

Normalisation des données & Score énergétique

Les données issues de la base de données sont toutes différentes : certaines sont exprimées en mètres carrés (surface), d'autres sont des nombres entiers (nombre de fenêtres) ou des booléens (présence de chauffage). Afin de pouvoir calculer un score énergétique cohérent et comparable entre toutes les salles, il est nécessaire de ramener l'ensemble de ces valeurs sur une même échelle.

Pour cela, on utilise une méthode de normalisation Min–Max. Cette méthode consiste à transformer chaque valeur réelle en une valeur comprise entre 0 et 1.

La surface de la salle et le nombre de fenêtres sont normalisés à l'aide de cette formule. Plus la surface est grande, plus la valeur normalisée est élevée, ce qui correspond à une consommation énergétique potentiellement plus importante. À l'inverse, un grand nombre de fenêtres améliore l'éclairage naturel et réduit les besoins en éclairage artificiel.

Les variables booléennes sont converties de la manière suivante :

- FALSE est converti en 0,
- TRUE est converti en 1.

Ainsi, la présence de chauffage ou de climatisation est directement interprétée comme un facteur de surconsommation.

Enfin, l'orientation de la salle est traduite en coefficient numérique afin de représenter l'impact de l'ensoleillement naturel :

- SUD = 1.0
- EST / OUEST = 0.6
- NORD = 0.2

Cette transformation permet de prendre en compte l'influence de l'exposition solaire sur les besoins en chauffage et en éclairage.

Grâce à cette normalisation, l'ensemble des données énergétiques est désormais exprimé sur une même échelle comprise entre 0 et 1, ce qui permet de calculer un score énergétique fiable et comparable pour chaque salle.

Score énergétique des salles

Le score énergétique est une valeur numérique comprise entre 0 et 100 qui permet d'évaluer la performance énergétique d'une salle. Il prend en compte plusieurs caractéristiques physiques et techniques de la salle, telles que la surface, le nombre de fenêtres, l'orientation, la présence de chauffage ou de climatisation.

Principe du calcul

Le score énergétique est basé sur la normalisation des données, afin de ramener toutes les variables sur une même échelle (0–1) avant de les combiner. Les étapes sont les suivantes :

1. Normalisation des variables continues

- La surface (surface_m2) et le nombre de fenêtres (nb_fenêtres) sont normalisés avec la formule Min–Max :

$$valeur_normalisée = \frac{valeur - valeur_{min}}{valeur_{max} - valeur_{min}}$$

Cette transformation permet de comparer directement des grandeurs de nature différente.

2. Conversion des variables booléennes

- Les indicateurs climatisation et chauffage sont convertis en 1 si présents, 0 sinon.

3. Conversion de l'orientation en coefficient

- L'orientation de la salle influence la lumière naturelle et donc les besoins énergétiques :
 - SUD = 1.0
 - EST / OUEST = 0.6
 - NORD = 0.2

4. Combinaison des facteurs avec coefficients en fonction

Chaque variable normalisée est multipliée par un coefficient reflétant son impact sur la consommation énergétique :

Le score énergétique d'une salle est une note sur 100 qui permet d'évaluer sa consommation d'énergie. Il combine plusieurs critères normalisés entre 0 et 1 pour pouvoir les comparer sur la même échelle.

La formule utilisée est :

$$\text{score} = 100 \times (0.4 \times (1 - \text{surfaceNorm}) + 0.3 \times \text{fenNorm} + 0.2 \times \text{orient} + 0.1 \times (1 - \text{chauffageVal}))$$

- Surface normalisée (surfaceNorm) : plus la salle est grande, plus la consommation est élevée. On prend (1 - surfaceNorm) pour que les grandes salles aient un score plus faible.
- Fenêtres normalisées (fenNorm) : plus la salle a de fenêtres, mieux c'est pour l'éclairage naturel, donc contribution positive au score.
- Orientation (orient) : coefficient basé sur l'exposition au soleil (SUD = 1.0, EST/OUEST = 0.6, NORD = 0.2). Une bonne orientation augmente le score.
- Chauffage (chauffageVal) : si la salle a un chauffage, la consommation augmente, donc (1 - chauffageVal) réduit le score.

Chaque critère est multiplié par un coefficient qui représente son importance relative :

- 40 % pour la surface, 30 % pour les fenêtres, 20 % pour l'orientation, 10 % pour le chauffage.

Enfin, on multiplie par 100 pour obtenir un score sur 100 : plus le score est élevé, plus la salle est économe en énergie, et plus il est faible, plus elle est énergivore.

Interprétation

Score	Interprétation
75 – 100	Salle très économe en énergie
50 – 75	Consommation modérée
25 – 50	Salle énergivore
0 – 25	Salle très énergivore

Source :

<https://datascientist.fr/blog/comment-la-normalisation-des-donnees-ameliore-les-performances-des-modeles-d-apprentissage-automatique>