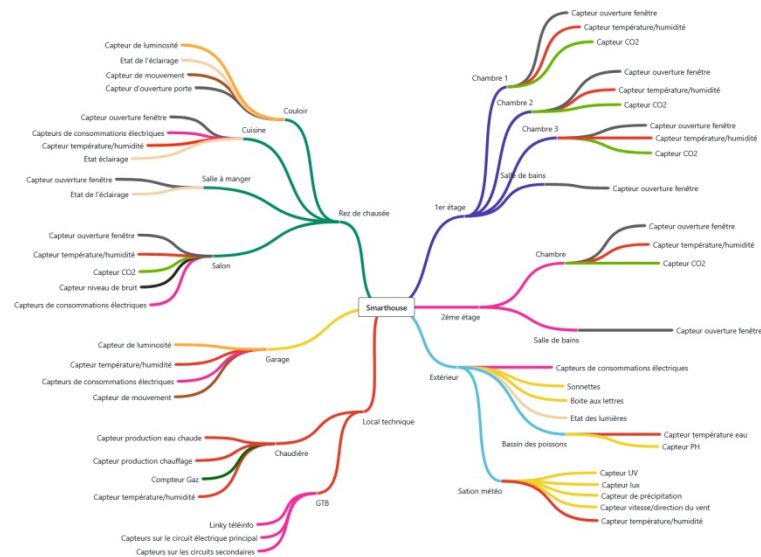


Introduction



Projet Expe-SmartHouse

Le projet Expe-SmartHouse a été lancé en 2018. Il permet l'acquisition de données provenant d'une habitation de 120m² où vivent plusieurs personnes. On peut ainsi avoir accès à plusieurs centaines de mesures en temps réel, qui sont ensuite stockées dans une base de données. Les logiciels utilisés pour ce projet sont en Open Source et permettent de mener plusieurs études différentes par plusieurs personnes différentes. Dans le cadre de notre projet en gestion d'énergie, nous allons donc pouvoir choisir un aspect à étudier et approfondir, et nous pourrions baser nos modèles sur des données réelles issues du projet Expe-SmartHouse.

Problème étudié

En observant les données mises à disposition par la plateforme Expe-SmartHouse, nous avons remarqué que les valeurs concernant le facteur de puissance dans l'habitation étaient très variables, passant même par moment en dessous de 0,4. L'objectif que nous avons choisi pour ce projet est donc de comprendre d'où viennent ces variations et par la suite de trouver une méthode pour les limiter et ainsi obtenir un meilleur facteur de puissance moyen, surtout pendant la nuit (période où ce facteur est le plus faible en général). Le fait d'améliorer le facteur de puissance peut être utile pour réduire sa facture d'électricité et diminuer les pertes. Voici un exemple des variations que peut connaître le facteur de puissance au cours du temps



Figure 1 : Variations du facteur de puissance au cours de 6 jours consécutifs

Pour parvenir à un modèle concluant, nous disposons de différents capteurs dans la maison, qui nous permettent d'avoir connaissance de la puissance apparente et réactive consommée, du facteur de puissance, et de la puissance consommée par divers appareils électriques de la maison.

Démarche prévisionnelle

Dans un premier temps, l'objectif sera de déterminer, à partir des données de chaque appareil, si un ou plusieurs appareils en particulier sont responsables des changements de facteur de puissance (ou au moins en fort lien avec). Ceci nous permettra, si l'on obtient un résultat, de cibler notre action sur ce même appareil et de prévoir à quels moments de la journée son utilisation sera la meilleure vis-à-vis de la puissance active/réactive.

Pour réaliser cette prévision, nous établirons un modèle de régression à l'aide de la librairie sklearn de python, qui nous permettra de prévoir les évolutions du facteur de puissance. Ainsi, il sera possible de savoir quand prévenir l'utilisateur que le facteur de puissance est faible.

Finalement, nous pourrions potentiellement créer une simulation pour savoir pour quel(s) appareil(s) il sera bénéfique de décaler sa période d'utilisation pour maintenir le facteur de puissance le plus haut possible.

Acquisition des données

Nous avons choisi dans un premier temps de prendre les données de 11 capteurs sur une durée d'un mois. Nous avons choisi les capteurs : de courant, de tension, de puissance apparente, de puissance réactive, et de facteur de puissance. Puis des capteurs donnant la puissance consommée

par les différents appareils de la maison : la TV, le four, les plaques à induction, machine à laver, lave-vaisselle, robot ménager.

