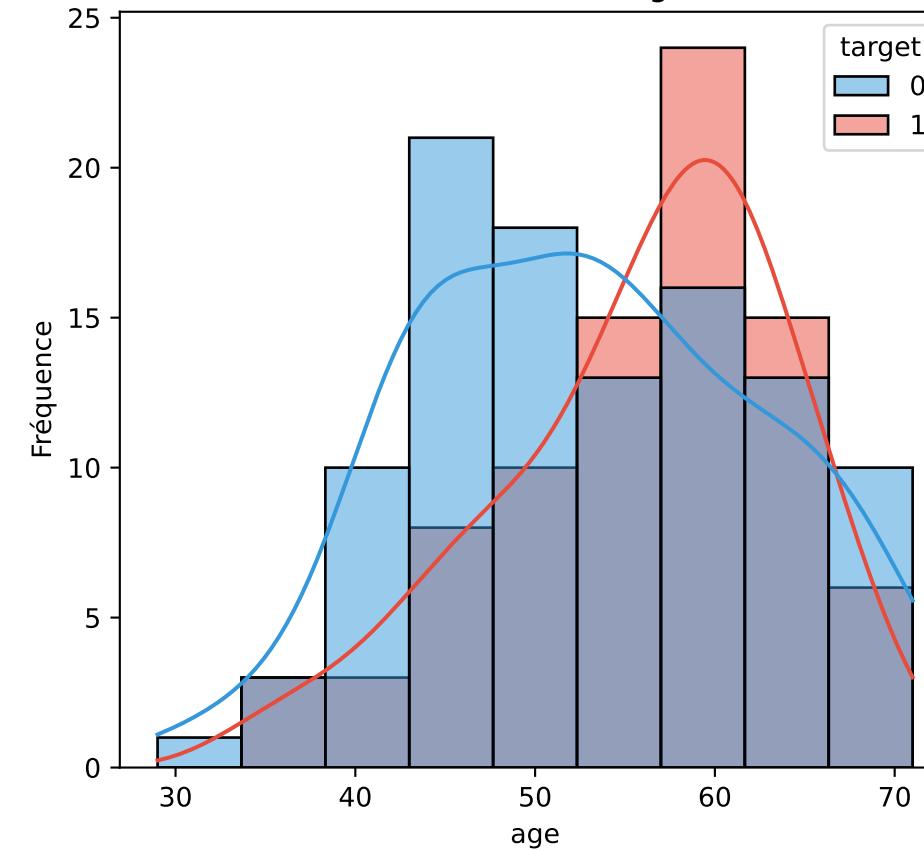


Analyse et Prédition des Maladies Cardiaques

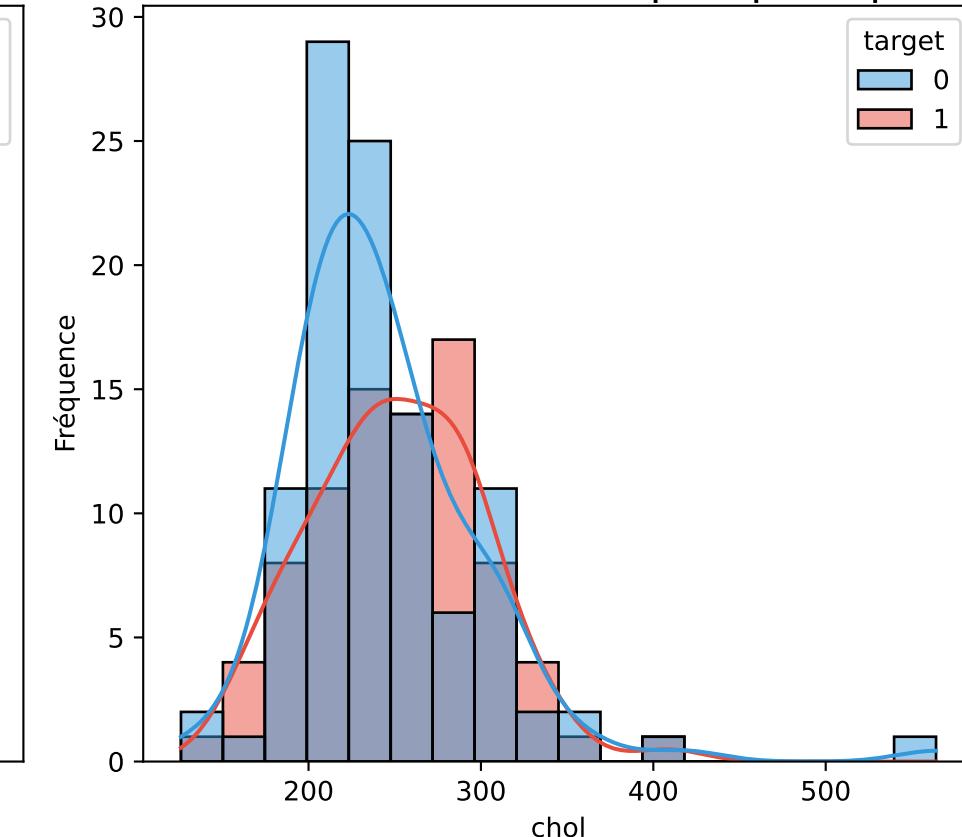
