

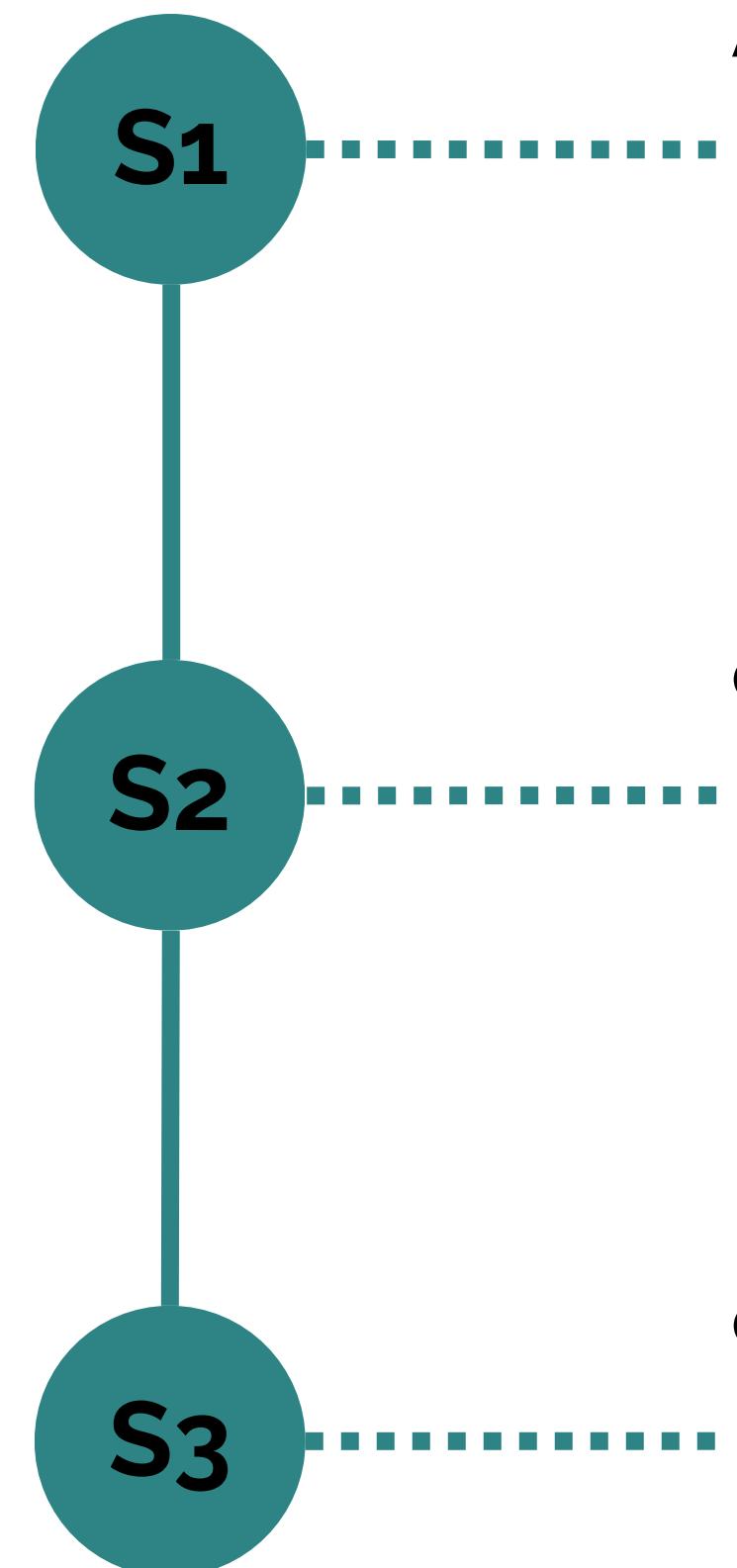


PRÉDICTION DES CHARGES D'ASSURANCE

Fatima, Umberto, Flora

Contexte

- élaboration d'un modèle d'intelligence artificielle
- prédition des charges d'assurance de nouveaux clients
- selon des critères médicaux et démographiques



Analyse de données clients :

- Fort impact du tabagisme sur les charges
- Charges élevées chez les fumeurs obèses
- Influence de l'âge sur les charges
- Asymétrie des charges

