour chaque cas, un cluster ne correspond pas à la même conclusion : par exemple pour le cas de 2 clusters, $y_{predict}=0$ correspond à des taux de CO2 élevés contrairement au cas pour 3 clusters. Le code *Abre_1er.py* permet d'obtenir ces résultats :



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	63487	4446	12244	6503	6358
y=1	94032	817	4575	2442	2320



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	48175	3827	6717	5463	5560
y=1	86336	817	3976	2288	2212
y=2	23008	6629	12244	7901	7825



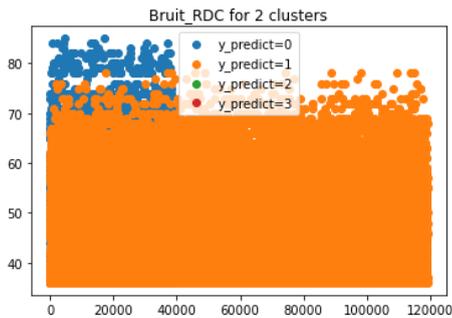
Cluster	Len	Min	Max	Moyenne	Médiane
y=0	19218	6974	12244	8113	8072
y=1	35145	2706	4860	3543	3450
y=2	61482	817	2740	1903	1839
y=3	41674	4641	7015	5879	5829

Le choix de la propriété CO2_sum semble judicieux car les clusters apparaissent clairement et donc peuvent être interprétés. Il peut être envisagé d'utiliser la méthode à 4 clusters qui semble assez précise; on peut imaginer que le cluster 2 en vert correspond à l'absence de personne au premier étage et que le cluster 0 en bleu correspond à la présence de quelqu'un. La méthode d'interprétation pour donner une valeur d'occupation correspondante à la valeur prise par $y_{predict}$ sera explicitée en troisième partie.

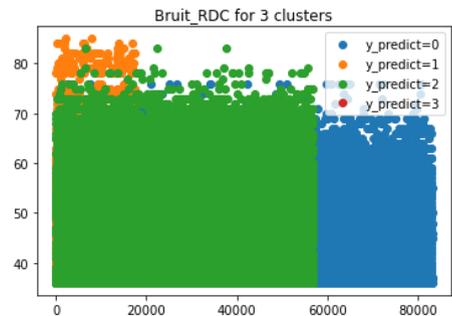
2.2 Résultat Rez-de chaussée

Dans le cas du rez-de-chaussée, deux propriétés semblent révélatrices d'une présence : le bruit et le taux de CO2. Il est possible d'observer les deux et de déterminer celle qui aidera à trouver la signification de chaque cluster. Le code *Arbre_RDC.py* permet d'obtenir ces résultats :

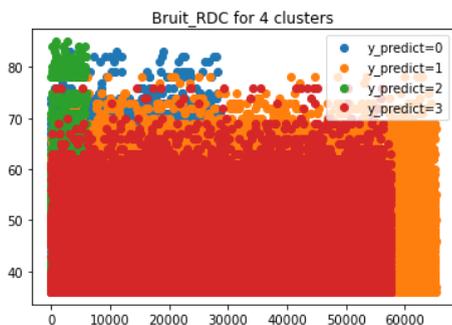
Bruit :



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	38444	36	85	46	44
y=1	118990	36	78	42	37



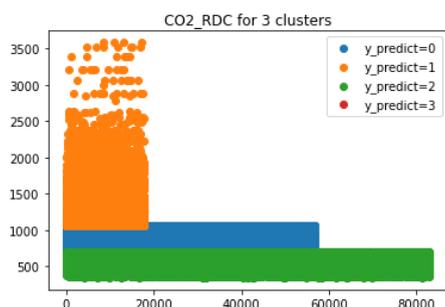
Cluster	Len	Min	Max	Moyenne	Médiane
y=0	83077	36	76	42	37
y=1	17562	36	85	46	43
y=2	56795	36	83	44	39



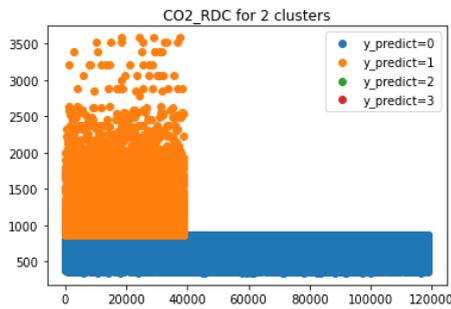
Cluster	Len	Min	Max	Moyenne	Médiane
y=0	28725	36	83	46	44
y=1	65153	36	78	43	38
y=2	5953	36	85	47	45
y=3	57603	36	76	41	37

A première vue, le bruit n'est pas une bonne propriété à observer : en effet, les clusters se superposent et il est donc difficile de les interpréter et d'en tirer une conclusion sur quel cluster correspond à quel taux d'occupation.

