



Design Project

“Machine Learning pour l’
énergétique des bâtiments
sur outil SIG”

Maxwell Bergström &
Sélène Ledain

Lausanne - 4.06.2021

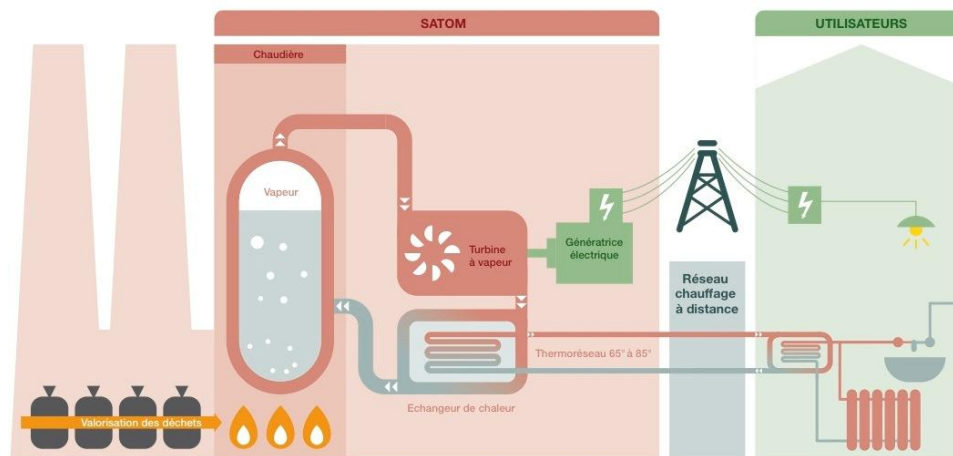
1.

Machine Learning pour l'énergétique des bâtiments

sur un outil SIG

Réseau de chaleur à distance centralisé qui fournit l'énergie à Collombey et Monthey

Principe de fonctionnement du thermoréseau



Source: SATOM SA

2.

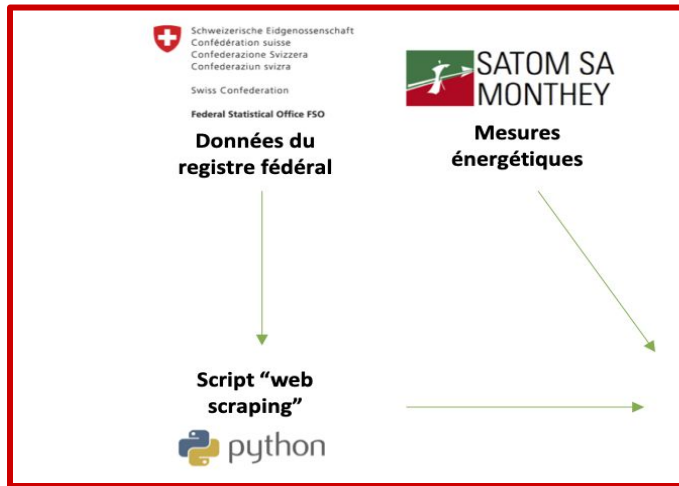
Machine Learning pour l'énergétique des bâtiments sur un outil SIG

Machine Learning pour l'énergétique des bâtiments

3.

sur un outil SIG

Workflow



Collection des données

Base de données



PostgreSQL

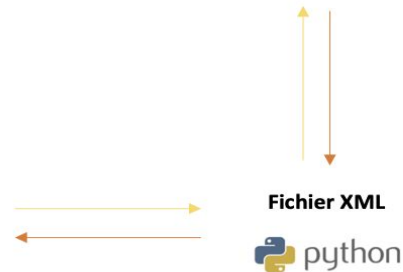
Modèles Machine Learning



Représentation des résultats



Simulation CitySim
Mesures SATOM



Données récoltées:

- Mesures énergétiques: situées dans la base de données de SATOM
- Registre fédéral: caractéristique du bâtiment
- Script "web scraping": collectionne les données depuis l'Office Fédéral Statistique (OFS)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Statistical Office FSO

