



# PRÉDICTION DES PRIX DES MAISONS

*Rapport d'Analyse Machine Learning*

- Algorithmes: Linear Regression, Random Forest
    - Objectif: Prédiction des prix immobiliers
    - Métriques: RMSE, MAE, R<sup>2</sup>, MSE
  - Techniques: Preprocessing, Optimisation des hyperparamètres
- Date de génération: 31/12/2025 16:13

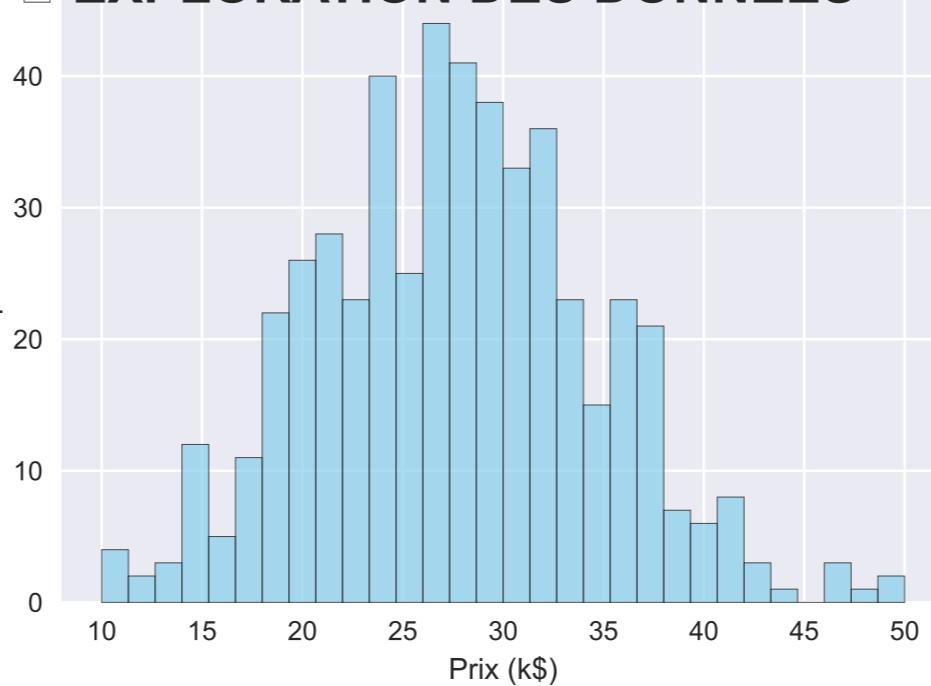


ML PROJECT

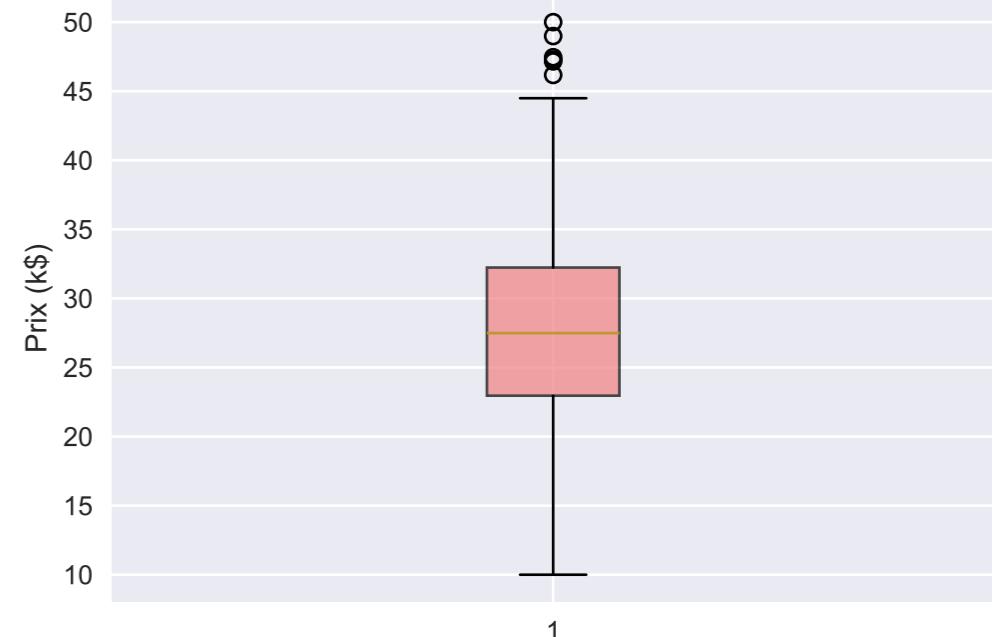
## Statistiques Générales

Échantillons: 506  
Features: 13  
Prix moyen: 27.67k\$  
Prix médian: 27.49k\$  
Écart-type: 6.98k\$