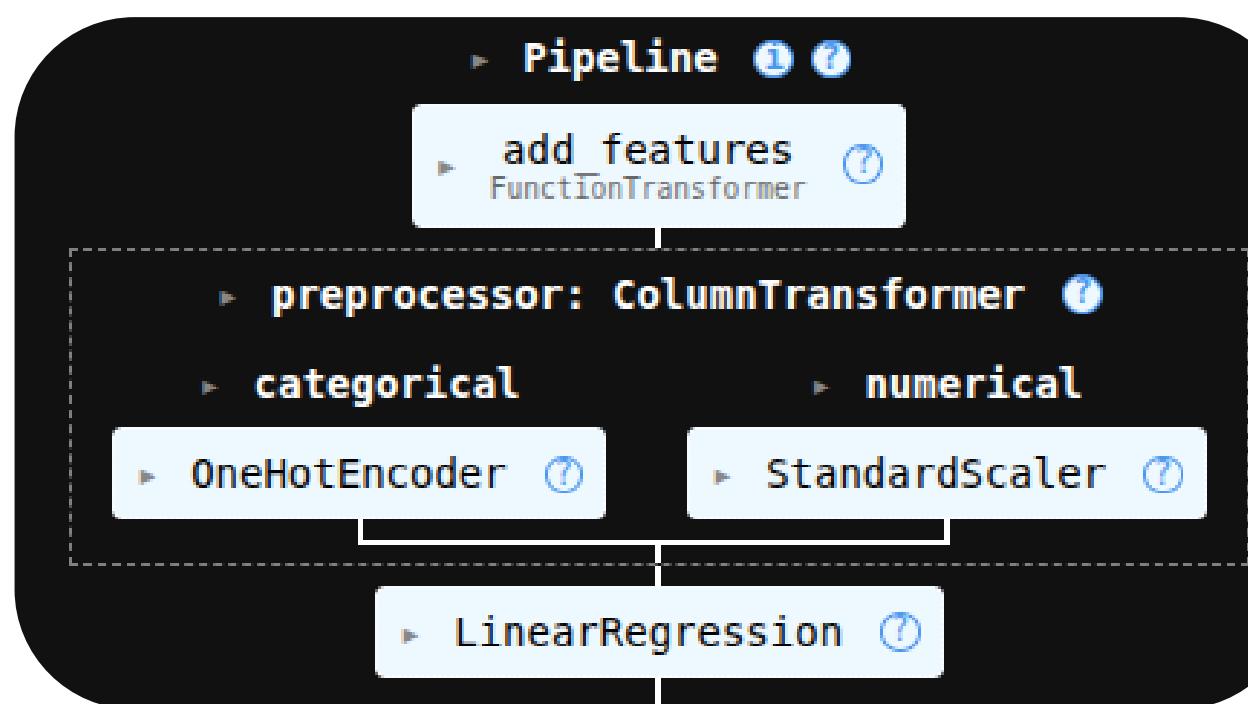
Construction d'un premier modèle :

- Nettoyage et encodage des données
- Entraînement et évaluation du modèle
- Ajout d'interaction : smoker x obese

Optimisation du modèle :

- Automatisation du traitement des données
- Comparaison de différents modèles
- Choix final et recommandations