Nous importons ces données au format csv, à partir de la plateforme jeedom, afin de pouvoir les manipuler plus facilement en python. Avant de les utiliser, nous devons traiter ces données, afin de supprimer certaines valeurs aberrantes et d'harmoniser les heures de mesure pour tous les capteurs. Tout d'abord, nous créons un fichier csv par grandeur mesurée. Ensuite, nous devons compléter les valeurs manquantes par interpolation et on peut également remplacer les valeurs aberrantes par des valeurs interpolées. On obtient alors une série de données normalisées et plus facilement exploitables. Nous avons choisi d'utiliser notre modèle avec une valeur toute les 30 secondes pour chaque grandeur. Nous avons donc effectué un resampling

Lors de la mise en forme des données, on a rencontré des difficultés car certains capteurs ne fournissent aucune donnée lorsque l'appareil correspondant est éteint pendant une certaine période. Grace à pandas on peut remplir ces périodes avec des 0, mais on ne peut pas le faire si la période sans données est à la fin ou au début. On obtenait donc des listes qui s'arrêtaient un ou plusieurs jours avant les autres, car le capteur n'a pas transmis de donnée sur les derniers jours ou que les données n'étaient pas stockées dans la base de données. De plus, la correction de certaines valeurs (comme la puissance consommée par le four) est complexe, car elles sont nulles la plupart du temps, et varient brusquement d'un coup. Il est donc difficile de détecter les valeurs aberrantes. Nous avons donc ajouté manuellement la date de fin et/ou de départ sur chaque enregistrement, précédée d'un 0 ou de la valeur cohérente afin que via le resampling on obtienne des listes de même taille avec une valeur toutes les 30s.

Estimation via apprentissage supervisé

On possède les données du capteur de facteur de puissance, on peut donc essayer d'adopter un modèle d'apprentissage supervisé. De plus, on veut prédire une valeur continue, donc on décide d'adopter un modèle de régression. Notre but est d'analyser l'influence des valeurs issues de chaque capteur sur la valeur prédite. Pour la régression, on utilise un arbre de régression et la régression linéaire de sklearn.

On décide également de procéder à une classification, via random forest, en multipliant les valeurs par 100 et en les considérant comme entières, ce qui crée un grand nombre de classes. On pourra comparer par la suite les résultats de ces deux méthodes. La classification paraît certes moins adaptée à ce genre de tâches, mais nous pensons qu'elle peut être utile pour mieux comprendre l'influence de chaque paramètre. Nous verrons la précision que cette méthode nous permet d'atteindre.

Dans les deux cas, on sépare nos données entre entraînement et test. On considère 70% des données totales pour l'entraînement et les 30% restants pour le test. On construit donc des listes Python de test et d'entraînement pour chaque grandeur mesurée. On peut ensuite appliquer les différentes méthodes d'apprentissage supervisé de scikit learn. On a choisi l'apprentissage supervisé car on disposait des labels et que la précision de ces méthodes est supérieure à celle des méthodes d'apprentissage non supervisé.

Importance de chaque paramètre

Tout d'abord, on vérifie que l'allure du facteur de puissance réel correspond bien à ce que l'on observait via Grafana, puis on test nos modèles :

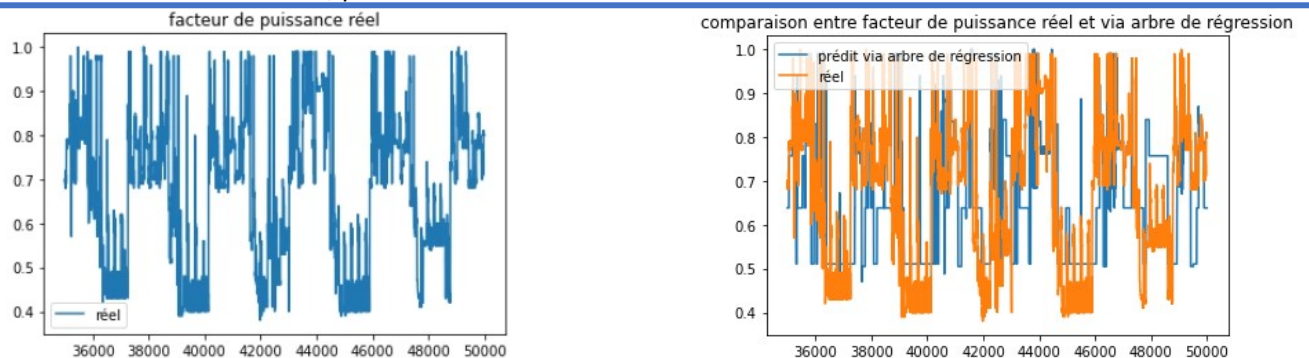
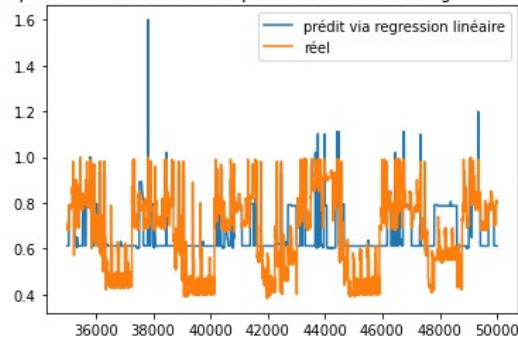


