

The background features a white canvas with several small, solid-colored circles in red, orange, blue, and purple scattered across it. Overlaid on these are several smooth, wavy lines in the same color palette, creating a dynamic, abstract pattern that frames the central text.

Soutenance en Simulation Monte Carlo

Charles Meldhine MADI MNEMOI – Youssef SAIDI – Lucas TERRA – Marwa TOURABI

ING1 GMA

10/05/2023

Sommaire

- 1) Choix des technologies
- 2) Présentation du jeu de données
- 3) Modélisation
 - a) Analyse des premiers résultats
 - b) Tests de différentes fonctions de covariance
 - c) Optimisation des hyperparamètres
 - d) Comparaison des modèles sur le jeu de test
- 4) Interprétation des résultats

Choix des technologies



Manipulation de tableaux multidimensionnels
et calculs mathématiques



Manipulation et analyse des
données en tableaux



python



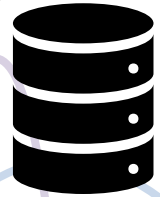
seaborn

Visualisation statistique de données



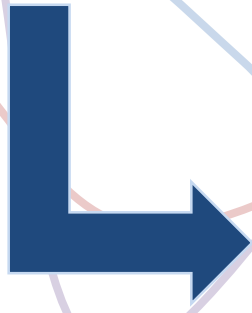
Visualisation 2D et 3D des données

Exploration des données



Fichier texte contenant :
2 colonnes et 205 observations

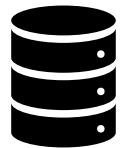
Variable explicative → Âge
Variable cible → Logarithme des revenus