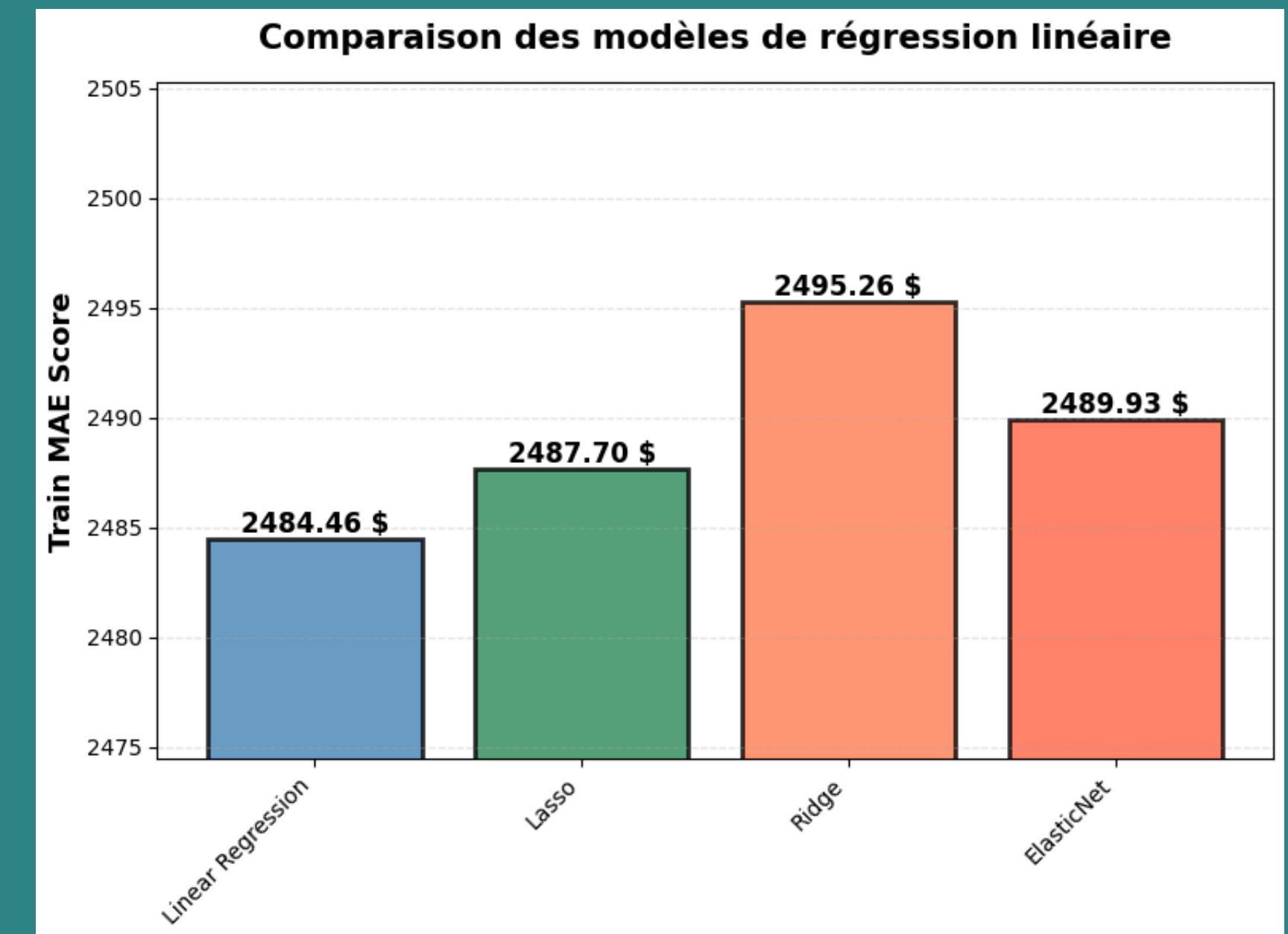
Modélisation automatisée



- Ajout de variables (age^2 , smoker x obese)
- Transformation des variables textuelles
- Mise à l'échelle des variables numériques
- Application du modèle de régression linéaire

