Proposition de Projet

ECO-ENERGY

Surveillance des appareils électroménagers à partir des lectures de la consommation d'énergie avec ML et DML

Auteurs :

Tristan Brunet (`ML Dev.`) 2652139@collegelacite.ca
Federic Bapoungue (`ML Dev. `) 2717680@collegelacite.ca
Georges Lemuel Balogog M.(`ML Dev. `) 2717615@collegelacite.ca

Table des matières

Table des figures

Liste des tableaux

1 Business Case

Présentement, Éco Énergie aide ses clients à réduire leur consommation d'énergie (mesurée en watts (W)) à l'aide de moniteurs (et des reseaux de capteurs) de consommation d'énergie connectés à chaque appareil consommateur. Par contre, ces clients n'ont pas de manière de recevoir un suivi sur leur consommation, seulement la lecture en temps réel. Notre solution, basée sur le "machine learning", permettrait à Éco Énergie d'offrir une solution à ces problèmes pour ses clients.

2 Problème et opportunités

Pour élaborer sur les problèmes rencontrés par les clients de Éco Énergie, la consommation est présentement mesurée par des moniteurs de consommation, et ces lectures sont stockées sur une carte informatique connectée à ces lecteurs. Par contre, nous n'avons pas de suivi sur la consommation générale. Pour résoudre ce problème, il nous faudrait une façon de visualiser nos données, ou une façon de détecter si notre consommation à un certain moment est normale, ou si elle est anormale.

3 Buts et objectifs

Notre solution proposée est basée sur le machine learning afin de classifier si la consommation à un moment donné est normale ou anormale. En analysant la consommation sur le circuit d'une résidence à travers le temps, on peut déterminer si la consommation courante est normale ou non.

La Fig. 1 est un exemple d'un graphique de consommation d'énergie. Nous pouvons observer les augmentations et diminutions de consommation d'énergie. L'intelligence artificielle serait donc capable d'observer si certaines données dans ce graphique sont normales ou anormales. Bien sûr, nous aurons beaucoup plus de données, ceci n'est qu'un exemple (échantillon).

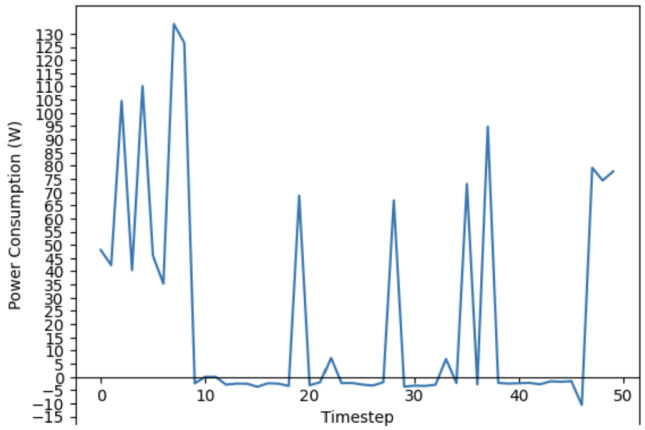


Figure 1 – Echantillon consommation de la puissance par unite de temps (timestep)

4 Environnement et contexte

Éco Énergie est une compagnie qui aide non seulement ses clients à réduire leur consommation d'énergie, mais également à mieux la comprendre. Ceci aide ses clients à mieux comprendre et contrôler leur consommation d'énergie.

Avec l'arrivée du télé-travail, nous consommons de plus en plus d'énergie, et il est facile de ne pas remarquer l'augmentation de consommation et d'être surpris par les factures d'électricité.

Présentement, les données énergétiques sont stockées localement sur une carte informatique et, à moins que les résidents mesurent la consommation pour chaque appareil, il peut être difficile de bien comprendre leur consommation et d'identifier les sources de consommation excessive. Pour régler ce problème, il leur manque une solution en temps réel qui donne une rétroaction aux utilisateurs (Fig. 2).