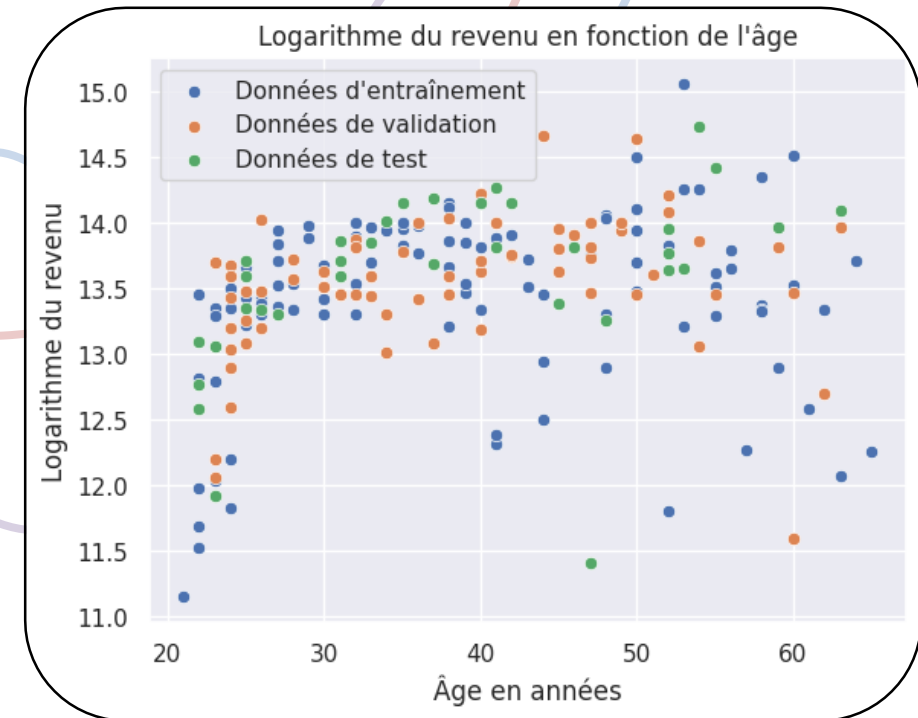
Jeu d'entraînement → 50%



Jeu de validation → 30%



Jeu de test → 20%

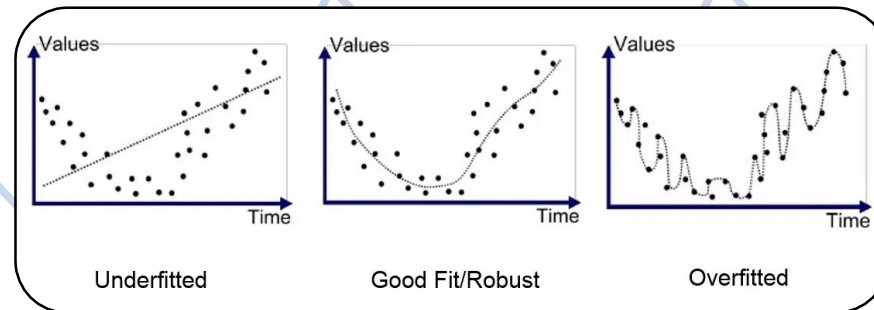


Modélisation

- Utilisation de l'EQM pour vérifier l'efficacité des modèles

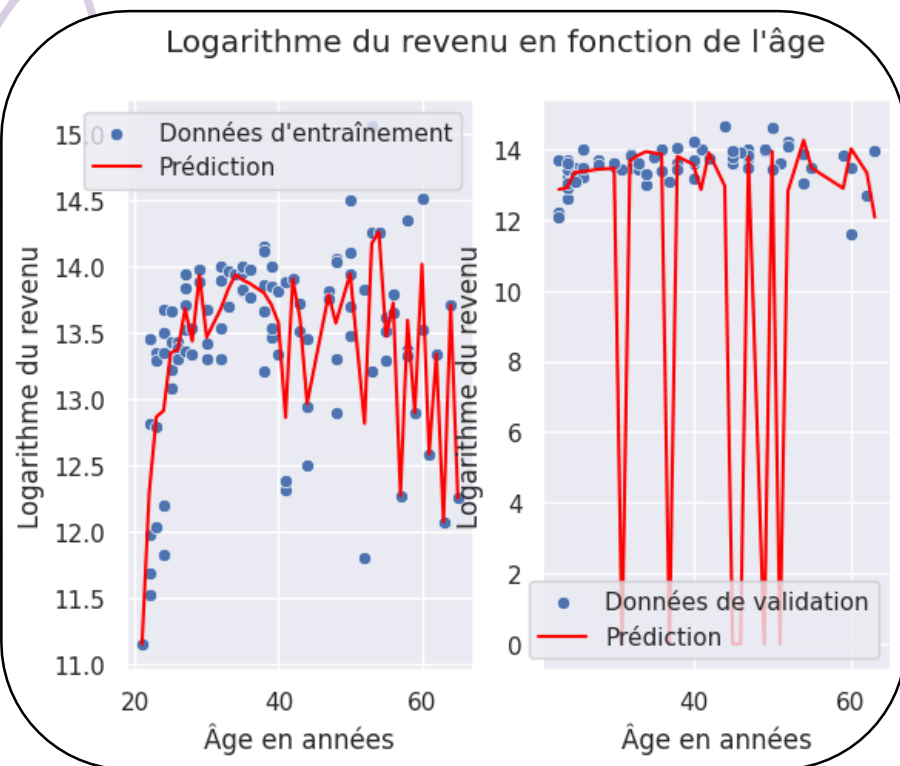
$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- Fixation des hyperparamètres :
 - θ à 0,1
 - σ_{noise} à 0,00001
- Vérification du sous ou sur-apprentissage pour chaque modèle