Rapport des Visualisations

Date: Avril 2025

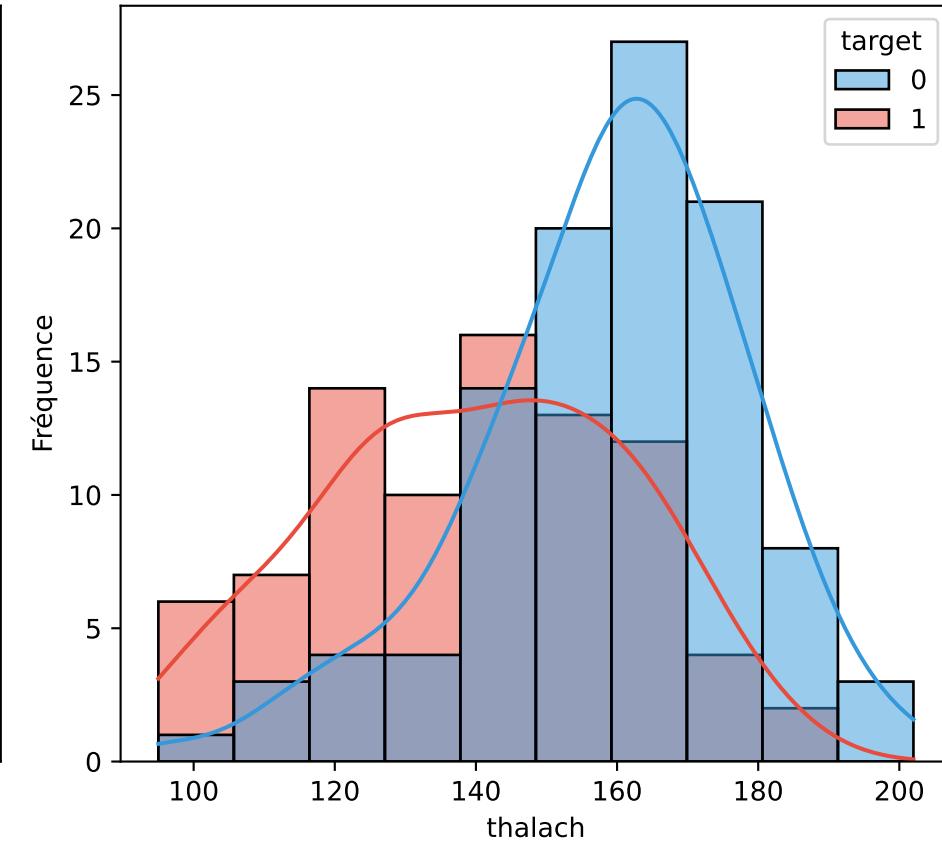
Distribution de age



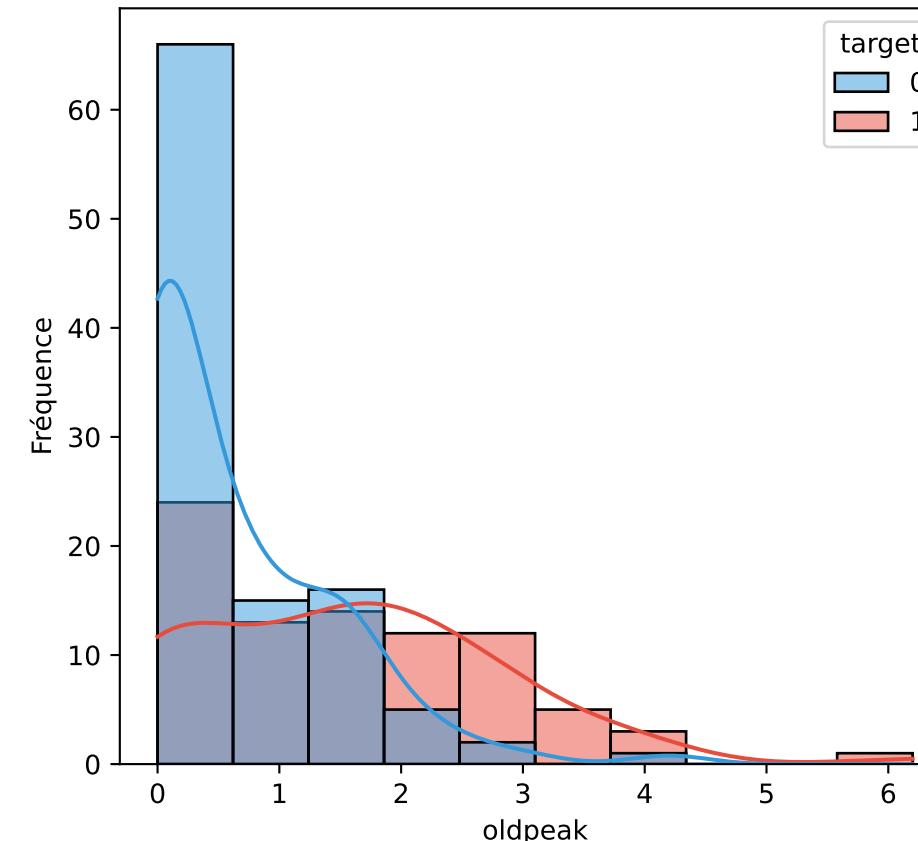