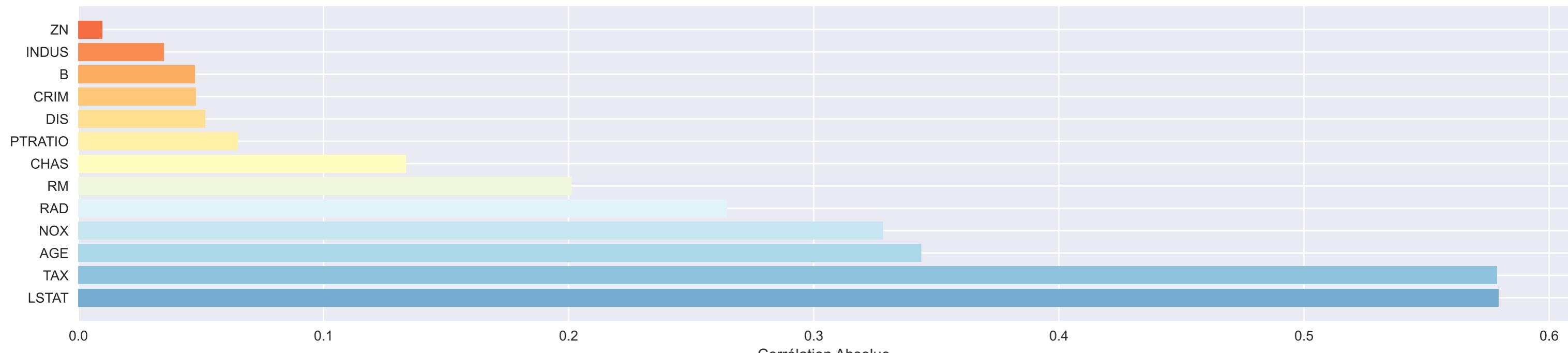
## DISTRIBUTION DES PRIX



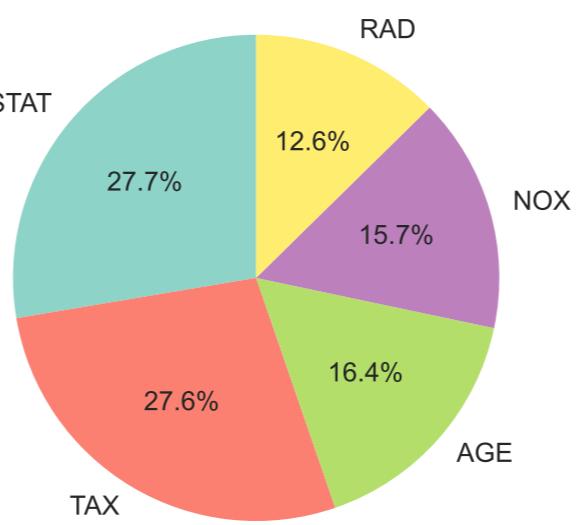
