

Prédire la performance énergétique des bâtiments

Apprendre les données des DPE

Marc Grossouvre¹ Jonathan Villot² Didier Rulière³

¹Ingénieur en sciences des données, Doctorant CIFRE, U.R.B.S. SAS,
marcgrossouvre@urbs.fr

²Co-fondateur - U.R.B.S. SAS

³Mines Saint-Etienne - LIMOS - Univ Clermont Auvergne

5 janvier 2023, LOCIE

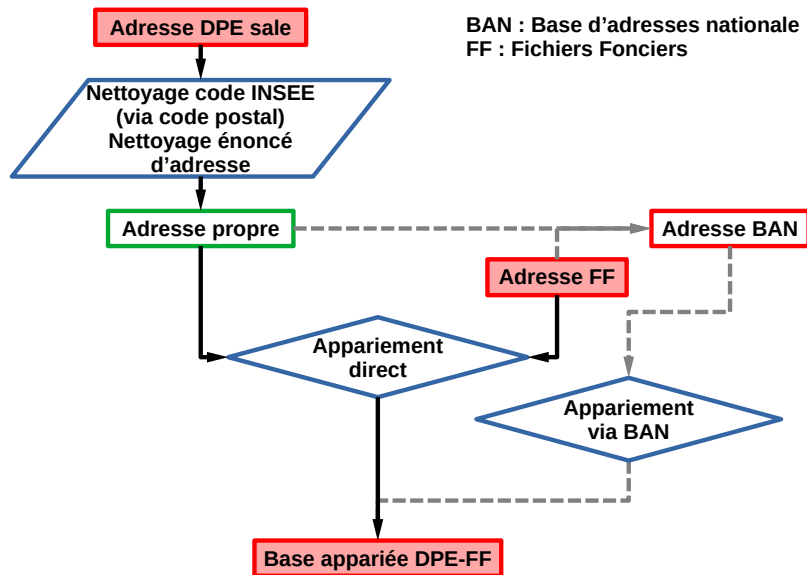


- 1 La base des DPE : beaucoup d'informations très incomplètes
- 2 Apparier les DPE à des bases complètes par les adresses
- 3 Quel cadre pour un apprentissage supervisé ?
- 4 Les plus proches "voisins"
- 5 Les résultats
- 6 Vers une prise en compte des données granulaires

Code de la construction et de l'habitat articles L134-1 à L134-5

*"Le diagnostic de performance énergétique d'un bâtiment ou d'une partie de bâtiment est **un document** qui comprend la quantité d'énergie effectivement consommée ou estimée, exprimée en énergie primaire et finale, pour une utilisation standardisée du bâtiment ou de la partie de bâtiment et **une classification en fonction de valeurs de référence afin que les consommateurs puissent comparer et évaluer sa performance énergétique.**"*

Apparier les DPE à des bases complètes par les adresses

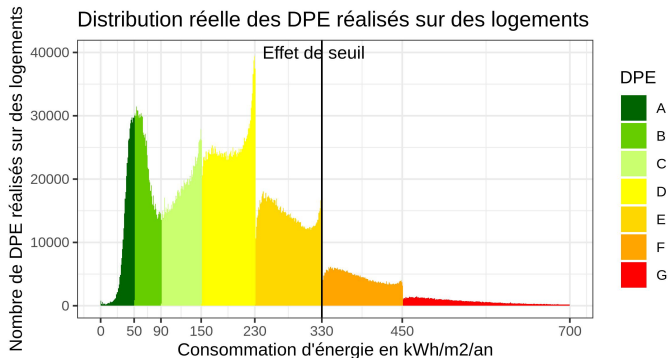


Quel cadre pour un apprentissage supervisé ?

Quoi prédire ?

Le dernier DPE réalisé sur le bâtiment (assimilé à une adresse), pris comme une valeur légale.

Classes d'effectifs hétérogènes, effet de seuil.



Balanced accuracy

$$\frac{1}{7} \sum_{x \in \{A,B,C,D,E,F,G\}} \frac{\text{nombre de bonnes prédictions } x}{\text{nombre de valeurs } x \text{ observées}}$$

C'est un nombre entre 0 et 1 que l'on cherche à maximiser.

Point fort Donne le même poids à toutes les étiquettes.

Point faible Ne tient pas compte de la "gravité" des erreurs.

Quel apprentissage ? Renoncer à un modèle physique

Les caractéristiques physiques détaillées ne sont pas disponibles partout et ne peuvent pas l'être sans inspection physique.



On s'appuie sur le lien statistique entre le DPE et des **variables connexes**.

Les petits appartements ont des mauvais DPE.
Les personnes très pauvres vivent plus souvent dans des passoires
énergétiques, sauf pour les logements sociaux.

...

Les plus proches "voisins" (1/2)

On s'appuie sur les bâtiments les plus ressemblants physiquement, géographiquement et socialement.

- Maisons en Normandie.
- Chauffées à l'électricité.
- Construites au XIXe siècle.
- Dans des carreaux de revenu moyen similaire.
- En monopropriété.
- Ne sont pas en zone prioritaire.



28 Rue Fernand Feron
76400 Ganzeville

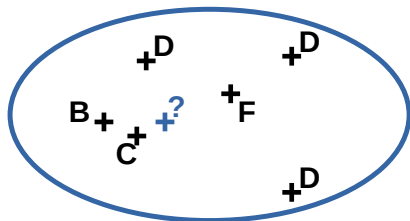


17 pl. de la République
27740 Poses

Les plus proches "voisins" (2/2)

Prédiction par les k plus proches voisins (KNN, k-nearest neighbours)

On prédit l'étiquette majoritaire parmi les k plus proches voisins. Et la consommation moyenne des plus proches voisins ayant l'étiquette prédite.



