

# Feature Engineering

Créer, transformer et sélectionner des variables

K. Kadri

- 1 Pourquoi le Feature Engineering ?
- 2 Types de transformations

- Transformer les données pour aider le modèle à apprendre.
- Améliorer performances, stabilité, généralisation.
- Réduire le bruit et révéler la structure.

## Idée clé

Un bon feature engineering peut valoir plus que changer d'algorithme.

# 1. Encodage des variables catégorielles

- One-Hot Encoding
- Ordinal Encoding
- Target Encoding
- Binary / Hashing encoding (datasets volumineux)

## Astuce

Pour les arbres (RF, XGBoost), One-hot est souvent inutile → Ordinal OK.

## 2. Normalisation / Standardisation

- MinMaxScaler :  $[0, 1]$
- StandardScaler : moyenne 0, variance 1
- RobustScaler : robuste aux outliers

À utiliser absolument avec

SVM, KNN, Logistic Regression, K-Means, PCA.

### 3. Feature Construction

- Polynomials features
- Ratios / interactions
- Domain knowledge features

#### Exemple

Prix / Surface  $\rightarrow$  densité de prix Très utile pour l'immobilier.

## 4. Feature Selection

- Variance Threshold
- Sélection univariée (ANOVA, Chi2)
- RFE (Recursive Feature Elimination)
- SelectFromModel (L1, arbres)

### Avantage

Moins de bruit → modèle plus rapide et plus général.