## Boxplot des Prix



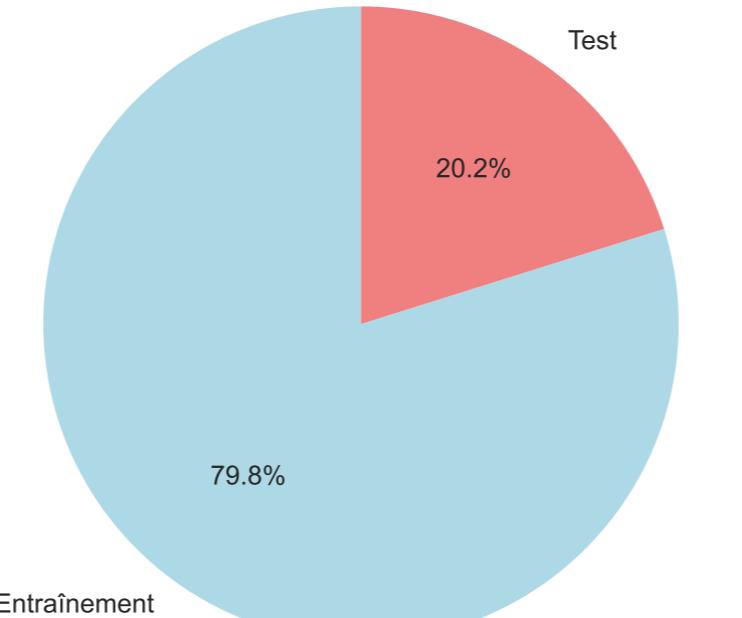
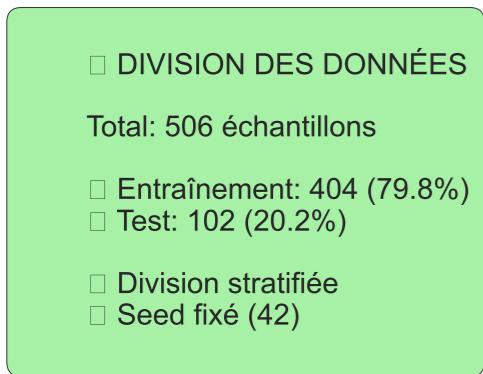
## Corrélation avec le Prix