Recherche du meilleur modèle

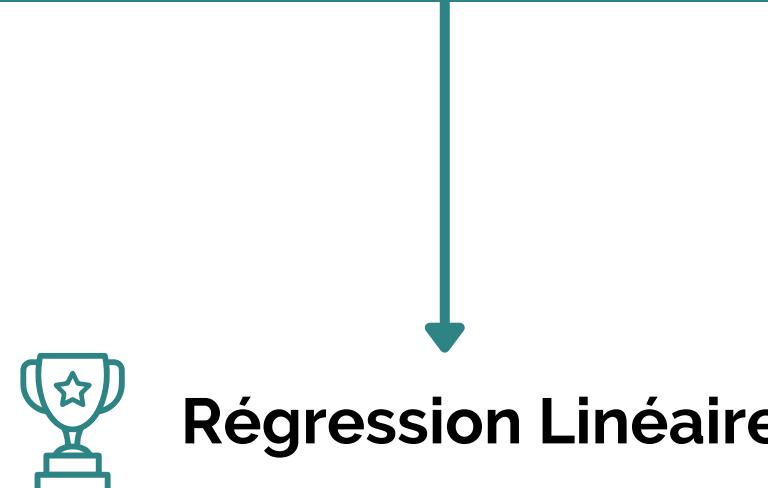
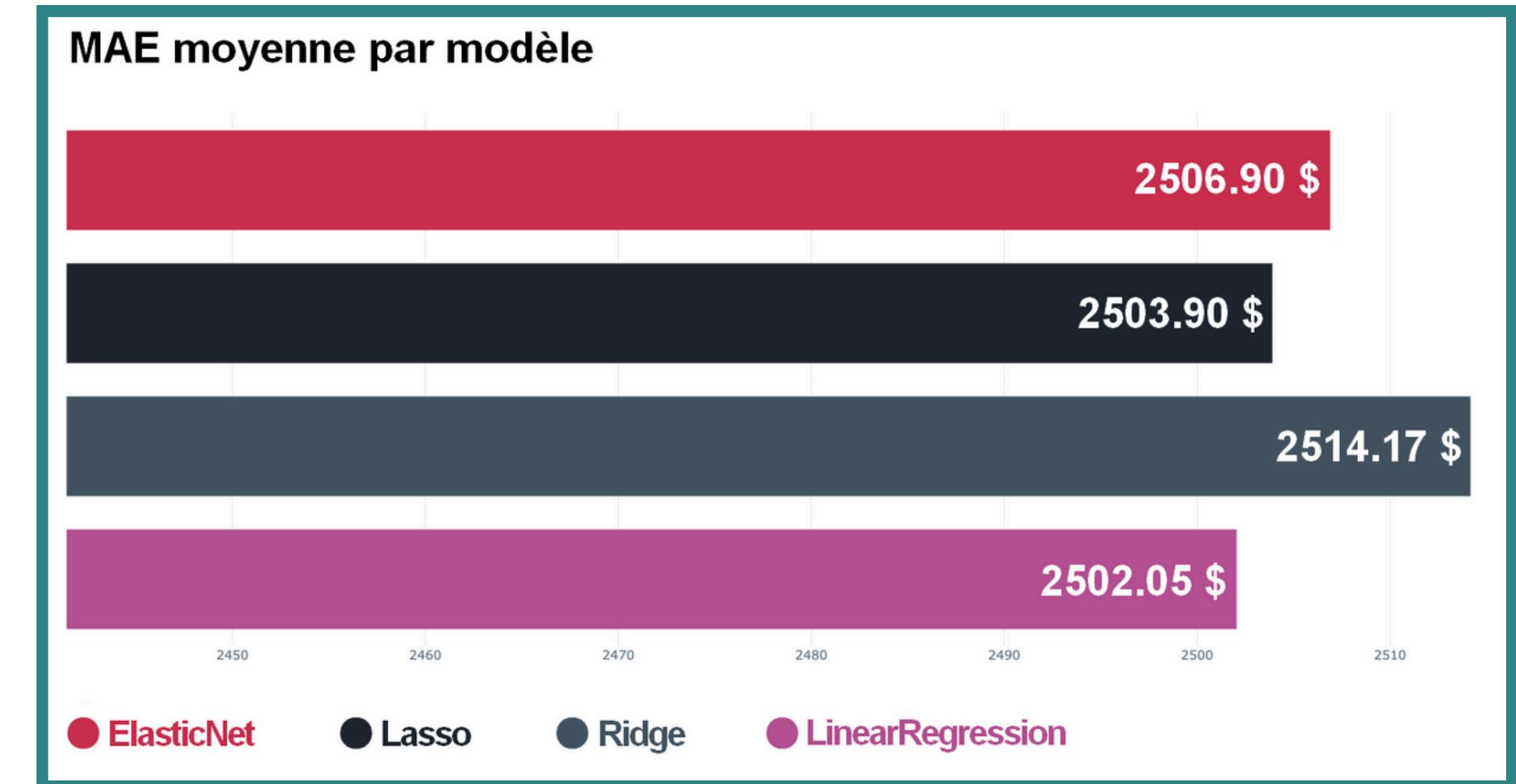
- Évaluation de 4 modèles :
 - LinearRegression
 - Ridge
 - Lasso
 - ElasticNet
- Recherche automatisée des meilleurs réglages pour chaque modèle



Comparaison des modèles

Après avoir trouvé les meilleurs réglages :