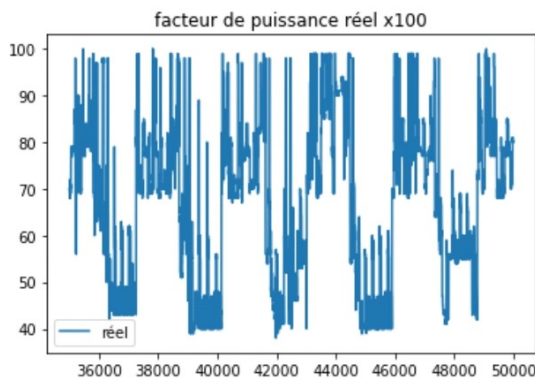
Figure 2 : Méthodes de régression

comparaison entre facteur de puissance réel et via régression linéaire



On remarque que la régression linéaire semble limitée à une valeur minimale de 0.6, mais que la périodicité des variations a tout de même été modélisée.

On a des résultats satisfaisants pour l'arbre de régression



comparaison entre facteur de puissance réel et via random forest

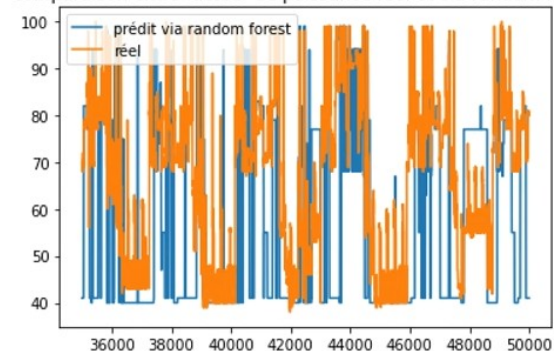
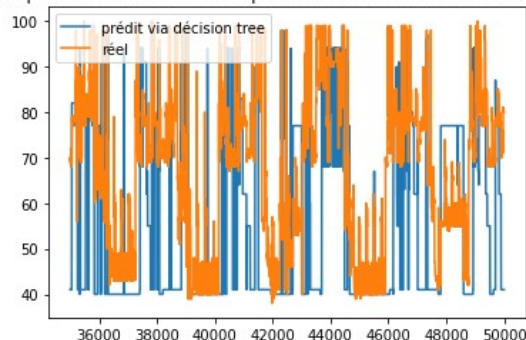


Figure 3 : Méthodes de classification

comparaison entre facteur de puissance réel et via arbre de décision



Pour la classification, on a multiplié les valeurs par 100 afin de créer des classes distinctes.

On voit que les deux méthodes fournissent des résultats cohérents et nous pourrions les analyser au regard des performances des méthodes de régression.

Nos valeurs semblent globalement correspondre à l'évolution d'un facteur de puissance au cours du temps. Pour juger de l'efficacité de nos modèles, on se base essentiellement sur l'erreur moyenne entre la prédiction et le réel. Cet indicateur nous paraît plus pertinent qu'une précision stricte et plus précis qu'une précision à x% près.

On obtient :

Tableau 1: Erreurs moyennes des différentes méthodes

	Modèles de classification		Modèles de Régression	
Modèle	Random Forest	Arbre de classification	Arbre de régression	Régression linéaire
Erreur moyenne	10.073	12,584	8.623%	9,491

Au vue de ces résultats, on peut considérer que nos différents modèles, sont assez précis pour refléter l'influence de chaque paramètre sur le facteur de puissance. On va désormais déterminer l'influence de chaque grandeur physique mesurée sur la grandeur qui nous intéresse : le facteur de puissance. Pour ce qui est de la régression linéaire, nous avons vu que même sur le jeu d'entraînement la valeur minimale est fixée légèrement inférieure à 0.6. Nous ne voyons pas ce qui peut expliquer ce comportement. Toutefois, les variations sont assez bien reflétée si on fait abstraction de la valeur minimale. Nous allons tout de même éliminer ce modèle au vue des résultats obtenus, afin d'avoir des courbes représentatives du niveau réel du facteur de puissance...