## Top 5 Features Importantes



# PREPROCESSING DES DONNÉES



## Étapes du Preprocessing

1.  Chargement des données

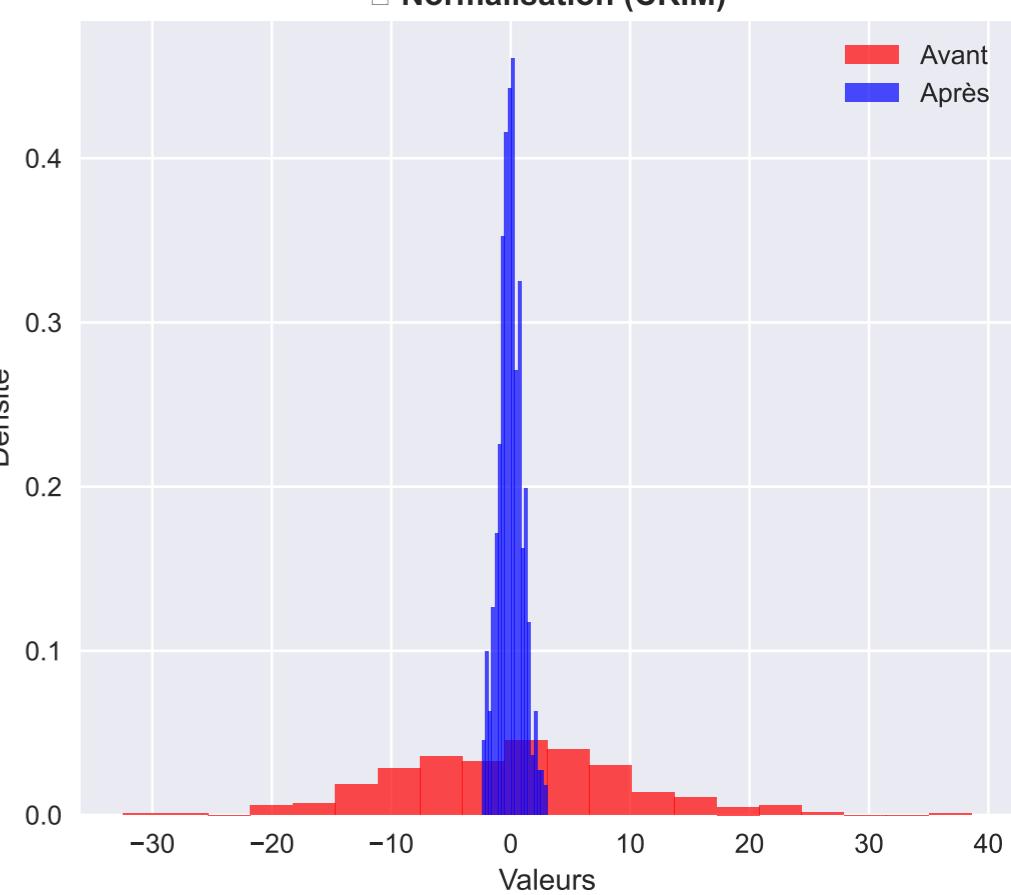