Distribution des caractéristiques principales



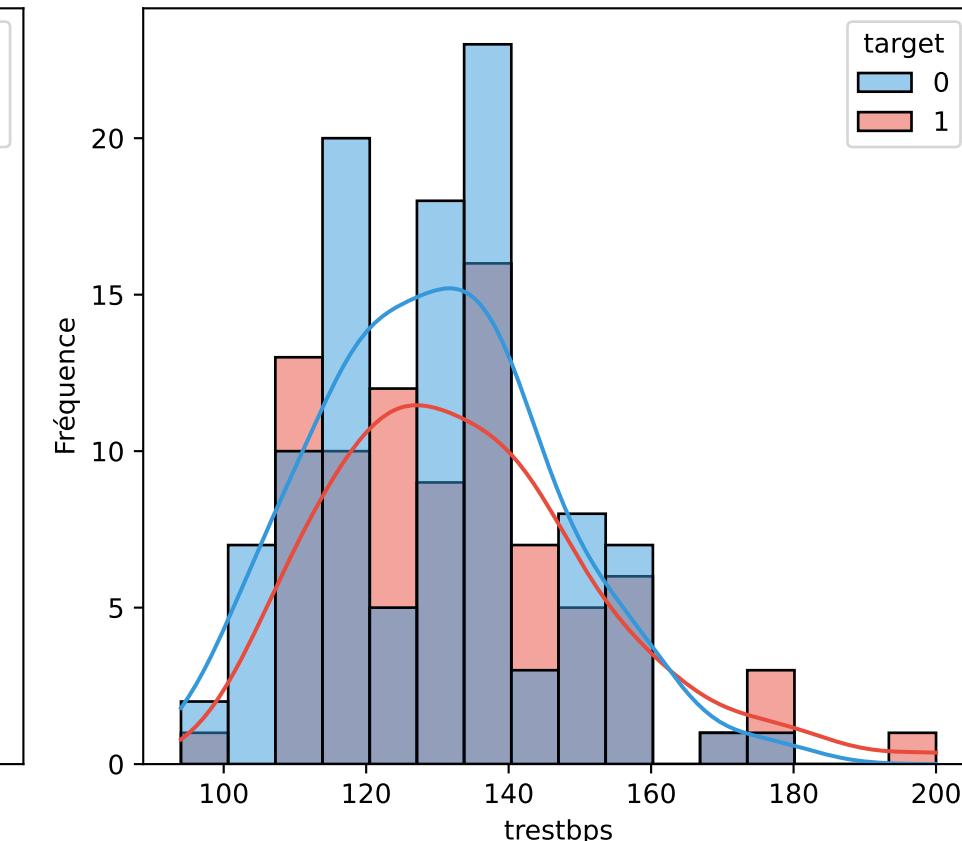
Distribution de thalach