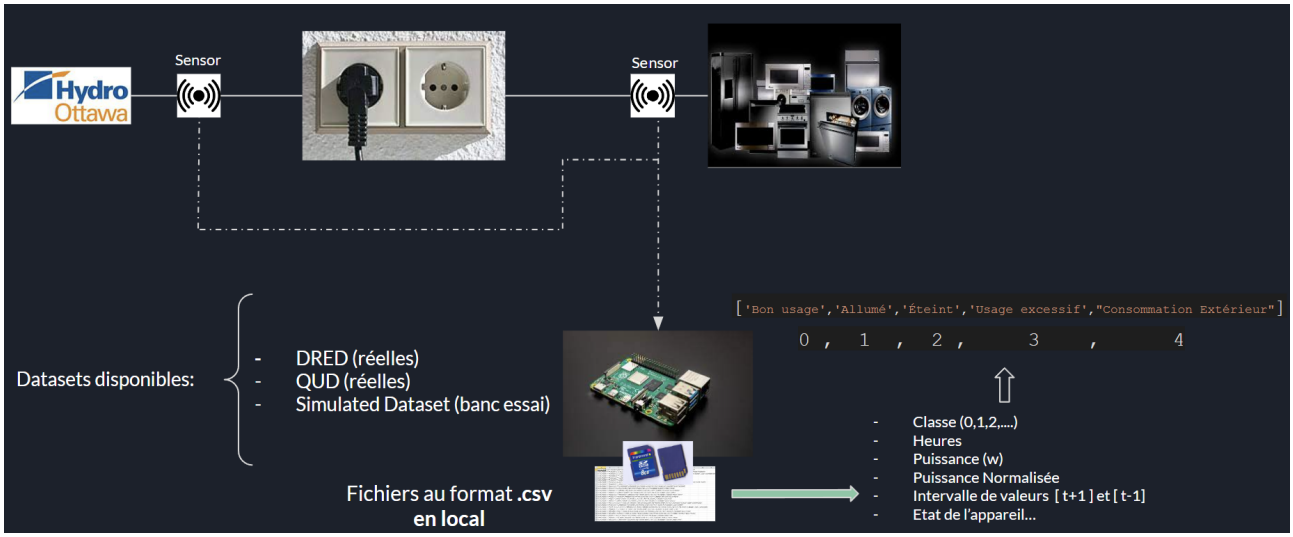


Figure 2 – Etat des lieux

Notre solution permettra à Éco Énergie qui souhaite offrir une rétroaction à ses clients avec une exactitude d'au moins 90% surtout pour les classe indiquant une consommation anormale, le tout déployé sur une carte informatique suffisamment puissante pour gérer un modèle déployé pour effectuer des prédictions à chaque lecture de consommation d'énergie.

5 Description de haut niveau de la solution

Comme mentionné plus haut, notre solution pour résoudre ce problème sera basée sur le Machine Learning. Plus précisément, nous allons créer une solution basée sur le "Machine Learning" simple et une solution basée sur la "Deep Machine Learning", soit des réseaux de neurones. Ci-dessous est un graphique représentant la solution de haut niveau