**Données du
registre fédéral**



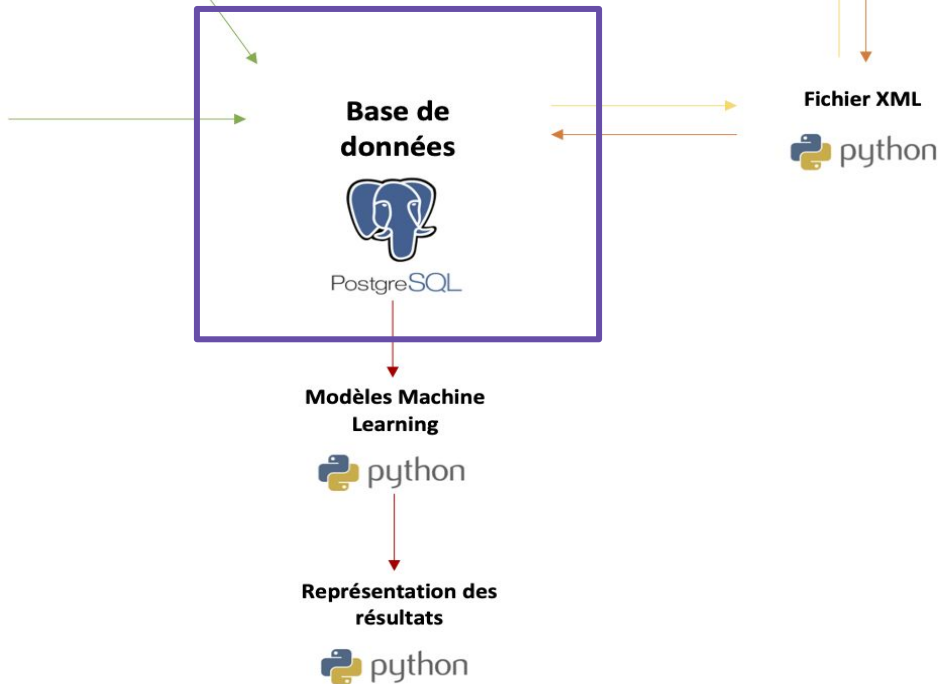
**Script "web
scraping"**



**Mesures
énergétiques**



Regroupement des données



Regroupement des données

- Données géométriques
- Données énergétiques
- Structure basée sur CityDB et CityGML
- Axe central du projet

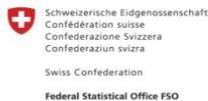


**Base de
données**



PostgreSQL





**Données du
registre fédéral**



**Script "web
scraping"**



**Mesures
énergétiques**



**Base de
données**

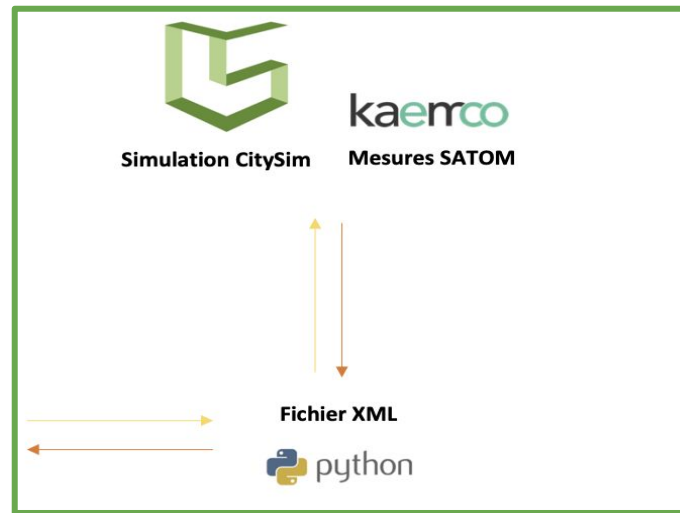


PostgreSQL

**Modèles Machine
Learning**

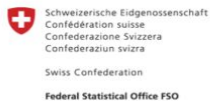


**Représentation des
résultats**



Préparations des données

**Développement des
modèles Machine Learning**



**Données du
registre fédéral**



**Script “web
scraping”**
python



**Mesures
énergétiques**



**Base de
données**



PostgreSQL

**Modèles Machine
Learning**



**Représentation des
résultats**



Simulation CitySim



Mesures SATOM



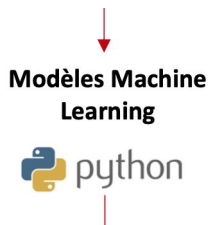
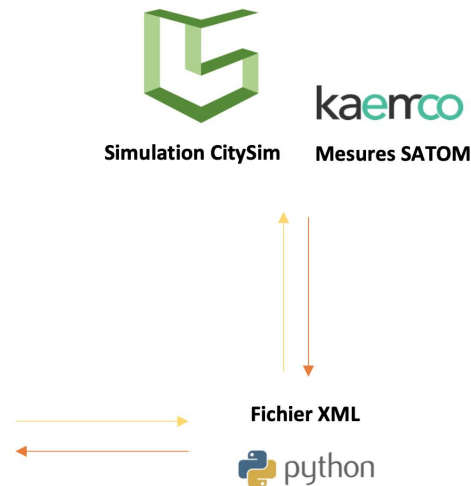
Fichier XML