Analyse des résultats :

On voit que certains paramètres ont une forte influence sur notre modèle : le courant et l'intensité. Toutefois, ces paramètres sont le résultat de l'utilisation des autres appareils. Ils impactent le facteur de puissance mais on ne peut pas agir directement sur eux. Nous avons donc décidé de relancer notre modèle sans prendre en compte ces données, afin de déterminer quelle est la cause des variations importantes du facteur de puissance.

Cette fois-ci, on obtient :

Tableau 2: Erreurs moyennes des différentes méthodes sans le courant et la tension

	Modèles de classification		Modèles de régression
Modèle	Random Forest	Arbre de classification	Arbre de régression
Erreur moyenne	15,282%	16,252%	11,167%

La erreur moyenne a légèrement augmenté, mais elle reste acceptable pour notre analyse. On utilise désormais des fonctions permettant d'évaluer l'importance de chaque paramètre initial dans la prédiction finale. Cela nous permettra de savoir pour quels appareils il serait préférable de changer le mode d'utilisation afin d'augmenter le facteur de puissance moyen. On trouve alors :

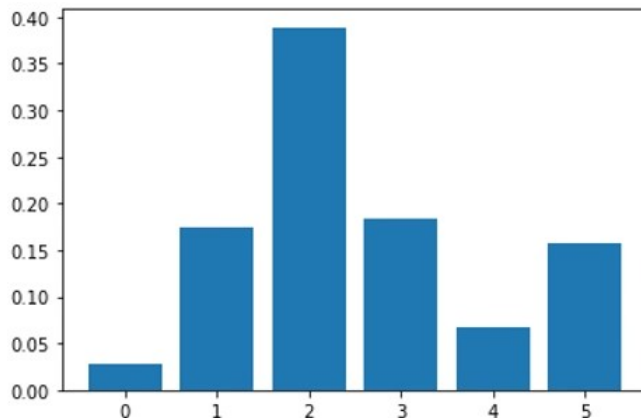


Figure 4 : Importance des différents paramètre pour Random Forest

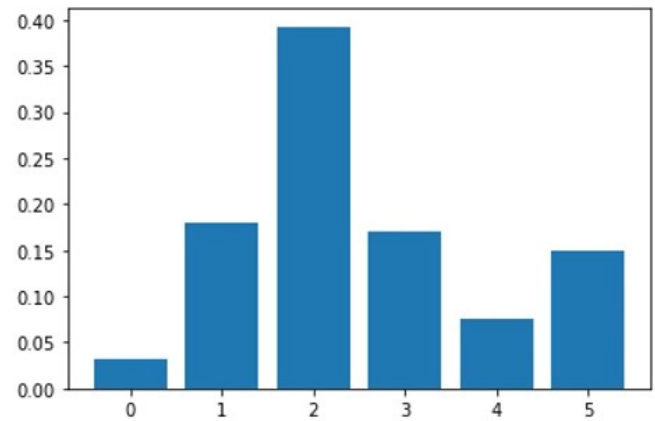


Figure 5 : Importance des différents paramètre pour un arbre de classification

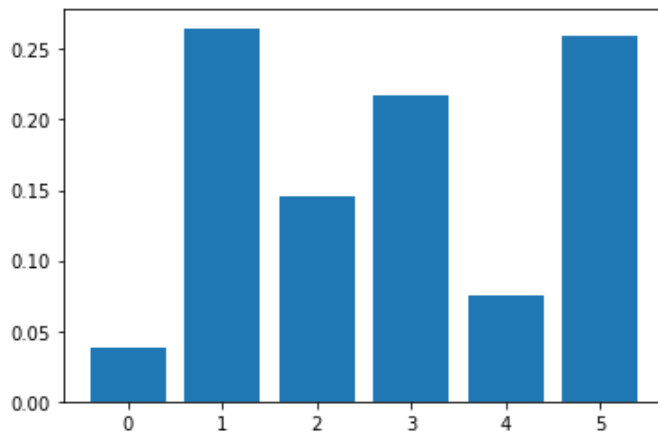


Figure 6 : Importance des différents paramètre pour un arbre de régression

Légende :

Feature n°	Paramètre représenté
0	Four
1	TV
2	Lave-linge
3	Lave-vaisselle
4	Plaques de cuisson
5	Robot ménager

On remarque donc que quelque soit le modèle, le four et les plaques de cuisson jouent un rôle mineure, ce qui peut s'expliquer par le fait qu'ils sont peu utilisés par rapport aux autres objets. Ils ne peuvent donc pas être responsables de la majorité des variations.