Etiquette prédite : D

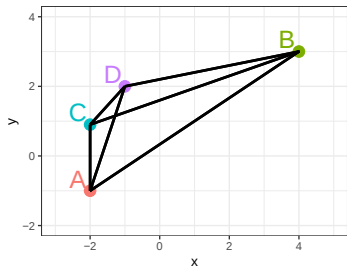
Consommation prédite : moyenne des consommations des points D.

Paramètres

Sélection des variables, optimisation du nombre de voisins, méthode de calcul de la distance. Optimisation par région.

Les résultats (1/2) : Résultats qualitatifs

- **Les variables les plus sélectionnées** : Matériaux (murs et/ou toiture), énergie de chauffage, période de réglementation thermique, type de propriétaire, type d'occupant.
- **Les distances testées** : Euclide, Manhattan, **Hassanat**.
- **Les transformations de variables testées** : variables brutes, ré-échelonnées entre 0 et 1, centrées normées, **rangs**, rangs normalisés.



d EuclideManhattanHassanat

CD	1.49	2.1	0.867
AC	1.90	1.9	0.655
AD	3.16	4.0	1.250
BD	5.10	6.0	1.083
BC	6.36	8.1	1.382
AB	7.21	10.0	1.657

Les résultats (2/2) : La matrice de confusion

Vraies valeurs	Valeurs prédites						
	A	B	C	D	E	F	G
A	88	40	32	84	42	12	1
B	38	109	114	161	60	15	1
C	25	44	503	863	222	48	1
D	24	35	377	1930	869	201	7
E	9	20	111	1084	1113	310	14
F	2	8	27	318	497	214	12
G	1	3	5	75	144	76	8

Prédictions correctes

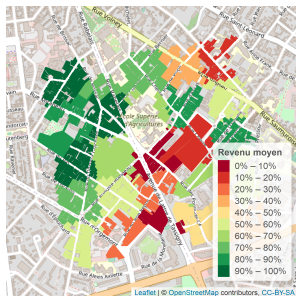
Part de toutes les
prédictions en %

+/- 0 étiquettes	39.7
+/- 1 étiquettes	82.9
+/- 2 étiquettes	95.5

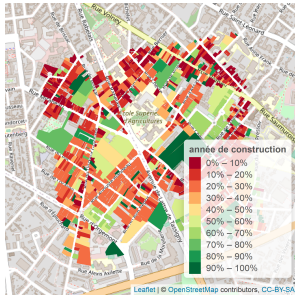
Balanced accuracy

29%

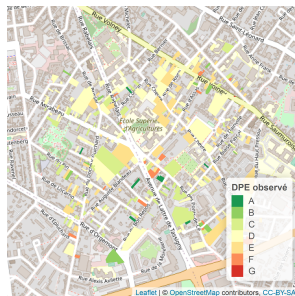
Vers une prise en compte des données granulaires



Revenu moyen au
carreau INSEE



Année de construction à
la partie d'évaluation



DPE à l'adresse
ADEME

Il est faux de dire que le revenu moyen des ménages du carreau est le revenu moyen des ménages de chaque bâtiment du carreau. Mais il est possible que le revenu moyen soit le revenu d'un des ménages du carreau. On introduit la notion de **distribution de mélange**.

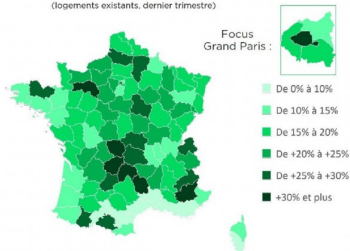
Conclusion (1/2)

- Les données socio-économiques contribuent à prédire le DPE.
- Oui mais il faut apparier des données à différentes granularités.
- Les DPE sont aussi une information géolocalisée.

Donc le problème appartient à la **géostatistique**.

Variables d'entrée	Balanced accuracy
année de construction	19%
année de construction, latitude, longitude	25%

Répartition géographique des DPE F ou G par département :
(logements existants, dernier trimestre)

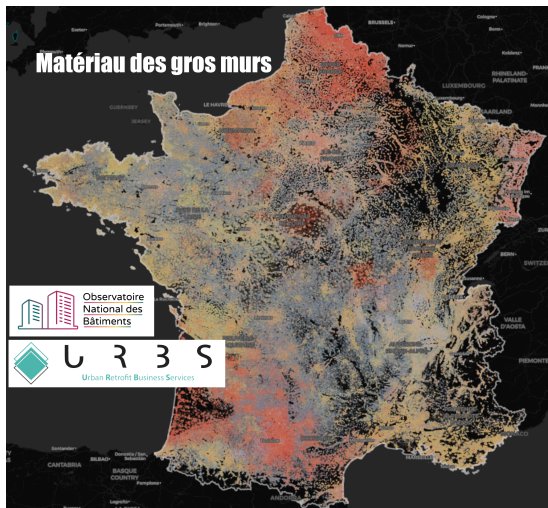


Le baromètre DPE de la Fnaim montre de fortes disparités de classe énergétique selon la situation géographique.

(www.batiactu.com, 09/12/2022)

Conclusion (2/2) : Merci!

<https://www.imope.fr/onb.html>



Merci de votre attention