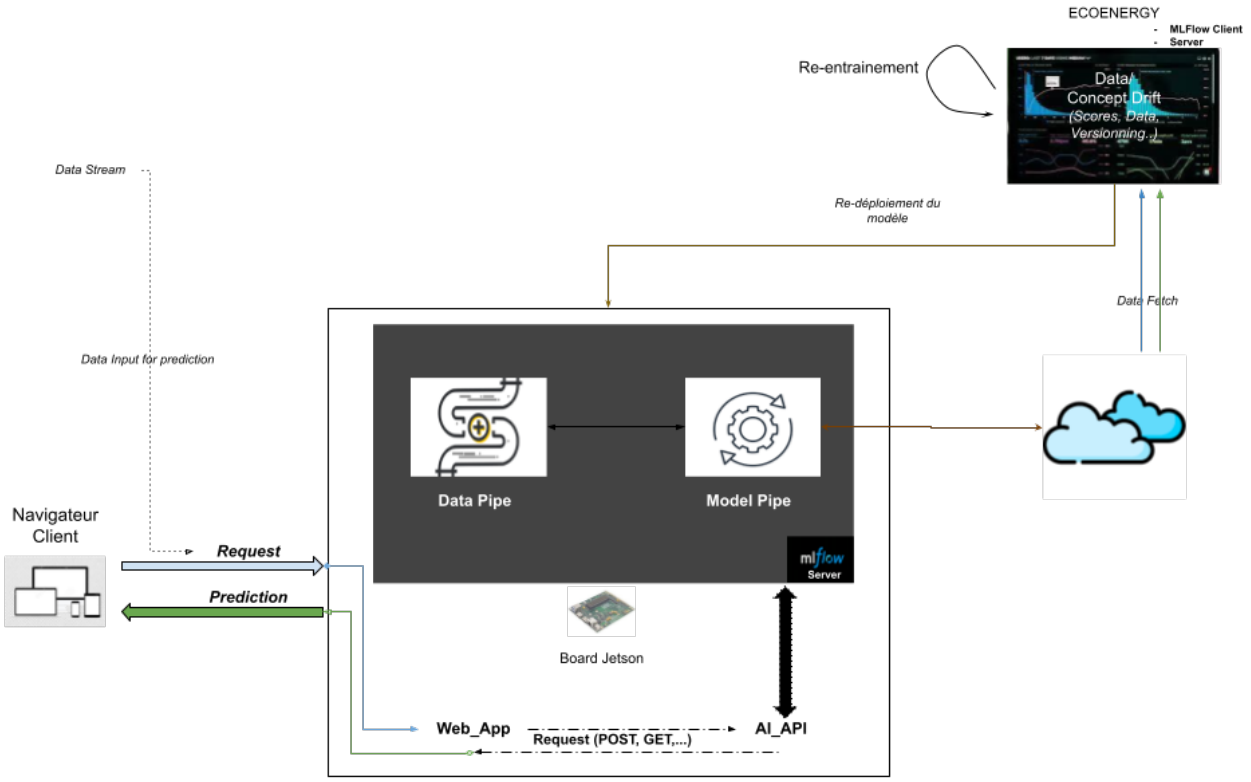


Figure 3 – Solution de haut-niveau

Les moniteurs de consommation vont fournir les données à la carte Jetson par le réseau local. Ces données seront fournies selon une fréquence spécifique, et sera considérée comme un flux de données. Ces données seront ensuite utilisées pour effectuer la classification par notre modèle IA. Ensuite les données et la prédiction effectuée seront envoyées au serveur de Éco Énergie afin d'effectuer de l'analyse. Par contre, nous devons demander aux utilisateurs leur permissions d'envoyer leurs données pour ce faire. Nous allons ensuite pouvoir mesurer la performance de notre modèle IA.

Avec le temps, la performance du modèle va déteriorer([2], [3]). Ceci est souvent dû au fait que les données sur lesquelles nous effectuons les prédictions divergent du dataset d'entraînement avec le temps, ce qui mène à une baisse de l'exactitude des prédictions. En effectuant la collecte des données, nous pourrons déterminer quand la performance du modèle n'est plus satisfaisante([1]).

Lorsque cette situation se produit, nous devons analyser la cause du problème. Parfois, un ré-entraînement du modèle sur de nouvelles données plus récentes sera suffisant pour que la performance soit de nouveau acceptable. D'autres fois, nous devons refaire l'architecture du modèle afin de résoudre ce problème.

Une fois que le nouveau modèle est ré-entraîné ou refait, nous allons pouvoir le re-déployer

	Prédiction				
	Bon usage '0'	Allumé '1'	Éteint '2'	U. excessif '3'	C. Extérieur '4'
Bon usage '0'	$TP_{1,1}$	$FP_{1,2}$	$FP_{1,3}$	$FP_{1,4}$	$FP_{1,5}$
Allumé '1'	$FP_{2,1}$	$TP_{2,2}$	$FP_{2,3}$	$FP_{2,4}$	$FP_{2,5}$
Éteint '2'	$FP_{3,1}$	$FP_{3,2}$	$TP_{3,3}$	$FP_{3,4}$	$FP_{3,5}$
U. excessif '3'	$FP_{4,1}$	$FP_{4,2}$	$FP_{4,3}$	$TP_{4,4}$	$FP_{4,5}$
C. Extérieur '4'	$FP_{4,1}$	$FP_{4,2}$	$FP_{4,3}$	$FP_{4,4}$	$FP_{5,5}$

Table 1 – Matrice de confusion

par internet aux cartes informatiques, qui sont connectées au réseau des utilisateurs. Cette opération devrait être presque transparente aux utilisateurs.

6 Mesures du rendement

Pour les métriques à utiliser, nous pensons utiliser le F1 score(1), qui est une métrique qui prend en compte les false positives et les false negatives (vus sur la matrice de confusion). Cette décision n'est pas fixe, il est possible de changer de métrique principale plus tard dans le cycle de développement ou intégrer un ensemble de métriques toutes a suivre.

$$F1 = \frac{2 \sum_{i=1}^{N} w_i \cdot p_i \cdot r_i}{\sum_{i=1}^{N} w_i \cdot (p_i + r_i)} \tag{1}$$

Où :
- N est le nombre de classes
- p_i est la précision pour la classe i
- r_i est le rappel pour la classe i
- w_i est le poids pour la classe i, qui peut être utilisé pour donner plus d'importance à certaines classes par rapport à d'autres. Si tous les poids sont égaux, on peut les définir comme $w_i = \frac{1}{N}$.

7 MileStone

Etape 1 : Cloture de la phase de préparation des données
Objectif : Finaliser la preparation, et le traitement des données nécessaires pour l'entraînement du modèle de prédiction relative a la consommation des equipements electriques.
Date de début : 01/01/2023.
Date de fin : 31/04/2023.
Livrables :

1. Un jeu de données nettoyé et préparé pour l'analyse
2. Un rapport de qualité des données décrivant la qualité, la pertinence et la cohérence des données collectées.

3. Un rapport d'analyse exploratoire décrivant les principales caractéristiques des données et les résultats de l'analyse precisant un ensemble de caractéristiques sélectionnées pour l'entraînement du modèle de prédiction.

Etape 2 : Construction et entraînement du modèle de prédiction
Objectif : Concevoir et entraîner un modèle de machine learning pour prédire le risque de crédit des clients en utilisant les données préparées et analysées précédemment.
Date de début : 01/04/2023.
Date de fin : 31/07/2023.
Livrables :

1. Un modèle de prédiction du risque de crédit entraîné et testé sur un ensemble de validation.
2. Un rapport de performance décrivant la précision et la performance du modèle.