De plus, en analysant les courbes, on remarque qu'il existe une fluctuation périodique : un pic toutes les deux heures environ pendant 20min environ. Ce pic apparaît également dans la puissance consommée globale, mais dans aucune des consommations des appareils que l'on étudie.

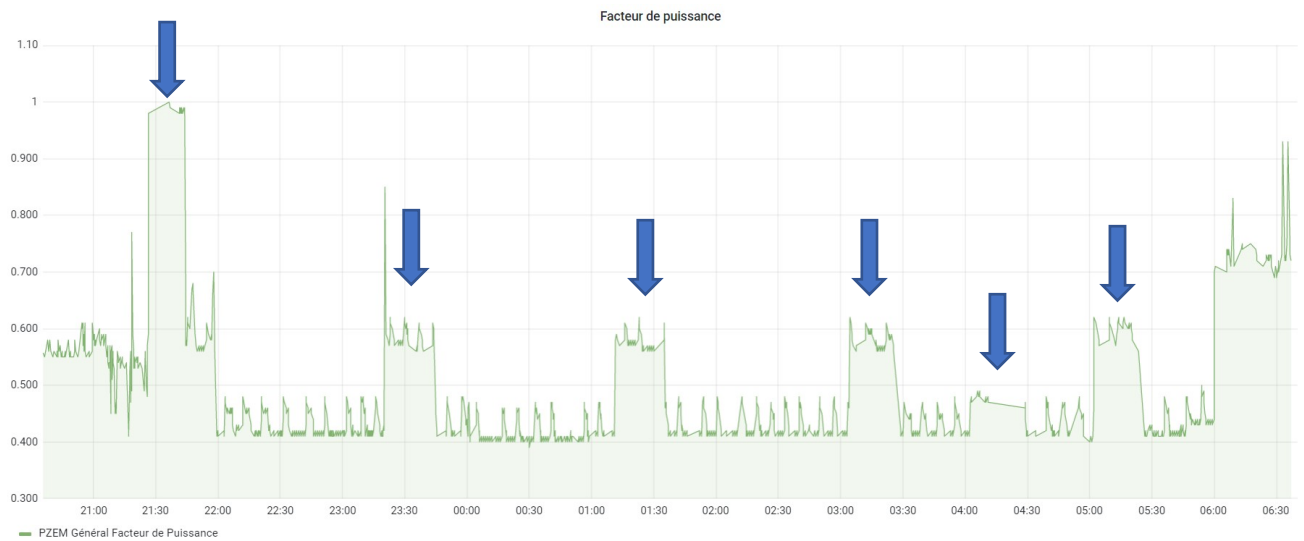


Figure 7

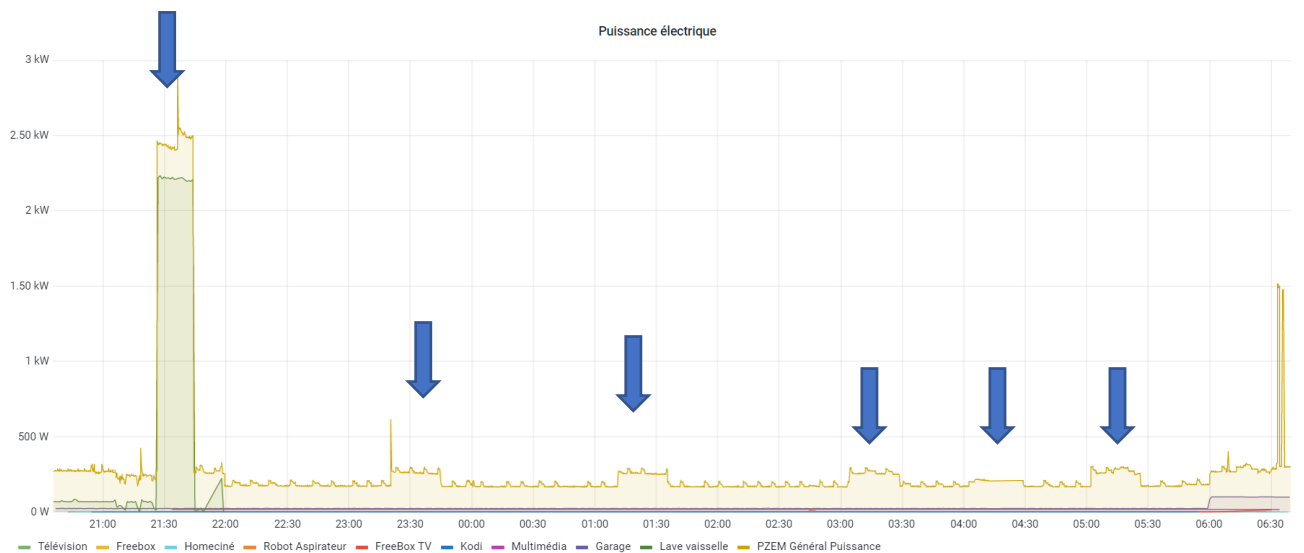


Figure 8

On peut faire l'hypothèse que ces piques sont dus au réfrigérateur, qui ne fait pas partie des objets pour lesquels nous avons des données. Le fait qu'un élément extérieur ait une influence non négligeable sur le facteur de puissance peut expliquer une partie de l'erreur de nos modèles.