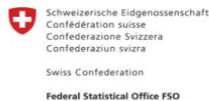


Représentation des données

Deux étapes:

- Préparation des données:
 - Raffinement de la granularité temporelle: simulation sur CitySim
 - Utilisation des mesures SATOM
- Développement des modèles de Machine Learning

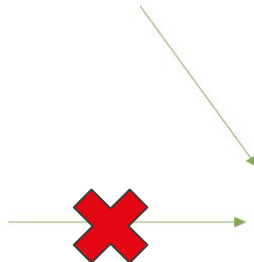




**Données du
registre fédéral**



**Mesures
énergétiques**



fusionnement des
propriétés physiques
des bâtiments et des
géométries existantes

**Base de
données**



**Modèles Machine
Learning**



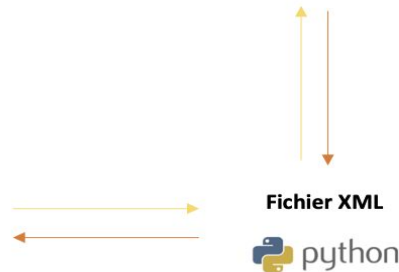
**Représentation des
résultats**



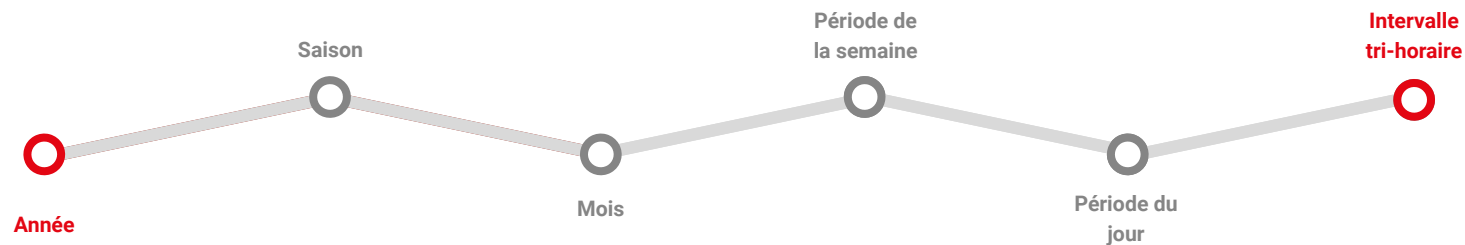
Simulation CitySim



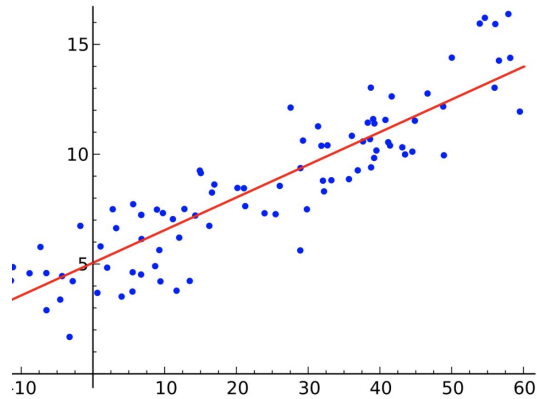
Mesures SATOM



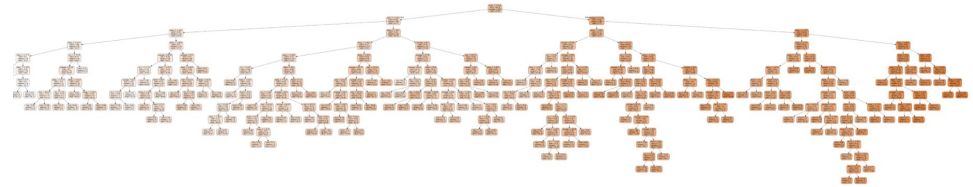
Comment prédire la demande énergétique?