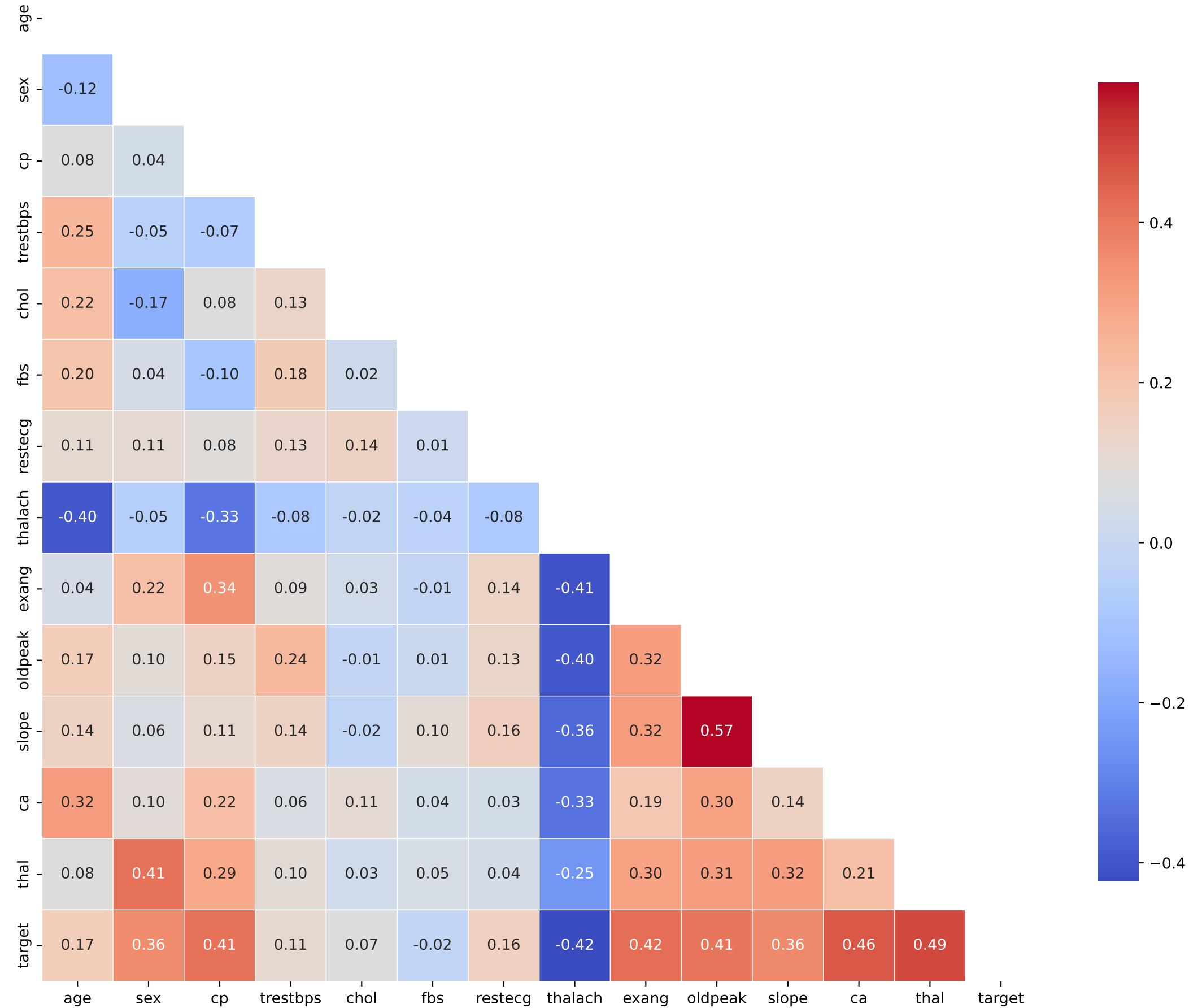
Distribution de oldpeak



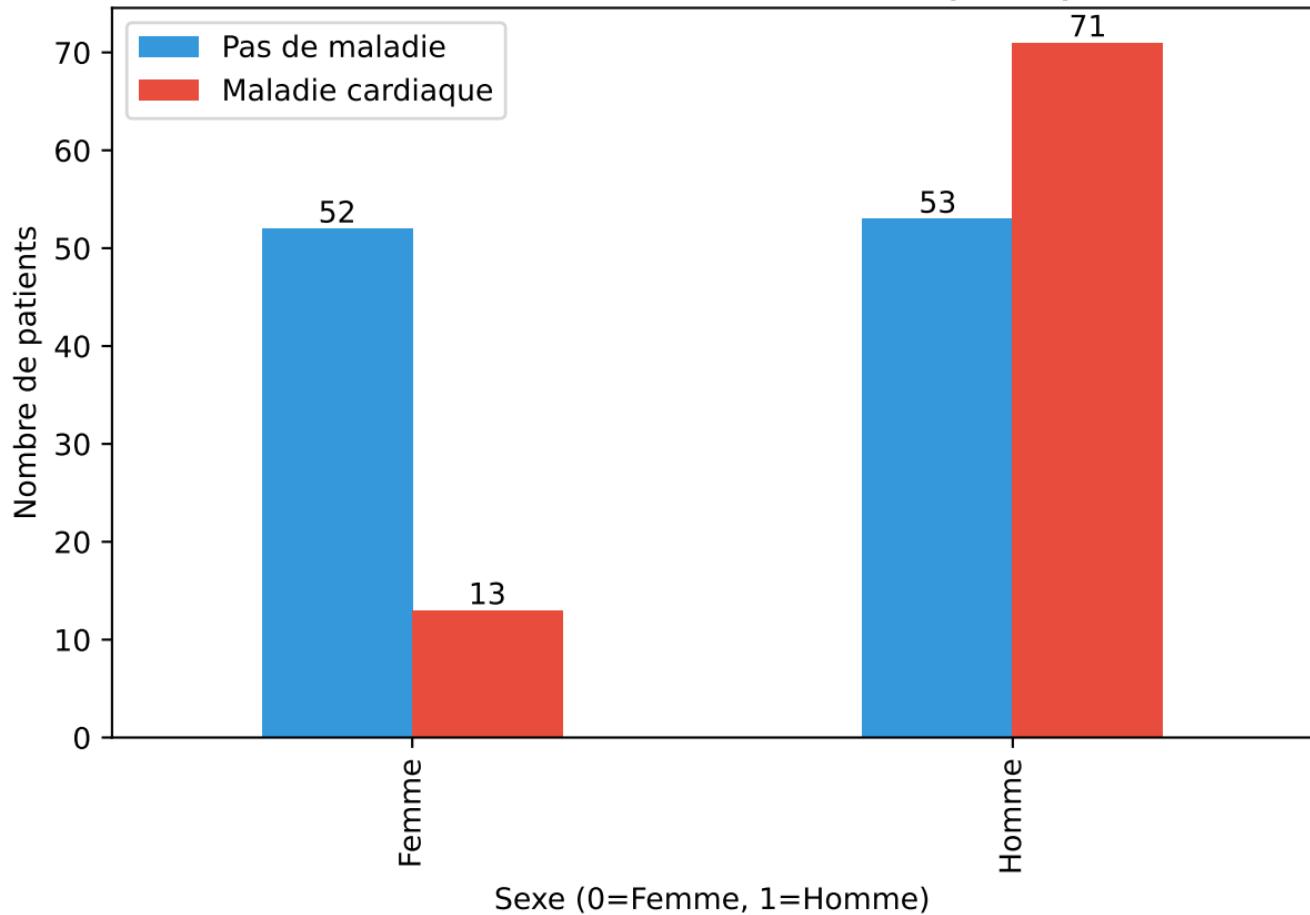
Distribution de trestbps