Analyse des premiers résultats

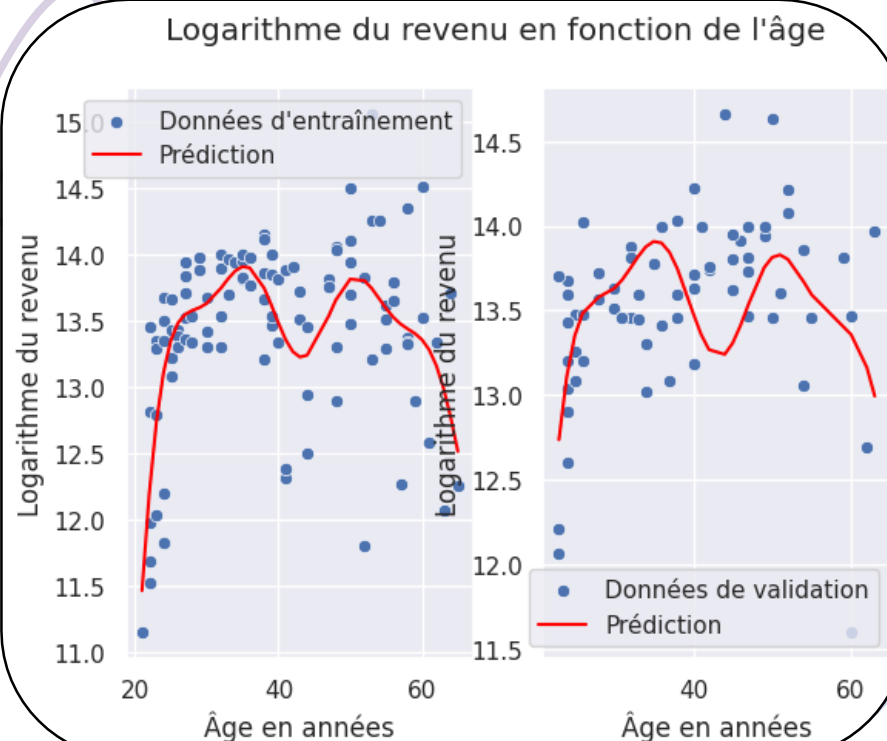
- Fonction matern :



- Beaucoup de fluctuations
 - Courbe de prediction “colle” les données d’entraînement
 - Mauvaise prediction sur les données de validation
 - Sur-apprentissage
- EQM jeu d’entraînement = 0,16
EQM jeu de validation = 27,95

Analyse des premiers résultats

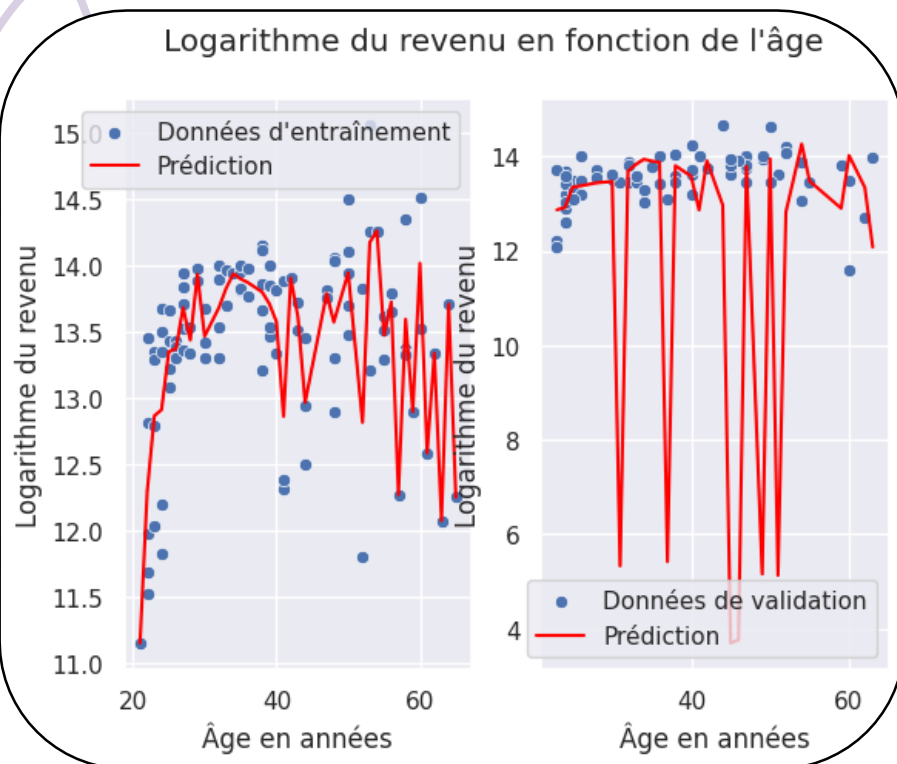
- Fonction gaussienne : $k(x, x') = \exp\left(-\frac{(x - x')^2}{2\theta^2}\right)$