Etape 3 : Optimisation et intégration du modèle
Objectif : Améliorer la performance du modèle de prédiction en optimisant les hyper-paramètres et en intégrant des techniques de régularisation.
Date de début : 01/08/2023.
Date de fin : 15/08/2023.
Livrables :

1. Un modèle de prédiction optimisé avec des performances améliorées
2. Un plan d'intégration du modèle pour une utilisation dans le système de décision d'EcoEnergie (Fig. 3)

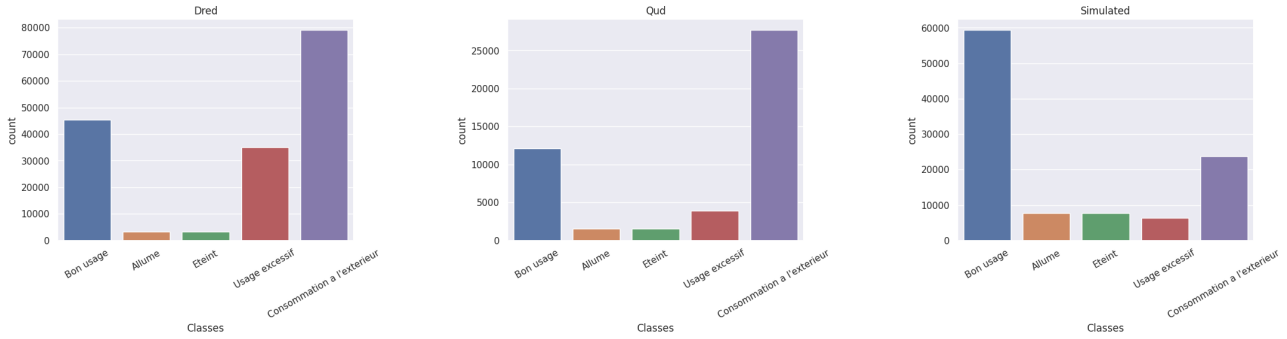
Etape 4 : Évaluation et validation du modèle
Objectif : Évaluer la performance du modèle sur un ensemble de test indépendant et valider son utilisation dans le système de recommandation d'EcoEnergie.
Date de début : 16/08/2023.
Date de fin : 30/08/2023.
Livrables :

1. Un rapport d'évaluation décrivant les performances du modèle sur un ensemble de test indépendant et les résultats de sa validation dans le système de décision de decision interne
2. Un modèle de prédiction validé et prêt à être utilisé dans le système de décision de EcoEnergie et deploye chez ses clients.

8 Résultats préliminaires

En effectuant notre exploration des données préliminaires, nous avons remarqué quelques problèmes avec le dataset que nous devrons résoudre.

Le premier problème que nous avons constaté est un débalancement des différentes classes de consommation d'énergie. Les classes signifiant *allumer* et *éteindre* l'appareil sont sous-représentés, et la classe pour "*Consommation à l'extérieur*" est sur-représentée pour DRED et QUD (Fig. 4a et Fig. 4b).



(a) **(b)** **(c)**

Figure 4 – Distribution des differentes classes dans les differentss dataset

Le deuxième problème rencontré sont les données erronées ou aberrantes. Normalement, lors de la mesure de la puissance, il n'est pas possible d'avoir une consommation plus basse que 0 watts. Par contre, puisque notre dataset est composé de données réelles, il y a quelques erreurs de lecture qui s'y sont infiltrées sous forme de valeurs de puissance négatives. On peut en observer dans le graphique ci-dessous.