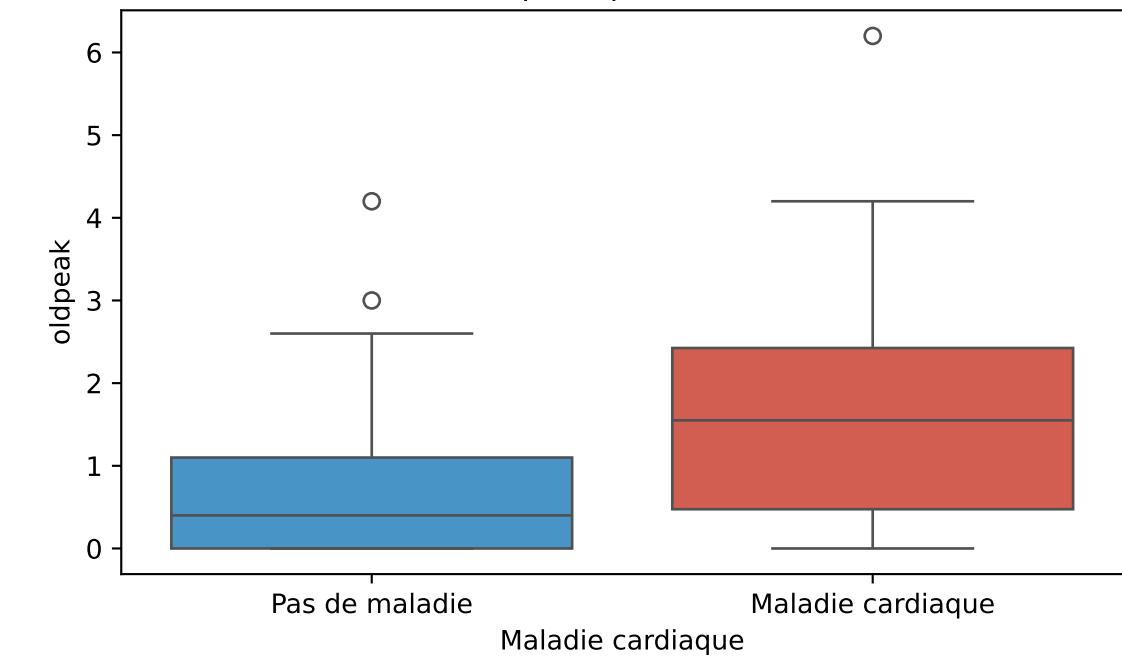
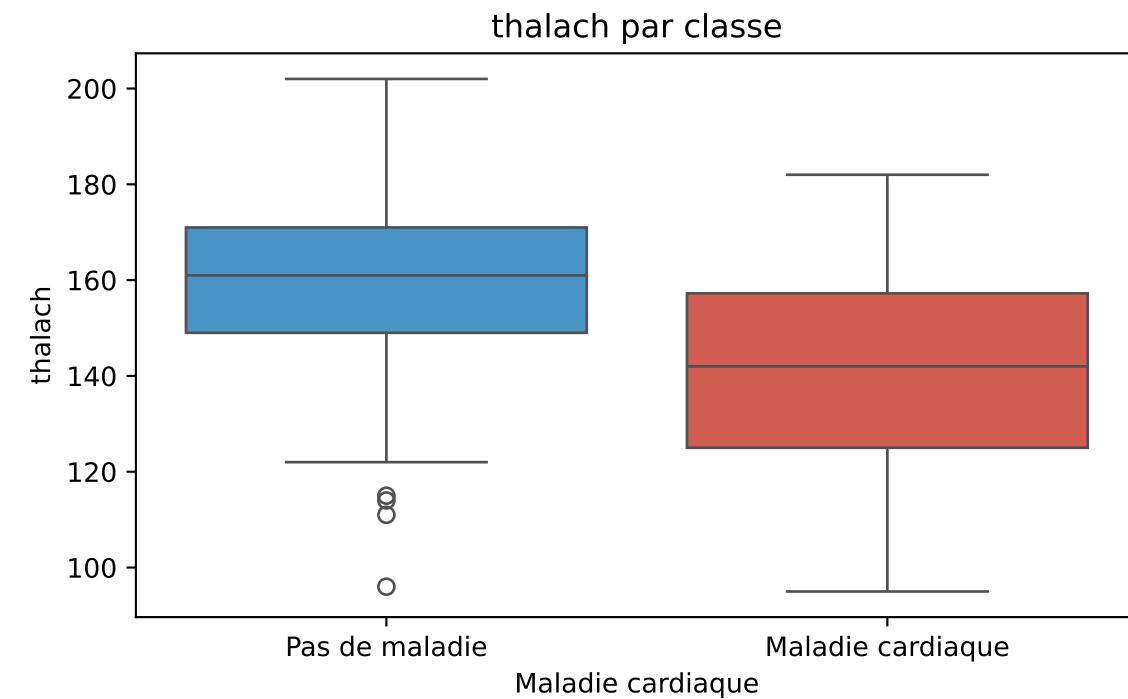
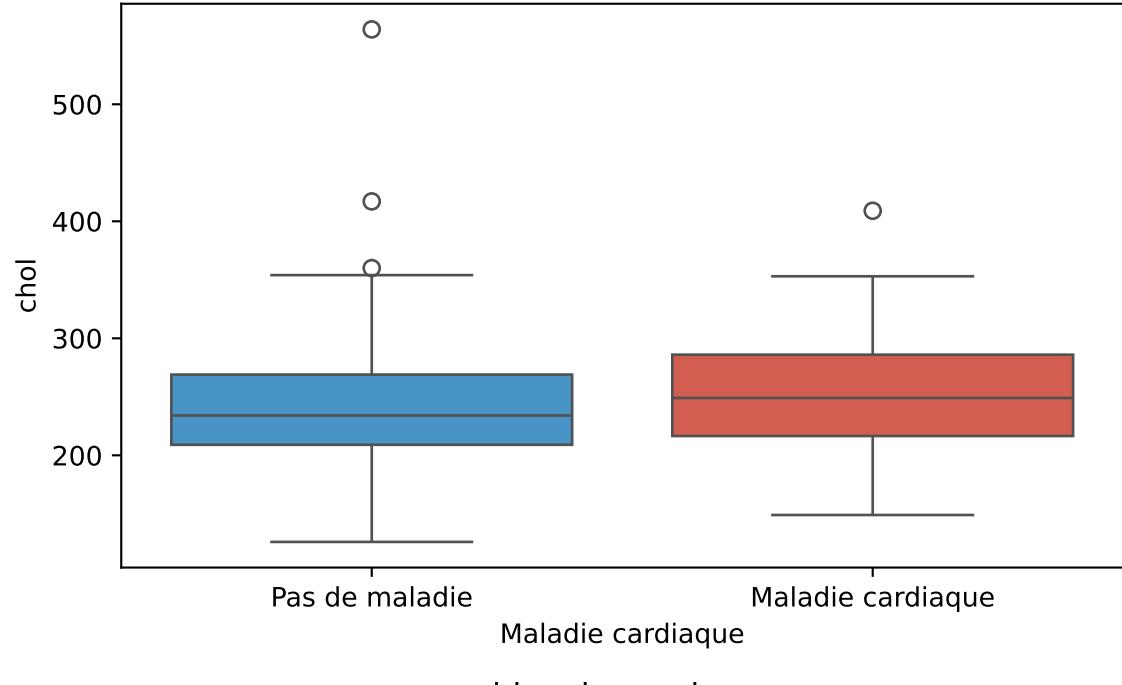
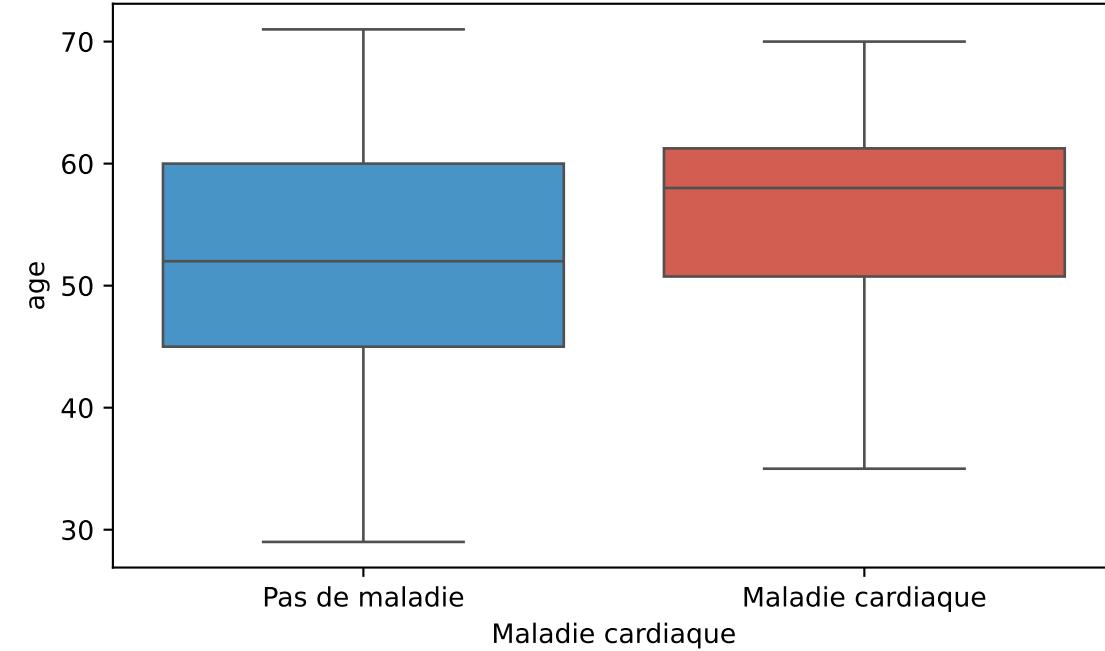