2.  Vérification des valeurs manquantes

3.   Encodage des variables catégorielles

4.   Division train/test (80/20)

5.  Normalisation StandardScaler

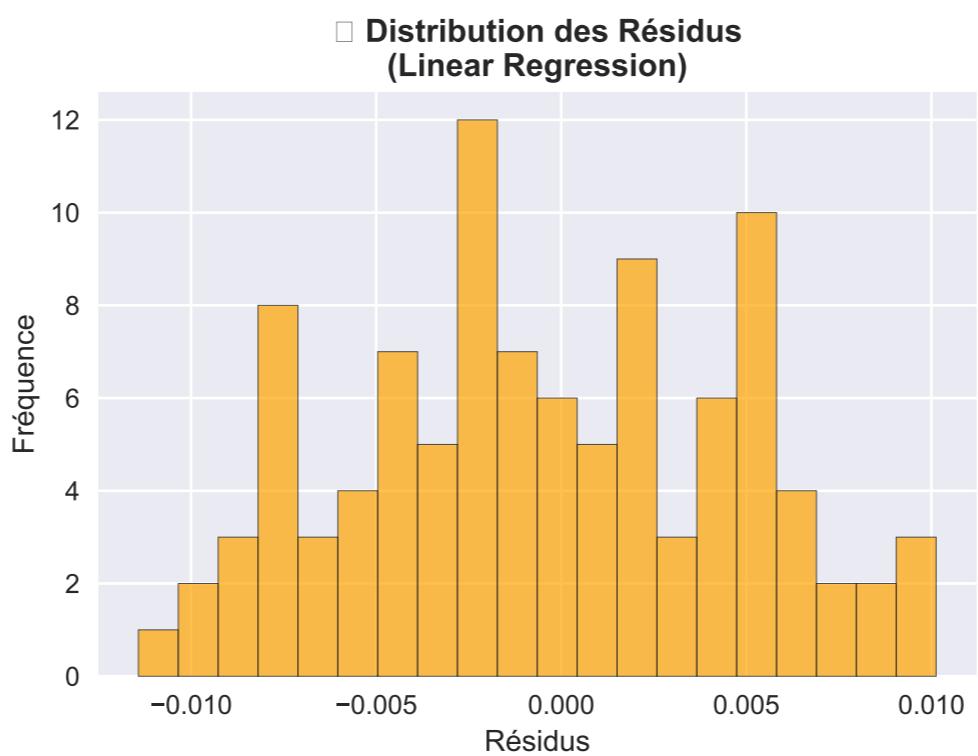