Modèle linéaire (baseline)



Random Forest



Données du registre:

- La période de construction
- La fonction du bâtiment
- Le nombre d'étage (utile pour déterminer le volume)



“web scraping”



| EGID | GBAUP | GBAUJ | GKLAS | GKAT |
|-----------|--------|--------|--------|------|
| 926301 | 8016.0 | NaN | 1110.0 | 1020 |
| 191637939 | 8017.0 | 1990.0 | 1110.0 | 1020 |
| 9014600 | 8017.0 | NaN | 1122.0 | 1020 |
| 9014527 | NaN | NaN | 1251.0 | 1060 |
| 190019246 | 8020.0 | 2004.0 | 1110.0 | 1020 |
| ... | ... | ... | ... | ... |
| 190013443 | 8020.0 | 2003.0 | 1110.0 | 1020 |
| 924454 | 8014.0 | NaN | 1121.0 | 1020 |
| 3110391 | 8019.0 | NaN | 1110.0 | 1020 |
| 926016 | 8017.0 | NaN | 1110.0 | 1020 |
| 3110808 | 8019.0 | NaN | 1110.0 | 1020 |

Données géométriques:

- Volume

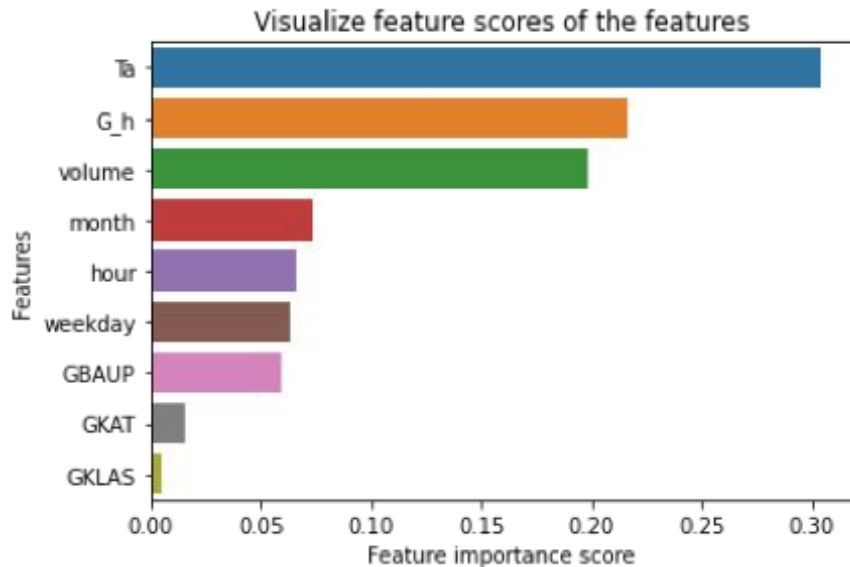
- Radiation globale horizontale
- Température de l'air
- Précipitation

Fichier climat
Aigle 2019



| | dm | m | h | G_h | Ta | Ts | RR |
|----|----|---|----|--------|------|------|-----|
| 0 | 1 | 1 | 1 | 2.00 | 0.84 | 4.40 | 0.0 |
| 1 | 1 | 1 | 2 | 2.00 | 0.72 | 4.33 | 0.0 |
| 2 | 1 | 1 | 3 | 2.00 | 0.45 | 4.30 | 0.0 |
| 3 | 1 | 1 | 4 | 1.83 | 0.13 | 4.30 | 0.0 |
| 4 | 1 | 1 | 5 | 2.00 | 0.45 | 4.23 | 0.0 |
| 5 | 1 | 1 | 6 | 2.00 | 1.27 | 4.20 | 0.0 |
| 6 | 1 | 1 | 7 | 2.00 | 1.57 | 4.18 | 0.0 |
| 7 | 1 | 1 | 8 | 6.17 | 0.32 | 4.10 | 0.0 |
| 8 | 1 | 1 | 9 | 94.33 | 0.78 | 4.10 | 0.0 |
| 9 | 1 | 1 | 10 | 211.83 | 2.50 | 4.02 | 0.0 |
| 10 | 1 | 1 | 11 | 332.67 | 3.98 | 4.00 | 0.0 |