Matrice de corrélation des caractéristiques



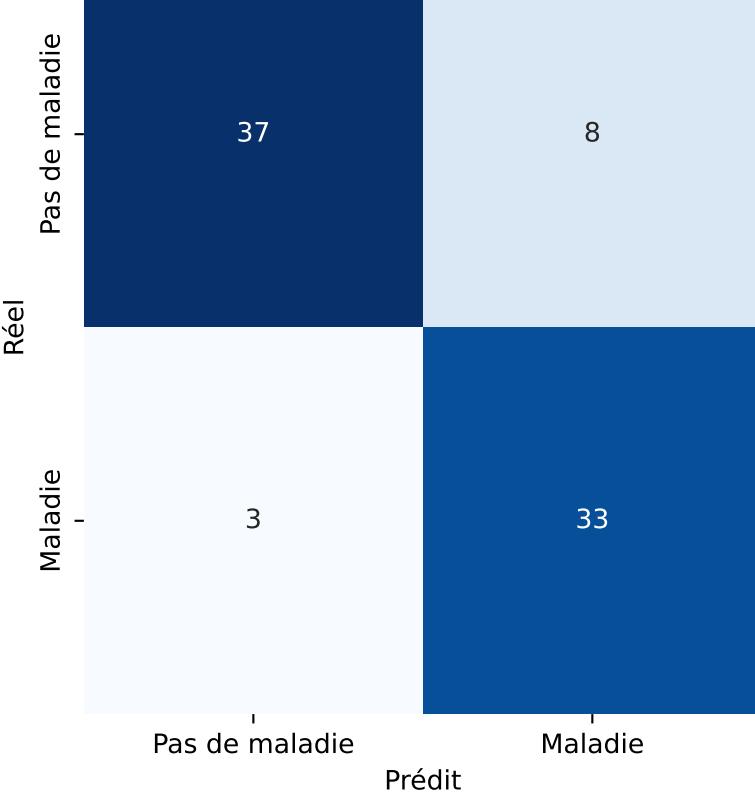
Distribution des maladies cardiaques par sexe