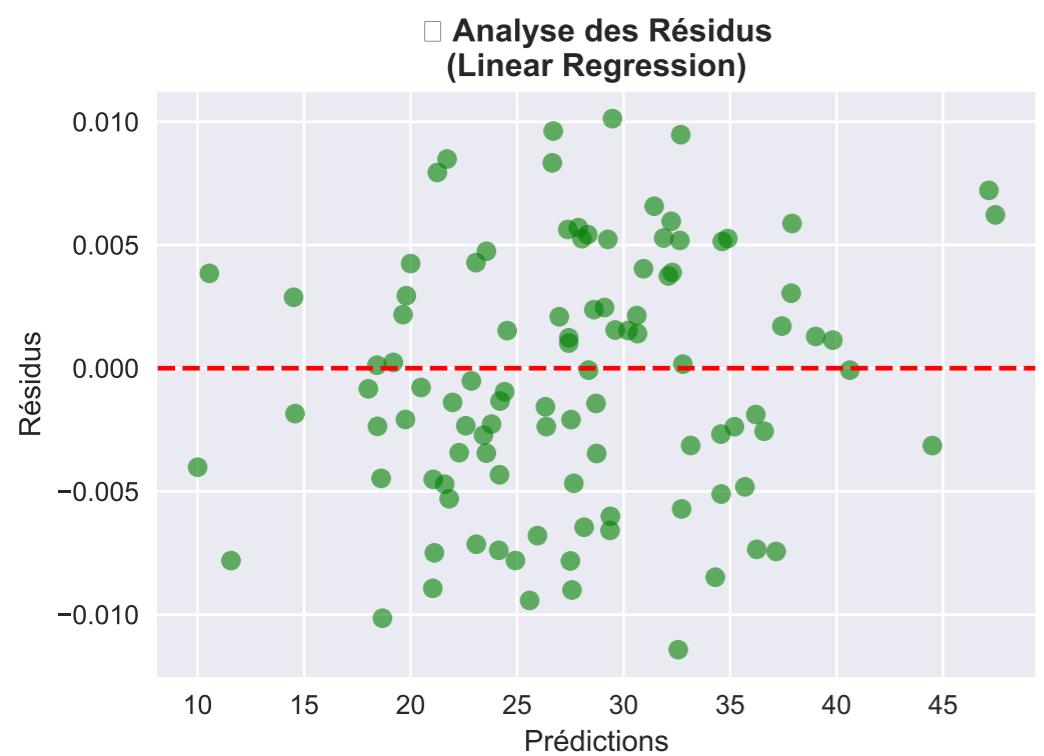
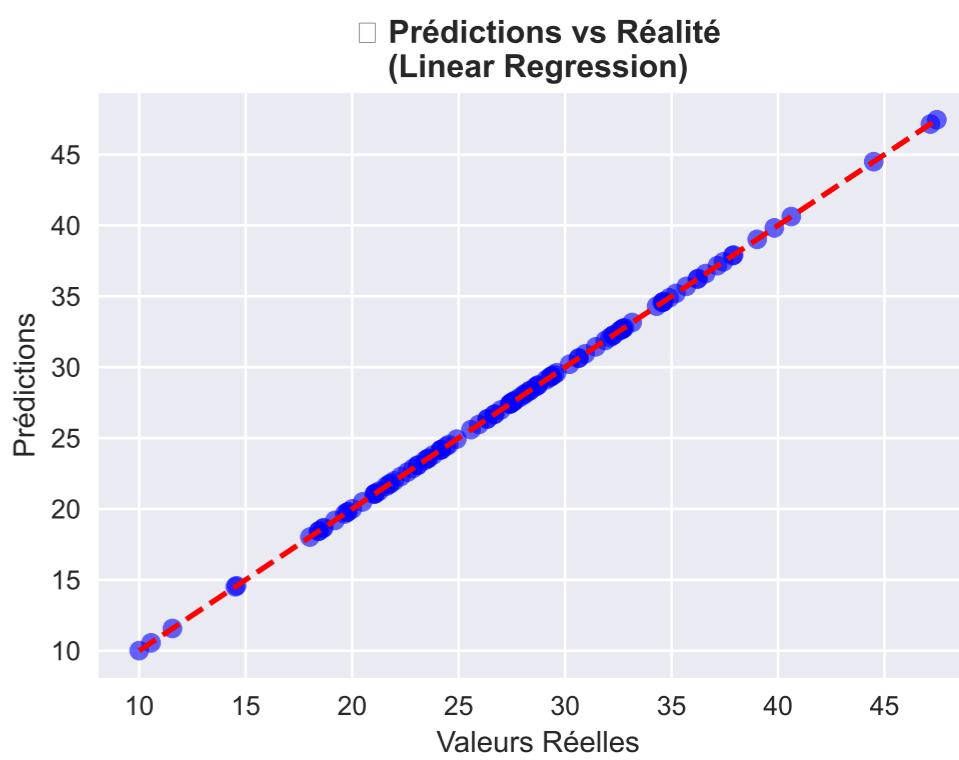
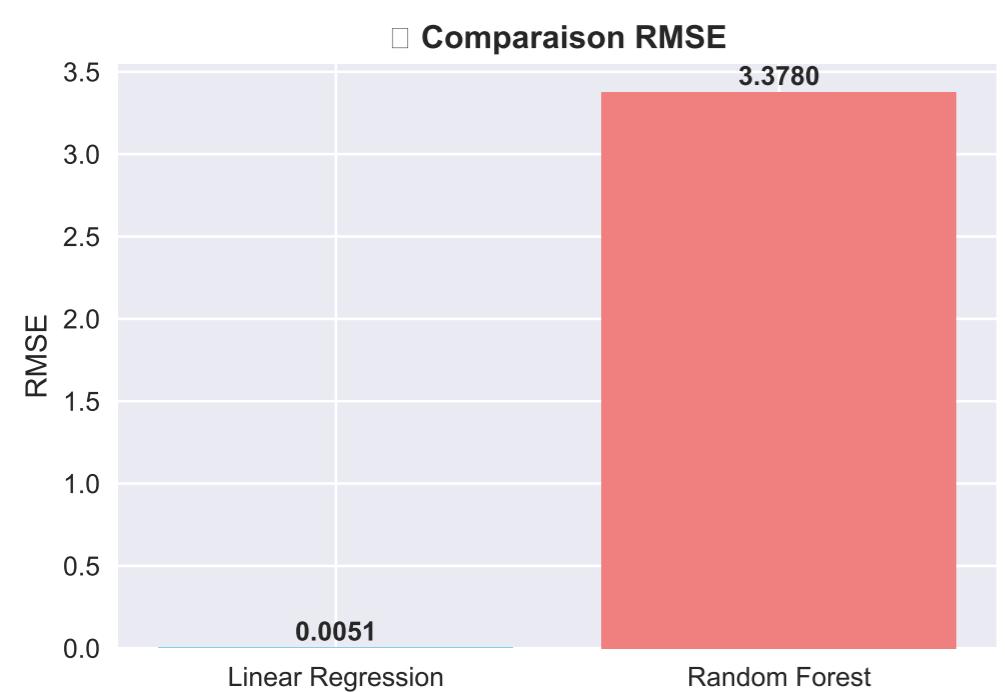
6.  Validation des formats