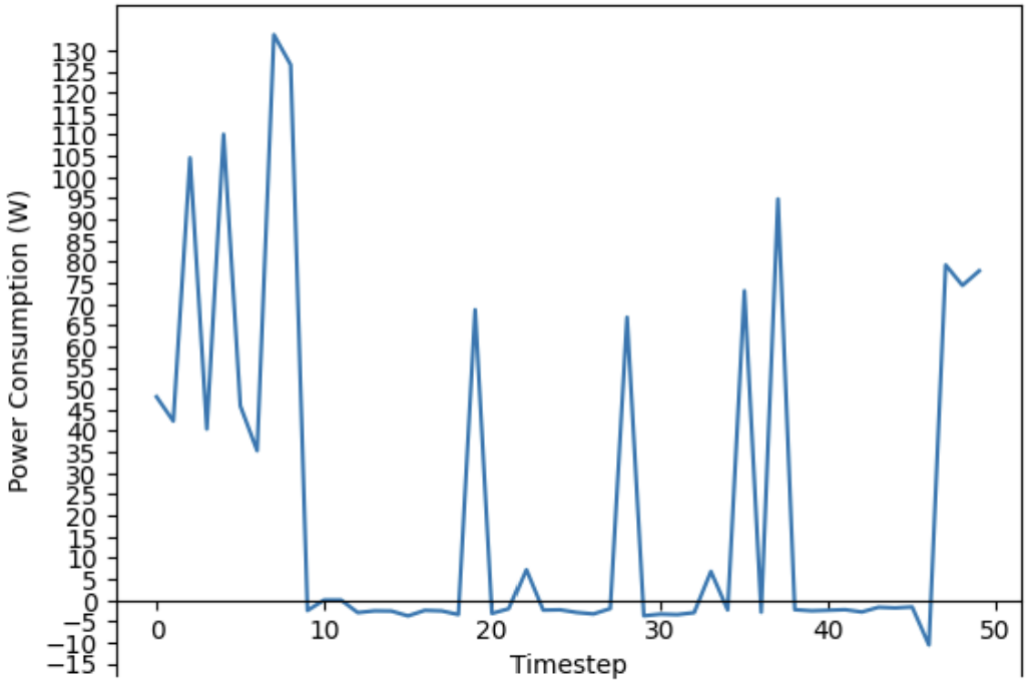
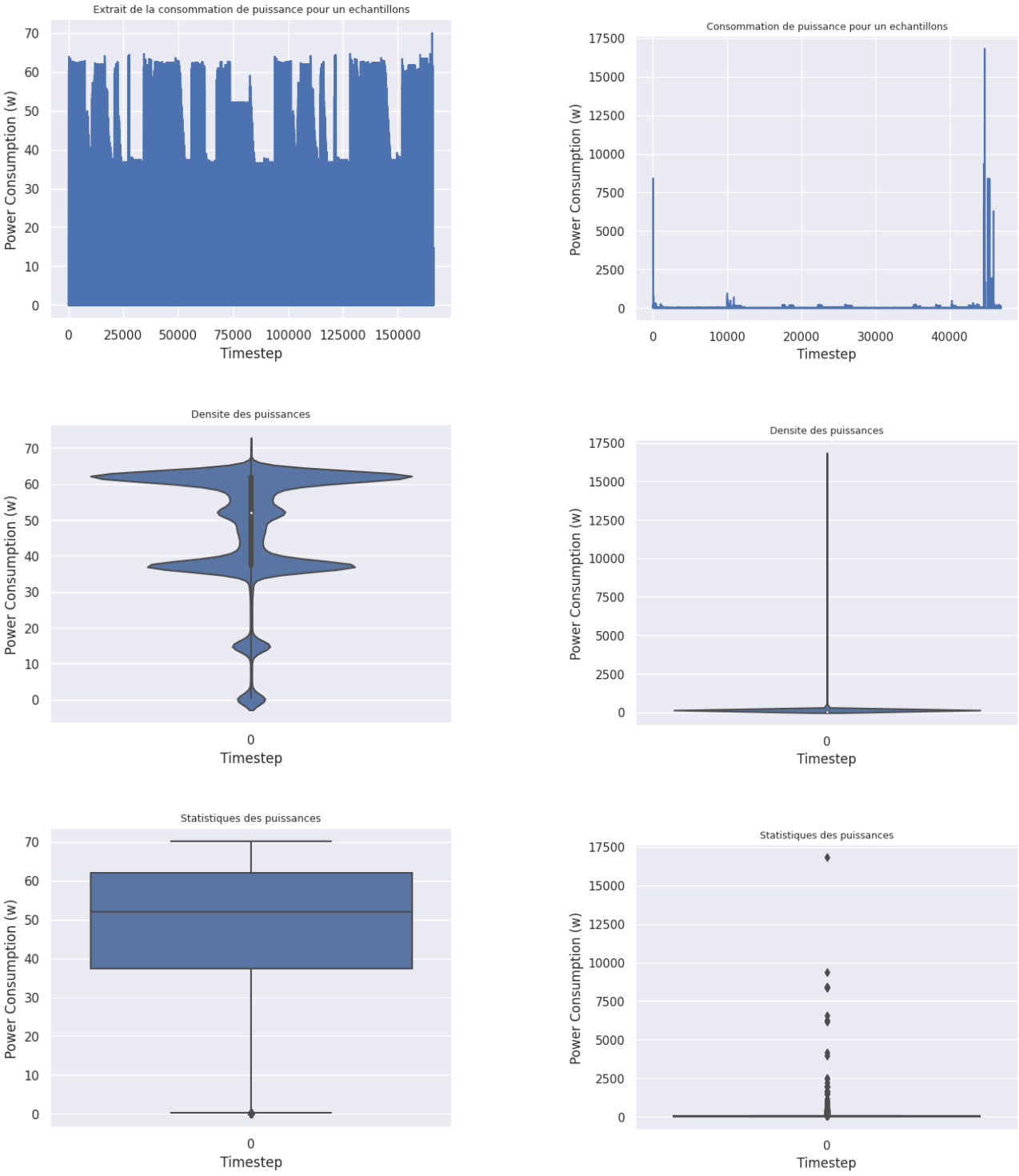


Figure 5 – Valeurs negatives de puissance dans Simulated_dataset

Nous avons également remarqué certaines données très élevées par rapport au reste du dataset. Nous pouvons considérer ces données comme étant aberrantes. Comme on peut l'observer dans le graphique ci-dessous, nous avons une lecture de consommation de plus de 8000 W, alors que le reste des lectures se trouvent entre 0 et 100 W.