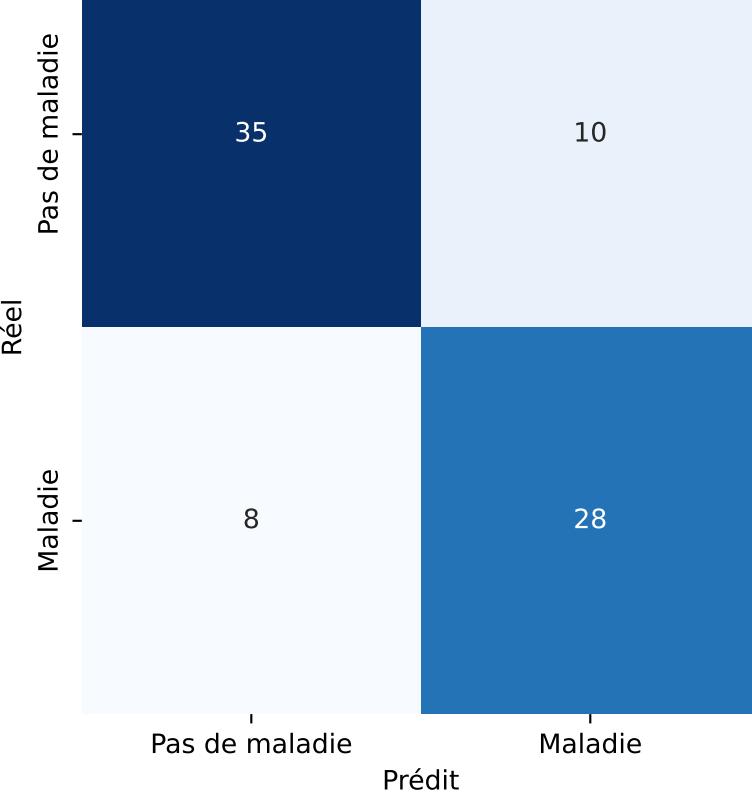
Distribution des caractéristiques par classe



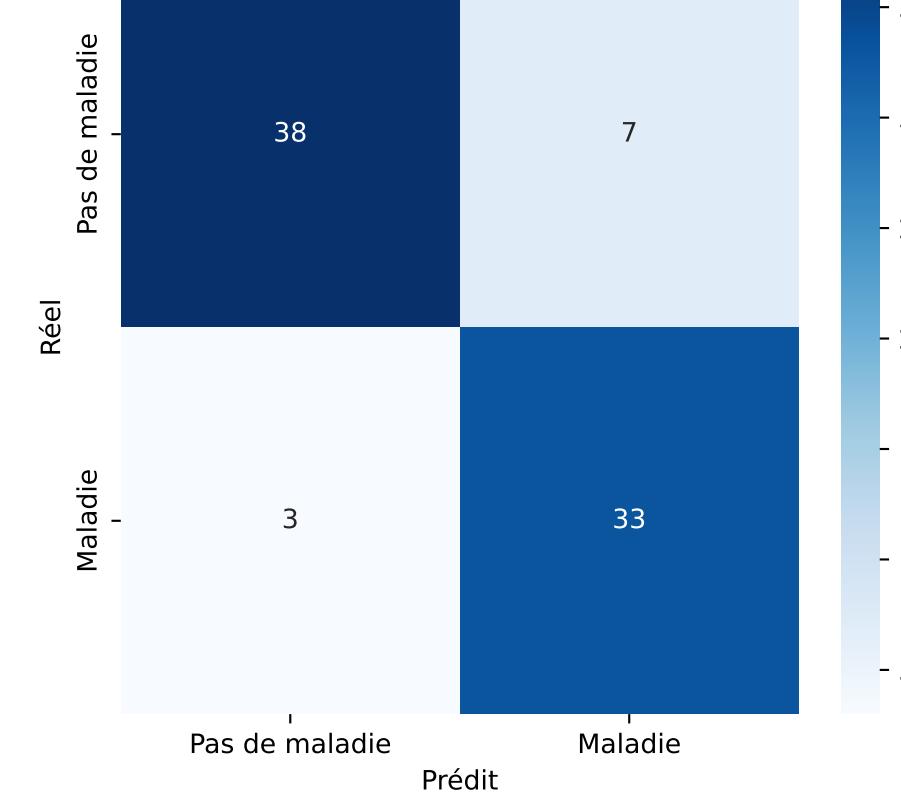