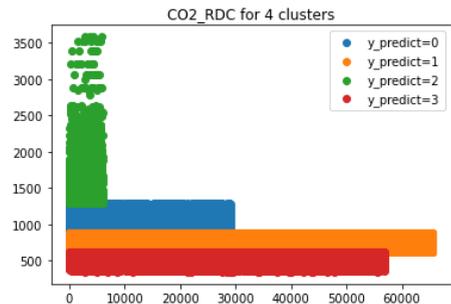
CO2 :



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	118785	345	864	621	631
y=1	38649	865	3582	1107	1044



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	56762	698	1064	841	818
y=1	17714	1065	3582	1285	1211
y=2	82958	345	698	556	558



Cluster	Len	Min	Max	Moyenne	Médiane
y=0	29158	891	1290	1048	1029
y=1	65535	621	890	733	718
y=2	5959	1291	3582	1531	1447
y=3	56782	345	620	508	507

Le taux de CO2 est une bien meilleure solution, comme il l'a été pour le 1er étage. Le modèle retenu sera celui avec 4 clusters afin d'avoir la plus grande précision. Cette fois-ci c'est le cluster 3 qui peut correspondre à une absence de personne et le cluster 2 qui semble présenter une occupation des lieux.

3. Interprétation des résultats

3.1 De cluster à taux d'occupation

La méthode K-means renvoie pour n clusters un entier compris entre 0 et n-1 qui ne correspond pas au nombre de personnes présentes dans la maison mais simplement au numéro du groupe. Pour identifier ce à quoi correspond chaque groupe nous avons décidé d'utiliser le taux de CO2 total de l'étage car il s'agit de la donnée ayant le plus d'impacte sur l'attribution des clusters. Les hypothèses étant que le groupe contenant les taux de CO2 les plus bas est le groupe représentant l'inoccupation de l'étage et les taux les plus hauts à l'occupation de celui-ci. Nous avons donc étudié les minimums, maximums et moyennes des différents clusters pour identifier visuellement à quel taux d'occupation correspondait chaque cluster. Il aurait cependant été possible d'automatiser ce processus grâce à un algorithme de comparaison des différents éléments afin de l'afficher directement sur les graphiques.

3.2 Économie d'énergie

Grâce au travail effectué lors de notre étude nous avons réussi à établir la présence ou non de personne dans les différents étages de la maison. Savoir si un étage est occupé ou non est une donnée importante afin d'avoir une maison la plus sobre énergétiquement.

Il sera donc possible dans un second temps de créer des méthodes d'automatisations afin de réduire la consommation en électricité comme une baisse du chauffage/de la température lorsqu'il n'y a

personne dans la maison. De plus, le mode veille de certains appareils n'est utile que lorsqu'il y a quelqu'un présent dans la maison, grâce à des prises connectés il sera alors possible d'éteindre des appareils comme la télé.

Il sera aussi possible de créer de nouveaux indicateurs qui aideront à rendre la maison plus intelligente. Tel qu'une routine de présence dans la maison afin d'adapter et régler les différents systèmes de manière à minimiser la consommation d'énergie tout au long de la journée. Comprendre l'impact du nombre de personnes présentes dans la maison sur les systèmes de chauffage, de consommation électrique et d'eau est nécessaire pour baisser les demandes en énergies. Ces différentes études pourront donc être effectuées dans un second temps.

Après calcul nous trouvons un taux d'inoccupation moyen de la maison de 37,5% (temps d'inoccupation de la maison/temps total des mesures). D'après l'ADEME pour chaque degré de moins de chauffage c'est en moyenne 7% d'économie. En supposant que le passage entre les différentes températures n'implique pas de surcoût, pour chaque degré de moins lorsque la maison est inoccupée, l'EXPE-SMARTHOUSE ferait une économie de 2,6% de chauffage.

4. Conclusion

Cette étude a établi un outil permettant de développer de nouveaux moyens de réduction de la consommation. Il permettra d'adapter les différents systèmes énergétiques de la maison en fonction du taux d'occupation sans jamais nuire au confort des habitants.

Par manque de temps, l'étude du deuxième étage n'a pas pu être menée. Cependant la même méthode aurait été suivie car le deuxième étage disposait des mêmes capteurs que les deux premiers : CO2 et lumières.

Afin de démontrer les résultats obtenus, il aurait été intéressant de tester en temps réel le modèle créé et d'en déduire sa précision et son taux d'erreur. De plus, pour créer un modèle le plus précis possible il aurait été nécessaire d'avoir le taux d'occupation réel sur une période test afin d'utiliser une méthode d'apprentissage supervisée.

Le modèle qui a été créé dans cette étude pourrait être amélioré afin d'obtenir le nombre précis de personnes dans la maison, voir dans chaque pièce ce qui impliquerait une gestion plus compliquée des systèmes énergétiques mais aussi plus efficace.