Toutefois, si on s'intéresse aux données fournies par nos modèles, on voit que le TV, le lave-vaisselle, le lave-linge et le robot ménager jouent un rôle important (même si cette importance varie selon les modèles) dans les variations du facteur de puissance. Pour les méthodes de classification, on voit que le paramètre le plus important est le lave-linge alors que pour la régression, les paramètres sont plus égaux. Si on essaie de voir les heures d'utilisation de ces différents appareils sur grafana, on a pour une semaine :

Lave-linge :

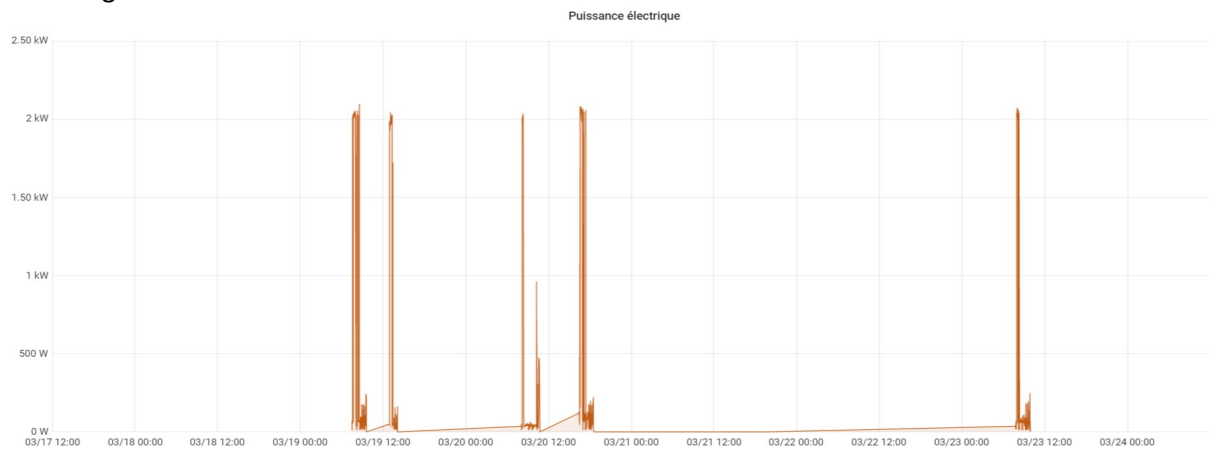


Figure 9

Lave-vaisselle :

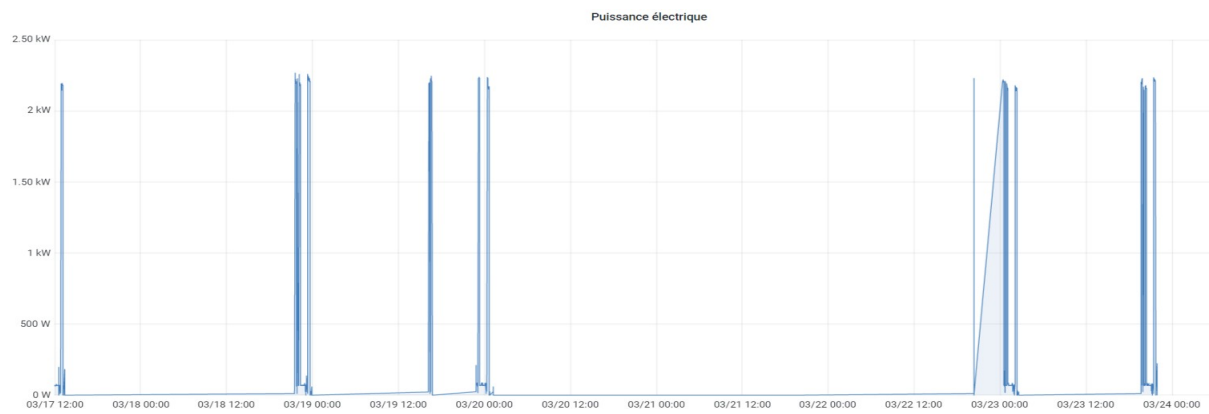


Figure 10

Robot ménager :

