

# Interprétabilité des modèles

## Méthode SHAP (SHapley Additive exPlanations)

K. Kadri

- 1 Pourquoi expliquer un modèle ?
- 2 Shapley Values
- 3 Visualisations SHAP
- 4 Quand utiliser SHAP ?
- 5 Référence

# Besoin d'interprétabilité

- Confiance dans les prédictions.
- Détection des biais.
- Validation réglementaire.
- Communication business.

## Idée clé

Un modèle opaque = un modèle inutilisable dans l'industrie.

# Origine des SHAP

- Inspiré de la théorie des jeux coopératifs.
- Mesure la contribution moyenne d'une feature.

## Formule (intention)

Contribution(feature) = gain moyen quand la feature “rejoint le jeu”.

## Propriétés

Équitables, cohérentes, additives.

# SHAP Summary Plot

- Importance globale des variables.
- Sens + direction de l'effet.



## Interprétation rapide

- Chaque point = contribution d'une feature pour un individu.
- La couleur représente la valeur de la feature (faible → bleu, élevé → rouge).
- L'axe horizontal = impact sur la prédiction (positif ou négatif).

- Explication locale pour un individu.
- Rouge = pousse vers la classe 1 Bleu = pousse vers la classe 0.

## Usage

Communication business → expliquer une décision individuelle.

# Cas d'usage recommandés

- Arbres de décision (RF, XGBoost, LightGBM).
- Modèles tabulaires.
- Diagnostics médicaux.
- Finance / scoring crédit.

## Attention

SHAP est coûteux pour les modèles non arborescents.

- Documentation complète SHAP : <https://shap.readthedocs.io/en/latest/>
- Exemples de visualisation : summary plot, force plot, dependence plot
- Implementations spécialisées :
  - TreeExplainer (RF, XGBoost, LightGBM)
  - LinearExplainer (modèles linéaires)
  - KernelExplainer (modèles black-box)

## Conseil

Toujours vérifier la compatibilité du bon explainer avec votre modèle. TreeExplainer est le plus efficace et le plus exact.