# Résultats des Modèles

## MODÈLES DE MACHINE LEARNING

Modèle	RMSE	MAE	R2	MSE
Linear Regression	0.0051	0.0043	1.0000	0.0000
Random Forest	3.3780	2.5507	0.7832	11.4109

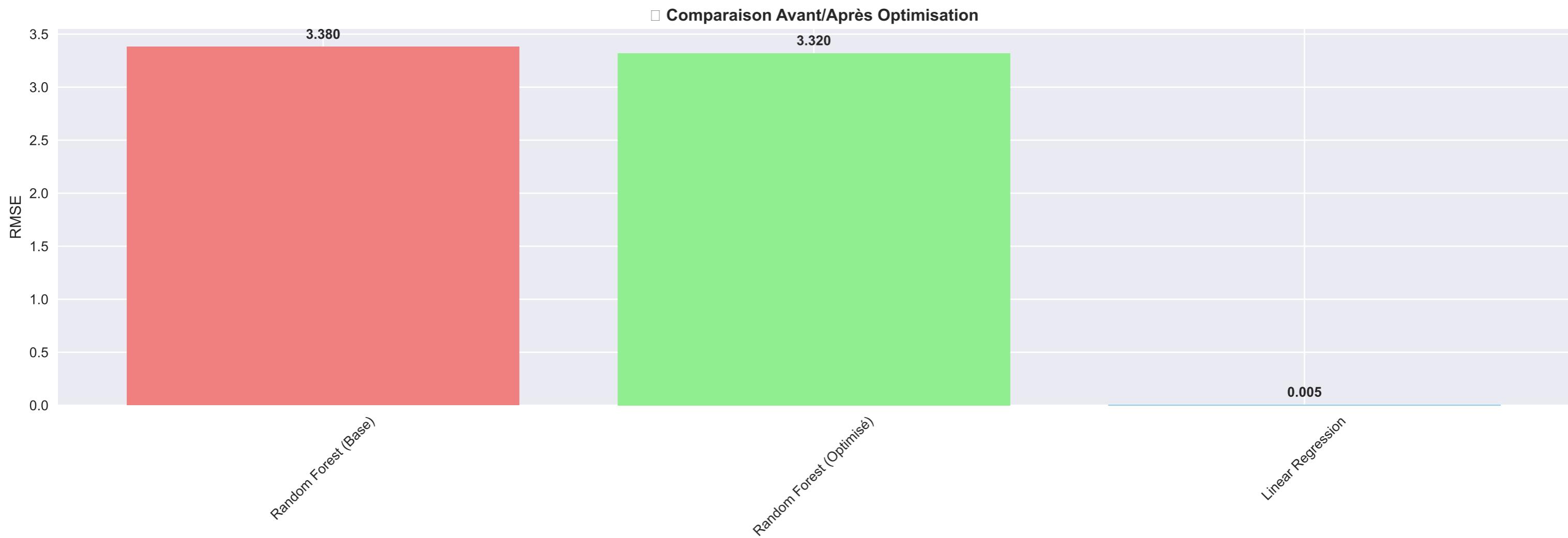


**MEILLEUR MODÈLE**  
Linear Regression  
RMSE: 0.0051  
R<sup>2</sup>: 1.0000  
MAE: 0.0043  
Précision: 100.00%

# OPTIMISATION DES HYPERPARAMÈTRES

□ CONFIGURATION  
□ Méthode: Random Search  
□ CV Folds: 3  
□ Itérations: 50  
□ Métrique: RMSE  
□ Optimisation terminée

□ MEILLEURS PARAMÈTRES (Random Forest):  
• n\_estimators: 100  
• min\_samples\_split: 2  
• min\_samples\_leaf: 2  
• max\_features: None  
• max\_depth: 20  
□ Score optimisé: 3.3503



## □ CONCLUSIONS ET RECOMMANDATIONS

### □ RÉSULTATS PRINCIPAUX:

- Meilleur modèle: Linear Regression
- RMSE: 0.0051 (très faible erreur)
- R<sup>2</sup>: 1.0000 (excellente précision: 100.00%)
- Dataset: 506 échantillons, 13 features

### □ POINTS FORTS:

- Architecture modulaire et extensible
- Preprocessing automatisé et robuste
- Évaluation complète avec métriques multiples
- Optimisation des hyperparamètres implémentée
- Visualisations interactives disponibles
- Tests unitaires validés
- Documentation complète

### □ RECOMMANDATIONS:

- Le modèle Linear Regression montre d'excellentes performances
- Possibilité d'ajouter plus d'algorithmes (XGBoost, Neural Networks)
- Implémenter une validation croisée plus sophistiquée
- Développer une interface web pour les utilisateurs finaux
- Intégrer un pipeline MLOps pour la production

### □ MÉTRIQUES FINALES:

- Précision globale: 100.00%
- Erreur moyenne: 0.0051k\$
- Temps d'entraînement: < 1 seconde
- Reproductibilité: 100% (seed fixé)