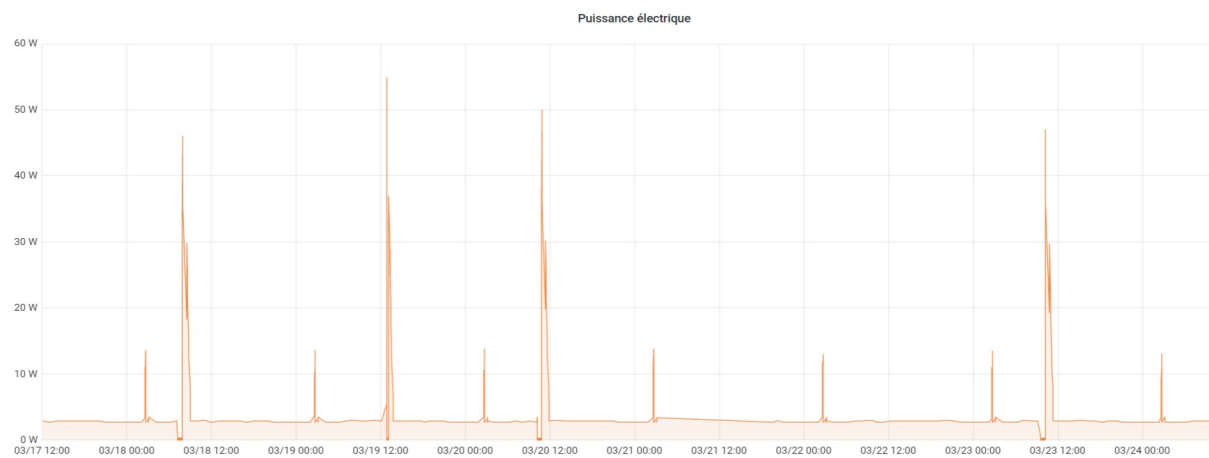


Figure 11

TV :

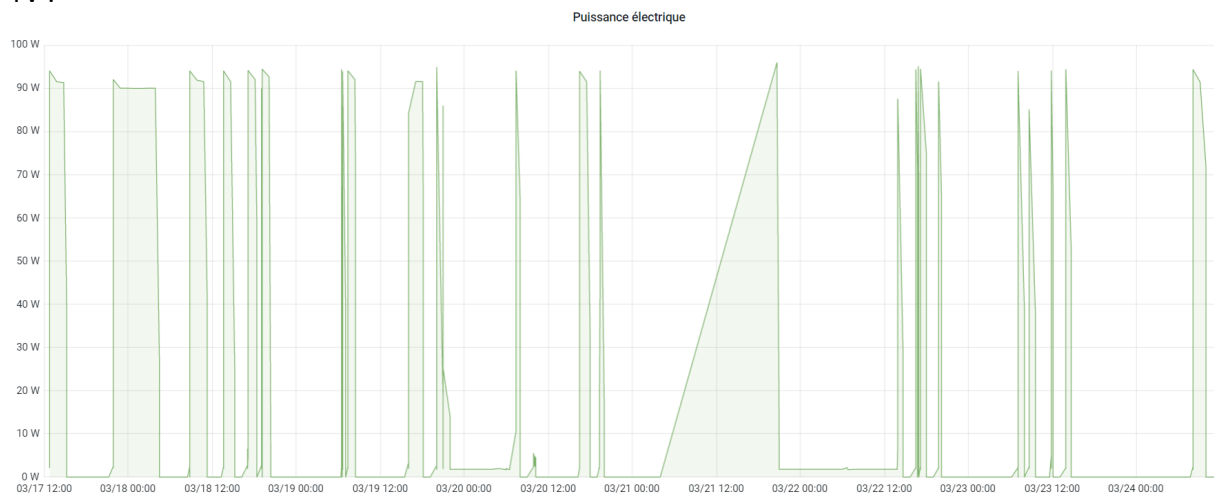


Figure 12

Via analyse de ces courbes et de celle du facteur de puissance sur la même période, on peut déduire plusieurs informations :

- L'allure créneau marquée par l'alternance jour/nuit n'est pas due à un appareil particulier mais à la baisse de consommation la nuit.
- Le facteur de puissance peut être augmenté la nuit si il y a plus d'utilisation des appareils(ex lave-vaisselle/lave-linge).
- Il sera difficile d'augmenter grandement le facteur de puissance en moyenne sans augmenter la consommation globale.