- Peu voire pas de bruit
- Courbe distinguable
- Courbe passe au niveau des masses de points
- Bon modèle
- EQM jeu d'entraînement = 0,2709
EQM jeu de validation = 0,2480

Analyse des premiers résultats

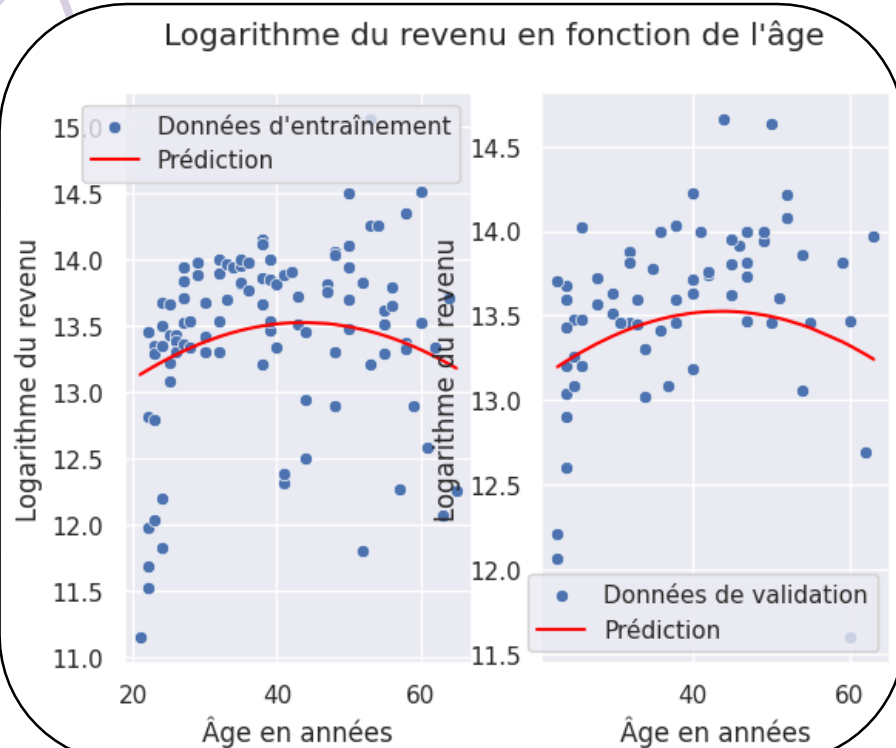
- Fonction rationnelle quadratique : $k(x, x') = 1 + \frac{(x - x')^2}{2\theta^2}$



- Beaucoup de fluctuations
- Courbe de prediction “colle” les données d’entraînement
- Mauvaise prediction sur les données de validation
- Sur-apprentissage
- EQM jeu d’entraînement = 0,16
EQM jeu de validation = 12,57

Optimisation des hyperparamètres

Optimisation de la fonction matern



- Avant optimisation :

- $\theta = 0,1$
- $\sigma_{noise} = 0,00001$

- Après optimization :

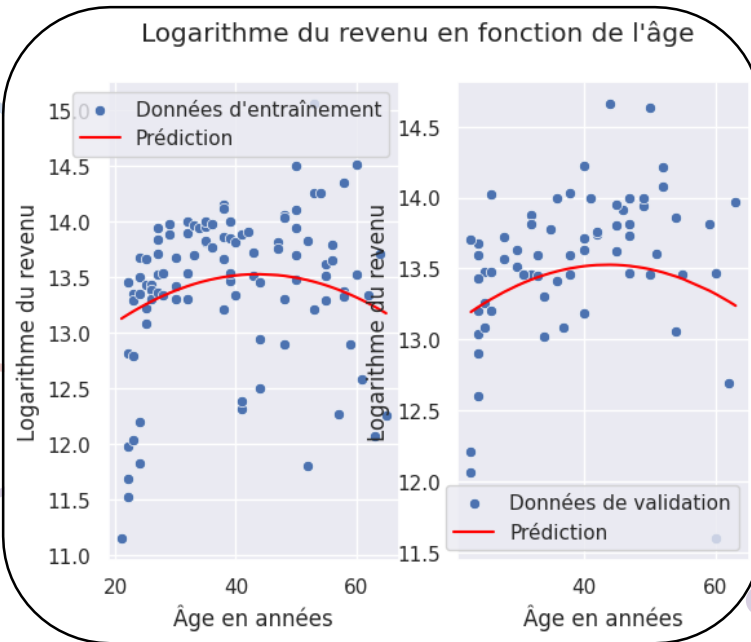
- $\theta = 119,7693$
- $\sigma_{noise} = 0,6972$