(a) Dred dataset **(b)** Qud dataset

Figure 6 – Distribution des valeurs de la puissance pour les differents dataset

Le troisième problème rencontré est que deux colonnes différentes du dataset ont une corrélation parfaite entre elles. Dans notre cas, ce sont les colonnes 'Power Consumption' et 'Normalized power consumption', comme on peut voir sur les graphiques Fig.7a et Fig.7b. Dans

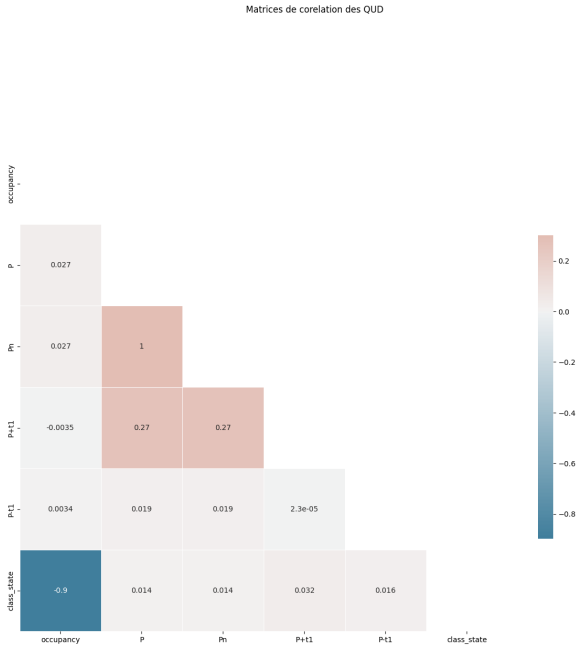
notre cas, nous avons identifié une des colonnes comme étant la **consommation d'énergie** en Watts qui a été normalisée avec la fonction (2) et rajoutée au dataset que nous avons reçu et conserver unique la *Puissance normalisee*.

$$P_n(t) = \frac{P(t) - \text{mean}(P)}{\max(P) - \min(P)} \tag{2}$$

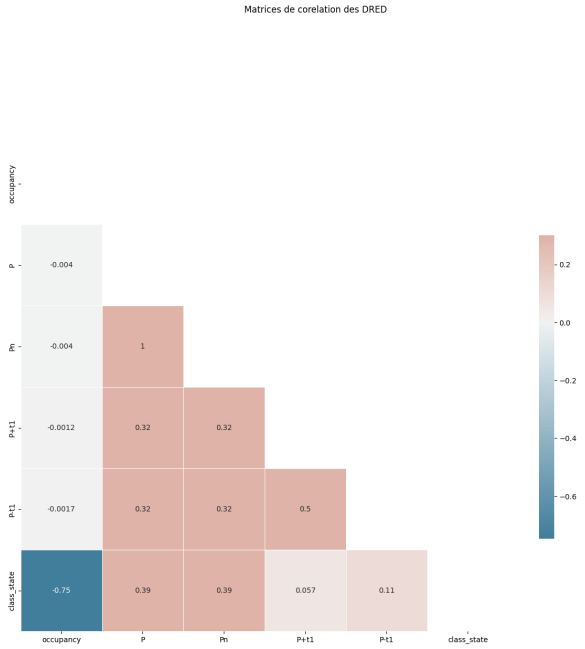
où :

— P(t) est la puissance consommée à un moment donné t,
— mean(P) est la valeur moyenne de la puissance consommée
— max(P) et min(P) représentent respectivement les valeurs maximale et minimale de la puissance consommée sur une période donnée