Importance des features



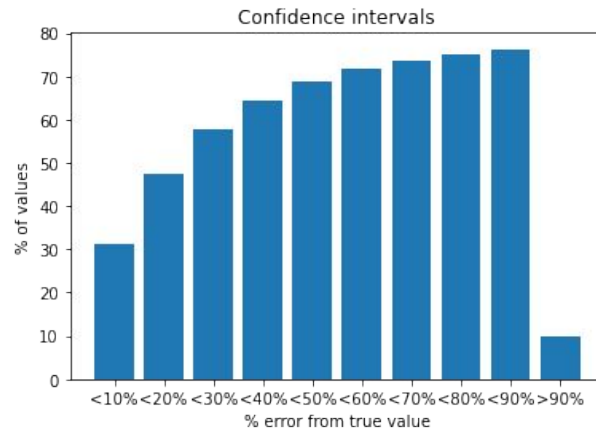
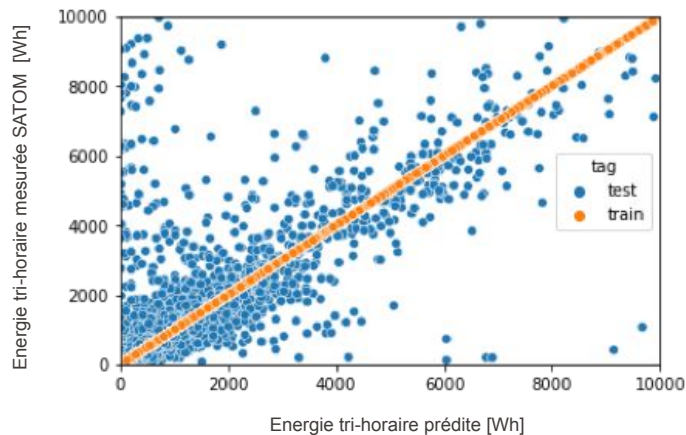
○ Intervalle
tri-horaire

Performance

| MAE [Wh/m ²] | MAPE [%] |
|--------------------------|------------------------|
| 36.7 | 2.65 * 10 ³ |