EQM jeu d'entraînement = 0,4100

EQM jeu de validation = 0,2355

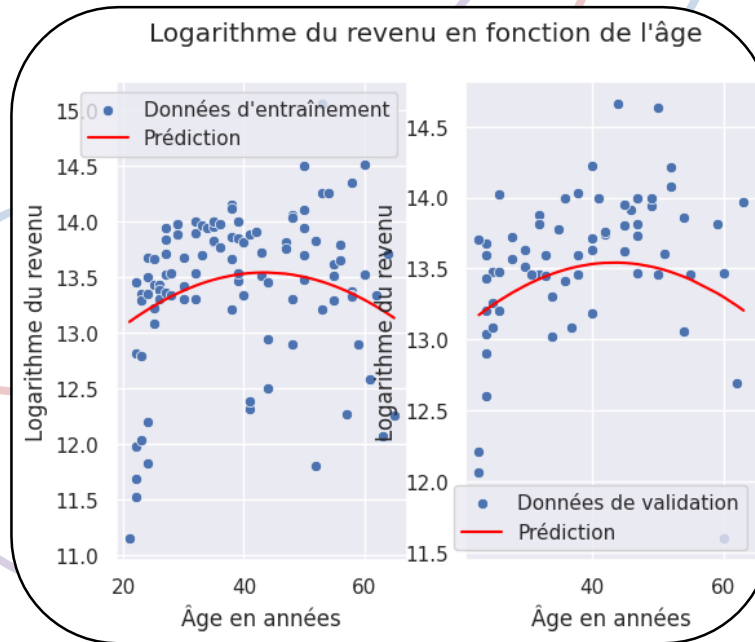
Optimisation des hyperparamètres

Optimisation de la
fonction gaussienne



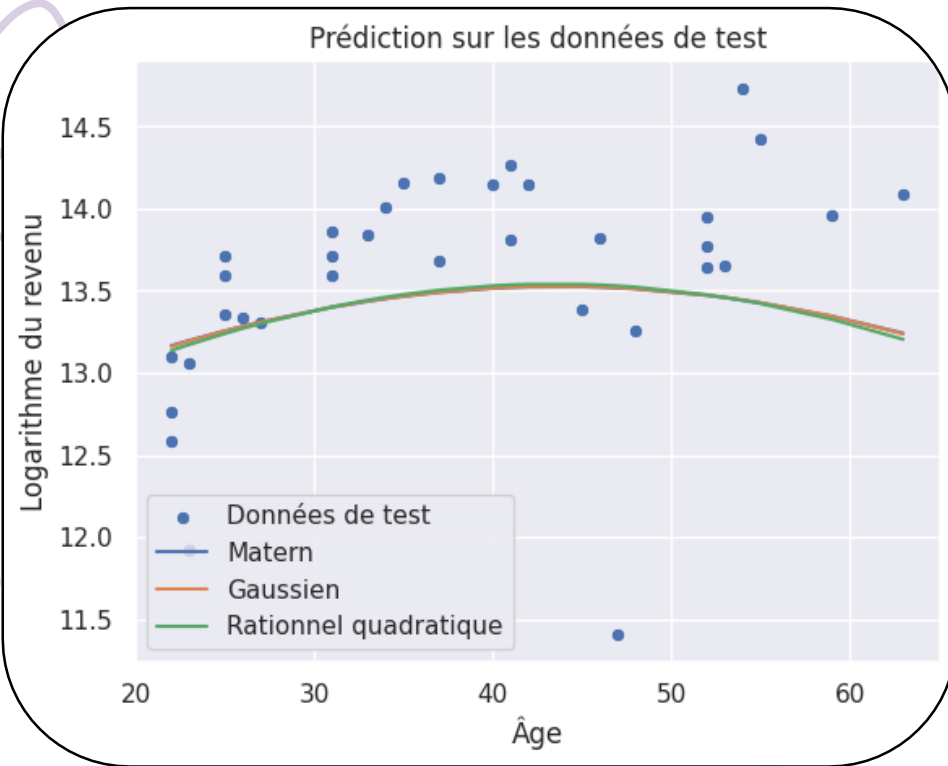
EQM E = 0,4089
EQM V = 0,2351

Optimisation de la fonction
rationnelle quadratique



EQM E = 0,4037
EQM V = 0,2310

Comparaison des modèles sur le jeu de test

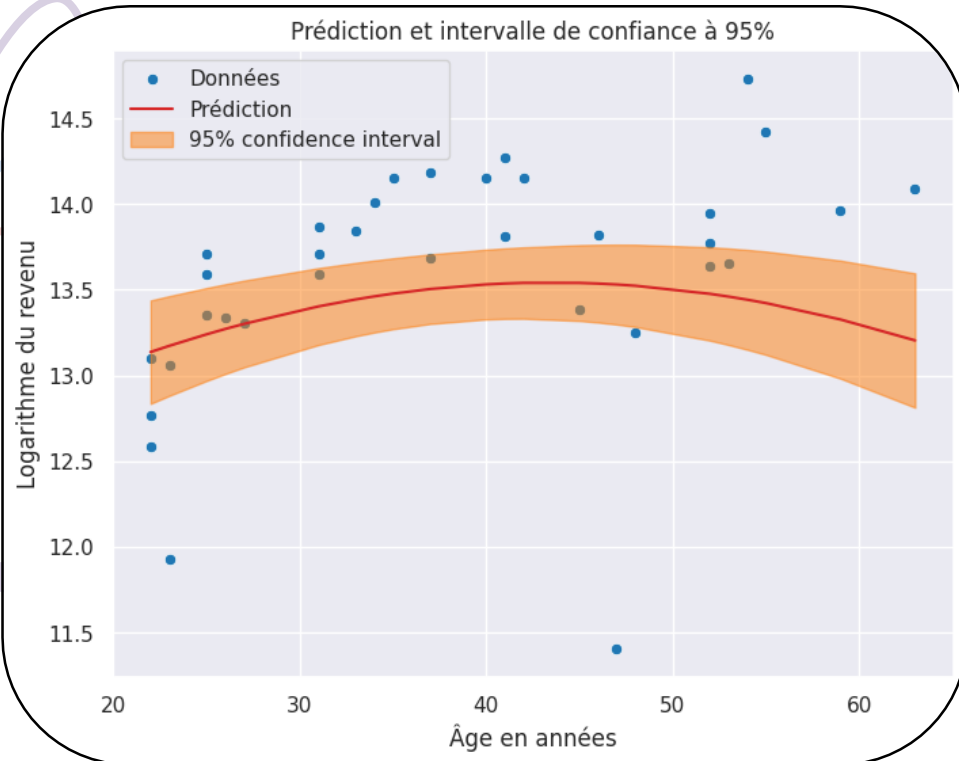


- Superposition des trois modèles
- Assez bonne prediction des données

Fonctions	EQM
Matern	0,4267
Gaussienne	0,4261
Rationnelle quadratique	0,4244

- Modèle le plus performant : RQ function

Interprétation des résultats



← Intervalle de confiance à 95% du modèle

- $RMSE \approx 0,62$
- Prédiction du logarithme du revenu à $\pm 0,62$

The background features a white canvas with several small, solid-colored circles in red, orange, blue, and purple scattered across it. Overlaid on these are three continuous, wavy lines in purple, blue, and red, which meander across the frame, creating a dynamic and artistic feel.

Merci de votre attention !

Charles Meldhine MADI MNEMOI – Youssef SAIDI – Lucas TERRA – Marwa TOURABI

ING1 GMA

10/05/2023