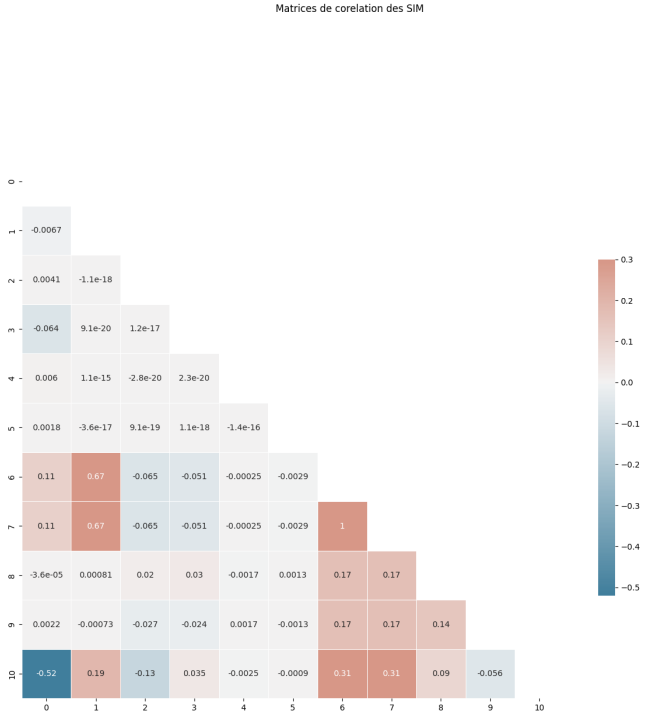
Dans ce cas-ci, nous pouvons donc **retirer la colonne de consommation d'énergie** ($P(t)$) du dataset avant l'entraînement.



(a) Matrice de corellation des colonnes du dataset DRED



(b) Matrice de corellation des colonnes du dataset QUD



(c) Matrice de corellation des colonnes du dataset SIM

Figure 7 – Correlation des differentes colonnes de chaque dataset

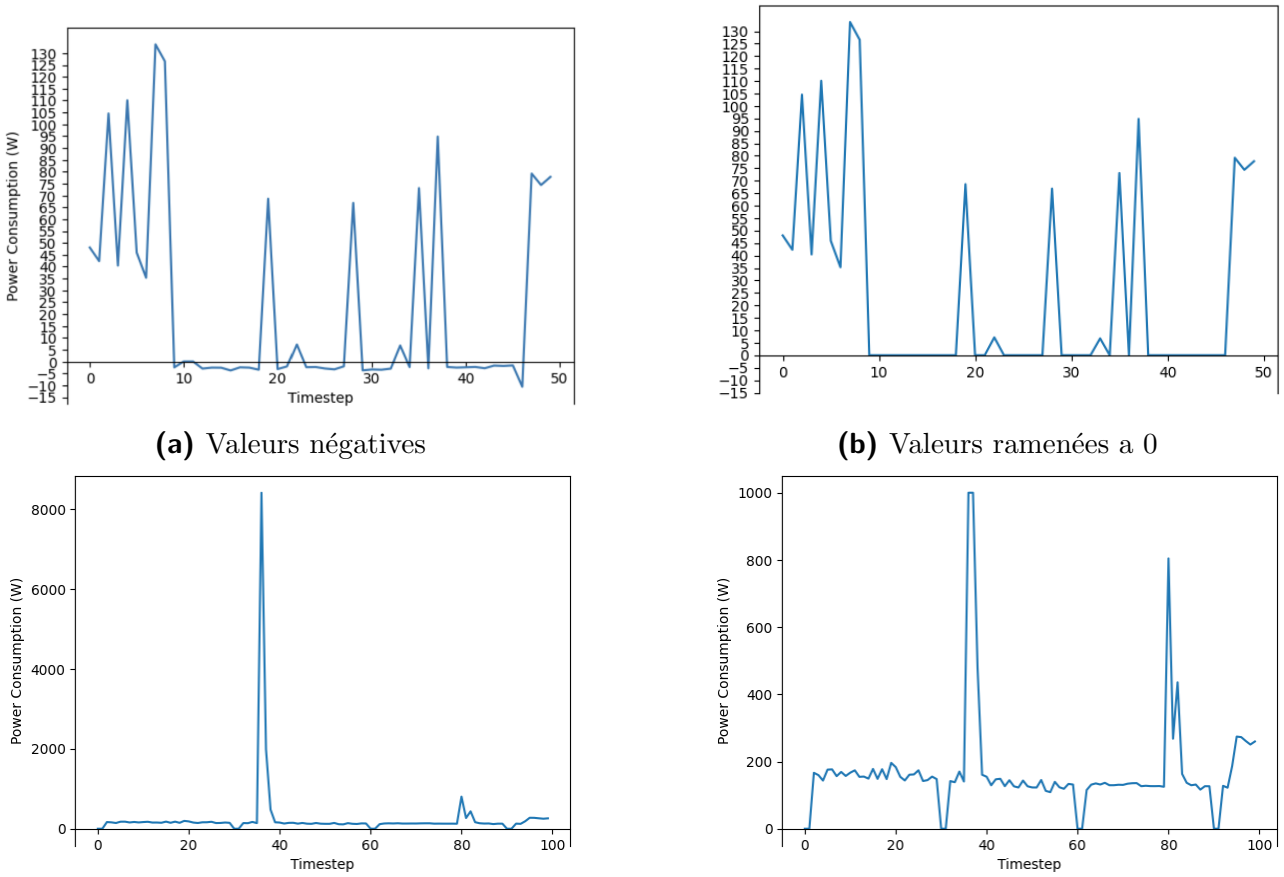
9 Conclusion et travaux futurs

9.1 Synthèse de la phase de pre-processing

9.1.1 Les données abbérantes

Probleme Quelle méthode a appliquer pour la gestion des outliers et comment les identifier.