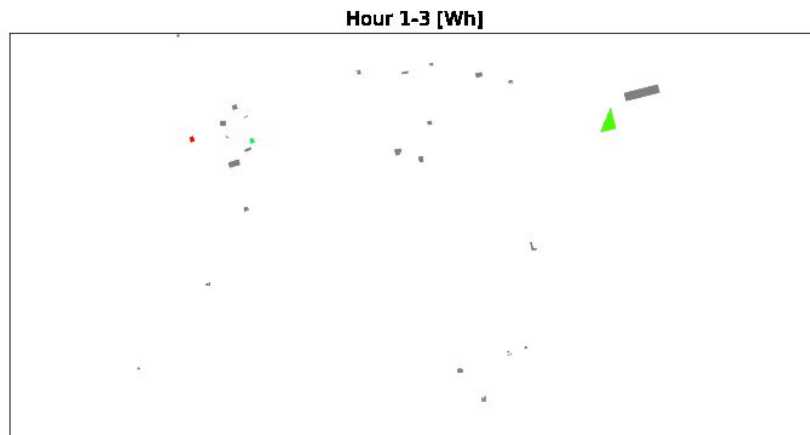


Intervalle tri-horaire



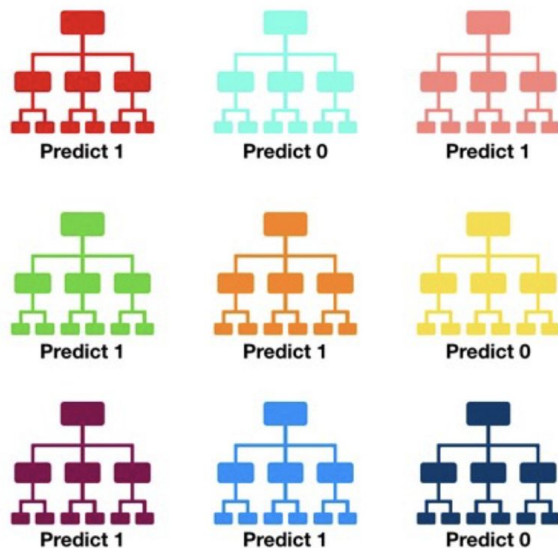


Intervalle tri-horaire



1er janvier 2019

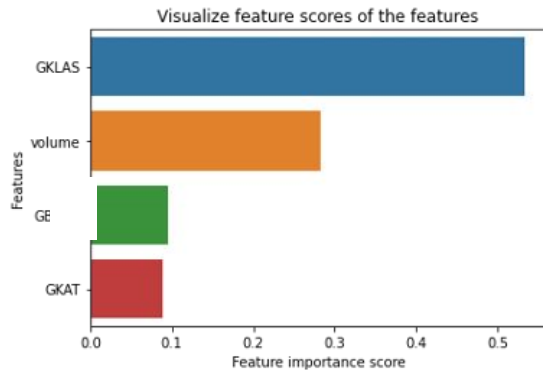
Choix de la meilleure combinaison de paramètres du Random Forest



Source: towardsdatascience.com

○ Année

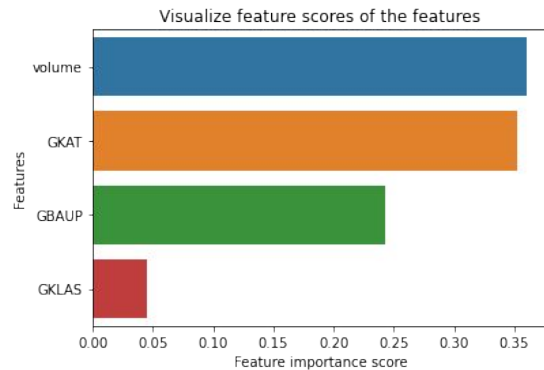
Par défaut



MAE [Wh/m²]

$4.86 \cdot 10^3$

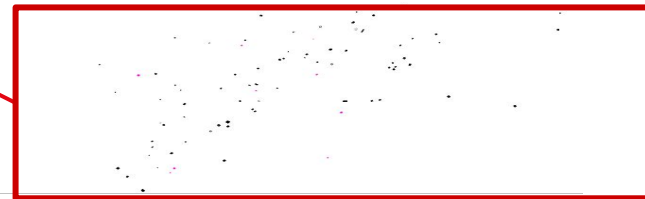
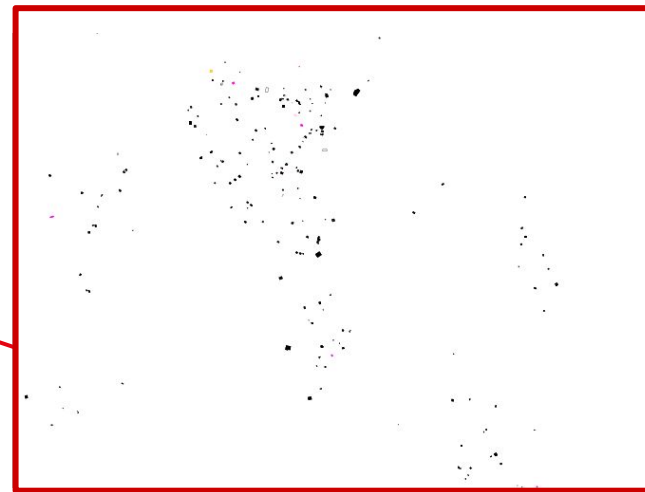
Hyperparamètres



MAE [Wh/m²]

$4.21 \cdot 10^3$

○ Année



Comparaison des modèles

| MAE tri horaire [Wh/m ² /an] | MAE annuel [Wh/m ² /an] |
|---|------------------------------------|
| $1.07 * 10^5$ | $4.21 * 10^3$ |

- MAE plus élevée dans le modèle tri-horaire
- Complexification d'un modèle n'implique pas nécessairement des meilleurs résultats
- Modèle tri-horaire reflète plus la réalité en terme de paramètres

Points forts

- Modèles adaptables aux features
- Granularité temporelle facilement modulable
- Adapté à la structure EnergyADE (CityDB)

Points faibles

- Normalisation biaisée par la surface
- Ambiguïté autour de 0
- Nombre réduit d'entrées dans les modèles

Recommendations & améliorations

- Harmonisation de la base de données
- Nettoyage des données
- Feature engineering
- Analyse spatiale plus approfondie

Nous souhaitons remercier

Prof. François Golay (EPFL)

Jérôme Kämpf (Idiap)

Giuseppe Peronato (Idiap)

Merci !