On a donc une meilleure idée du lien entre l'activité dans la maison et le facteur de puissance.

Conclusion

Retour sur notre problème

Pendant ce projet, nous avons pu nous attarder sur l'évolution du facteur de puissance global de la maison au cours du temps. Nous avons pu mettre en évidence certains paramètres, et nous avons pu trouver de nouvelles questions. Nos modèles étaient assez précis pour refléter l'impact des différents appareils sur le facteur de puissance. Cependant, une erreur non négligeable demeurait et nous avons cherché à identifier sa provenance.

On a donc voulu tester notre modèle en essayant de prédire la puissance consommée ($P = \sqrt{S^2 - Q^2}$) à partir des mêmes paramètres. Nous avons pu voir que le résultat obtenu pour cette prédiction est très proche de la réalité (erreur moyenne <5% pour tous les modèles).

En revanche, nous avons pu voir que notre prédiction était beaucoup moins fiable pour la puissance réactive(erreur moyenne >20% voir 25%). On peut donc supposer que l'erreur que nous avons sur le facteur de puissance provient en grande partie de la puissance réactive.

L'autre élément qui a pu augmenter l'erreur est le fait que nous ne disposions pas des données pour le frigo. Nous avons tout de même pu identifier les pics engendrés par celui-ci : 20min toutes les 2h environ.

Ensuite, nous avons pu constater que quatre des appareils que l'on étudiait avaient une réelle influence sur les facteur de puissance : TV, lave-vaisselle, lave-linge et robot ménager. Leur utilisation permet d'augmenter ponctuellement le facteur de puissance. Le problème étant la faible valeur de celui-ci pendant la nuit, on peut donc recommander de programmer les lave-linge et si possible le lave-vaisselle et le robot pour des heures de fonctionnement nocturnes.

Toutefois, nous avons constaté que la forme créneau globale de facteur de puissance était due principalement à la baisse d'activité la nuit. Donc les pertes engendrées par la chute du facteur de puissance interviennent lorsqu'il y a peu de consommation et donc n'augmentent pas significativement la facture d'électricité. A l'inverse, la valeur du facteur de puissance est élevée en période de forte consommation. Il n'y a donc que peu de gains à faire puisque augmenter la consommation la nuit améliorerait certes le facteur de puissance mais augmenterait aussi la facture globale. La seule amélioration serait donc de décaler des usages existants, en particulier pour les quatre appareils cités ci-avant.

D'un point de vue plus technique, on a pu constater que (comme attendu) la régression est bien plus adaptée à ce genre de tâche que la classification, car en créant un grand nombre de classes on diminue la précision des algorithmes de classification (à quantité de données d'entraînement égale).

Ce que l'on aurait pu faire

Pour valider notre modèle et aller plus loin, on aurait pu simuler l'allure du facteur de puissance avec une utilisation décalée de certains appareils. Mais la précision de notre modèle ne donne pas de résultats significatifs pour un seul appareil en utilisation décalée.

On aurait également pu essayer de rajouter d'autres facteurs dans notre étude comme par exemple l'occupation globale de la maison, ou identifier la pièce de la maison où les activités ont le plus d'impact sur le facteur de puissance

Du point de vue de la simulation, on aurait pu prendre d'autres modèles plus complexes de régression (non linéaires par exemple) ou essayer de prendre des données plus anciennes pour entraîner notre modèle et comparer les différents résultats plus précisément.

On aurait également pu essayer de retirer la composante que l'on attribue au réfrigérateur via un « filtrage » sur les valeurs du facteur de puissance afin d'entraîner notre modèle sur les seules fluctuations dues aux données que l'on a.

Ce que l'on retient

On a pu réaliser l'importance et la puissance des algorithmes de machine learning dans la prévision et la gestion de la consommation d'énergie. Ce projet nous a permis d'étudier un cas concret et de mieux comprendre la chaîne de fonctionnement allant du choix des capteurs au choix des modèles informatiques et à l'analyse des résultats en passant par le stockage et la mise en forme des données.