Solution On définit un seuil/plafond, 0 pour les puissances négatives et Pmax fourni par un expert



(a) Valeurs négatives

(b) Valeurs ramenées a 0

(c) Avant Ecretage

(d) Apres Ecretage Pmax=1000 w

Figure 8 – Solutions proposées pour les Outliers

9.1.2 Echantillonage

Probleme Quelle stratégie a mettre en place pour prendre en compte les différentes périodes d'échantillonnage (relevé des données de puissance) pour chaque dataset.

Solution :

1. On va normaliser le pas d'échantillonnage (timestep) a 3 secondes (min{3 , 5} secondes qui sont respectivement les pas d'échantillonnage de QUD et DRED)
2. On va rajouter une colonne qui va mentionner le timestep utilisé dans chaque data-frame {3 | 5} dépendamment du dataset que l'on traite.

9.1.3 Sauts de lecture lors du stream de donnees

Probleme Comment gerer les “sauts de lecture” éventuels qui pourraient apparaitrent, ce qui traduirait un defaut/defaillance de la chaine d’acquisition des donnees.

Solution :

1. Le client nous garantit de la qualité de la lecture des données (dataset) actuelles
2. Dans le data pipe, nous allons intégrer une valeur par défaut qui sera la valeur précédente en cas de saut de lecture
3. On renforce la robustesse du systeme en controlant le format des streams recus

9.1.4 Prise en compte des multiples datasets

Probleme Résoudre la question de la nécessité du merging des datasets

Solution : La question ne se pose même plus parce que les données étant des TimeSeries, elles sont indépendantes les unes des autres ; On entraînera des ensembles de modèles sur ces différents datasets

9.2 Ebauche des modeles

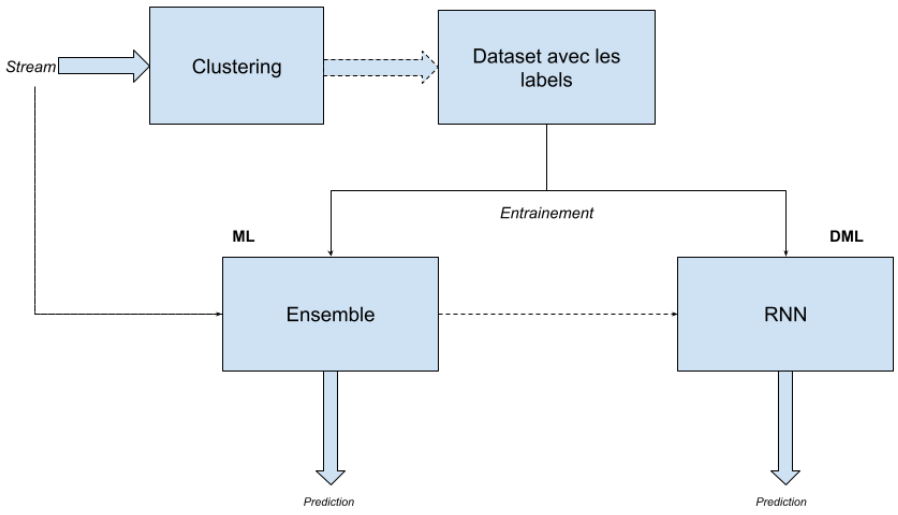


Figure 9 – Architecture des modeles

10 Conclusion

EcoENergie ayant a coeur la satisfaction de ses client s'est donné pour mission d'augmenter son portefeuille de service offert a ceux-ci quant a la gestion efficace de leur consommation énergetique.

Actuellement la retro-action n'étant pas effectuée sur la consommation d'énergie, le client d'Ecoénergie n'a aucun control en temps réel sur celle-ci. La solution proposée a EcoEnergie viendra donc palier a ce probleme : un outil de ML et DML pour automatiser cela. EcoEnergie met a notre disposition 03 datasets collectés par leur soins.

Le travail presenté dans ce rapport, présente la phase de planification du dit projet ; Cette proposition de projet sera présentée au Comité Exécutif le 25 Avril 2023. Prévoir ensuite 3 semaines pour modifications suite au feedback. Puis poursuivre avec la rédaction du plan de projet a proprement parlé, en même temps qu'on lance une investigation technique sur les ébauches de modeles et ainsi que l'utilisation des RNN pour les recommandations, et une discussion avec le Comité Ethique de Écoénergie sur le développement responsable de l'IA dans ce projet. En cas de validation, ce rapport marquera ainsi le debut de la phase d'implémentation si et seulement si la direction donne sont accord.

Références

[1] Mlflow documentation.

[2] Whylab documentation.

[3] whylogs documentation.