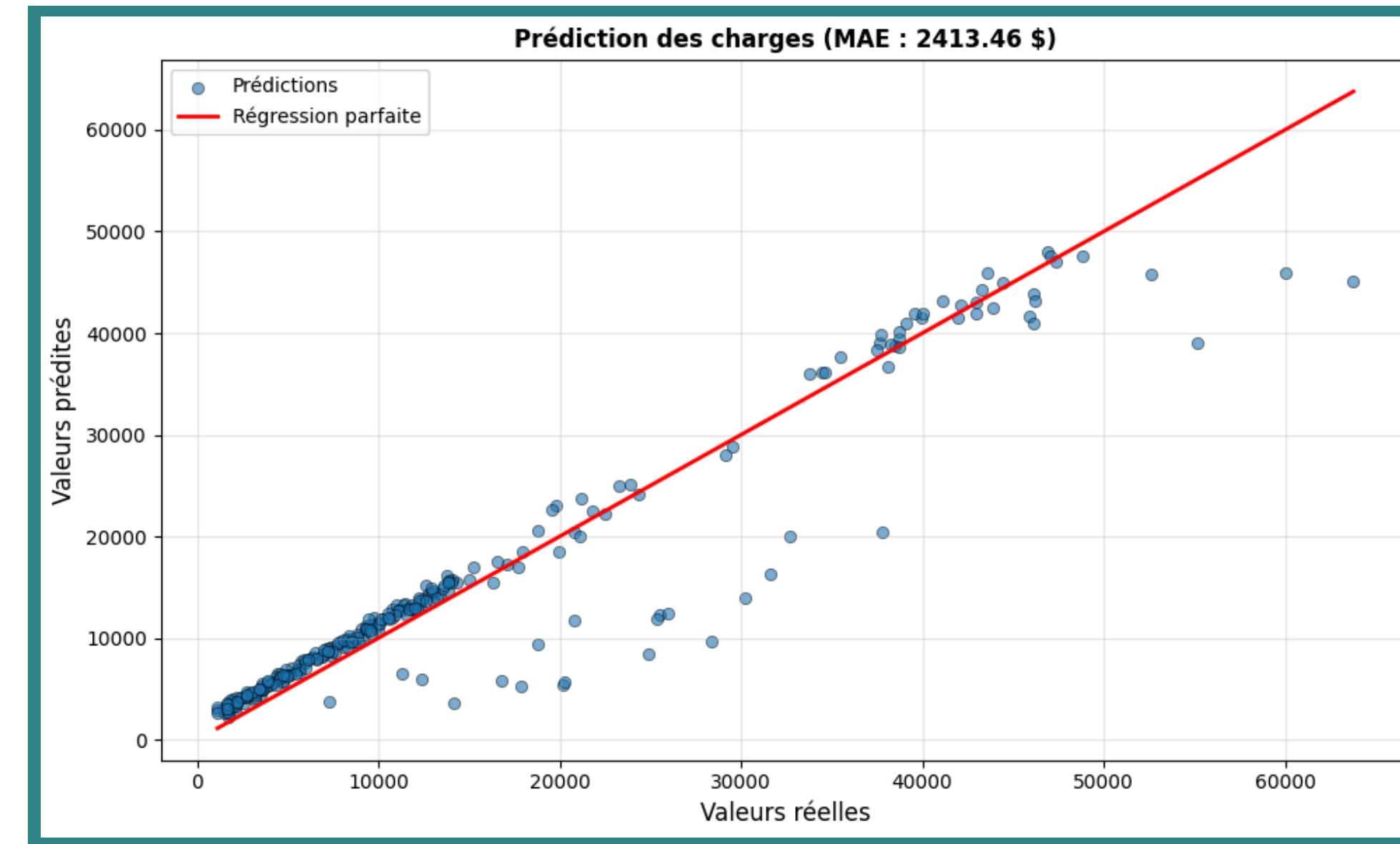
- Comparaison des 4 modèles
- Scores moyens sur chaque sous-ensemble des données d'entraînement



Régression Linéaire

Écart moyen entre charges prédites et charges réelles

2413 \$



Conclusion



Modèle prêt à être déployé

Fonctionnalités de sauvegarde et chargement du modèle pour le déploiement.

Axes d'amélioration

Analyse des erreurs résiduelles

- Quelques charges sont fortement sous-estimées par le modèle
- Pas d'explication apparente en se basant sur nos données ?



Analyse de nouvelles données

- Entraînement sur une plus grande quantité de données clients
- Modèle V2 : autres critères pouvant expliquer les grandes erreurs

Encore un peu d'attention !

Petite démonstration