Matrice de confusion - Logistic Regression



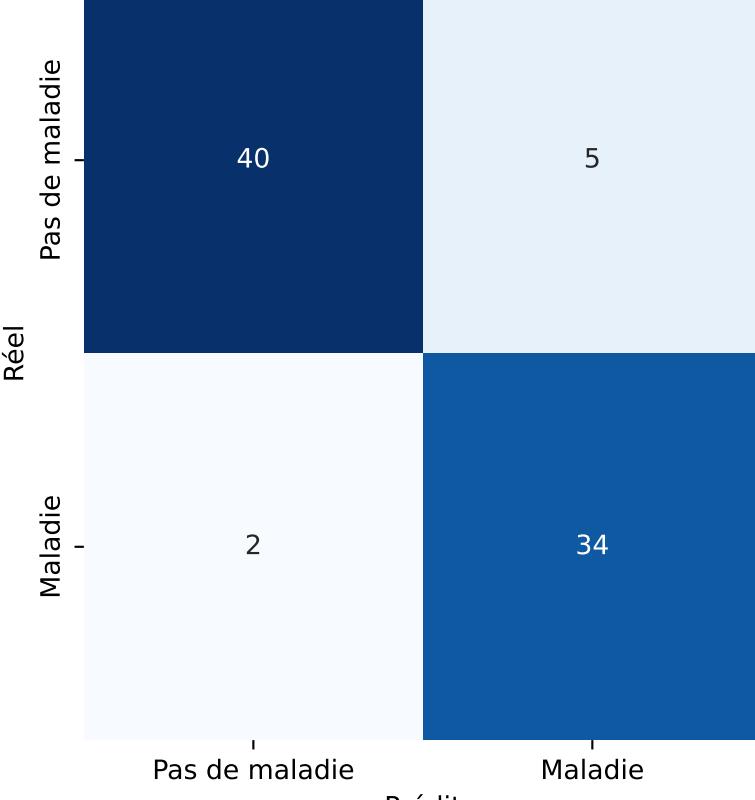
Matrice de confusion - Decision Tree



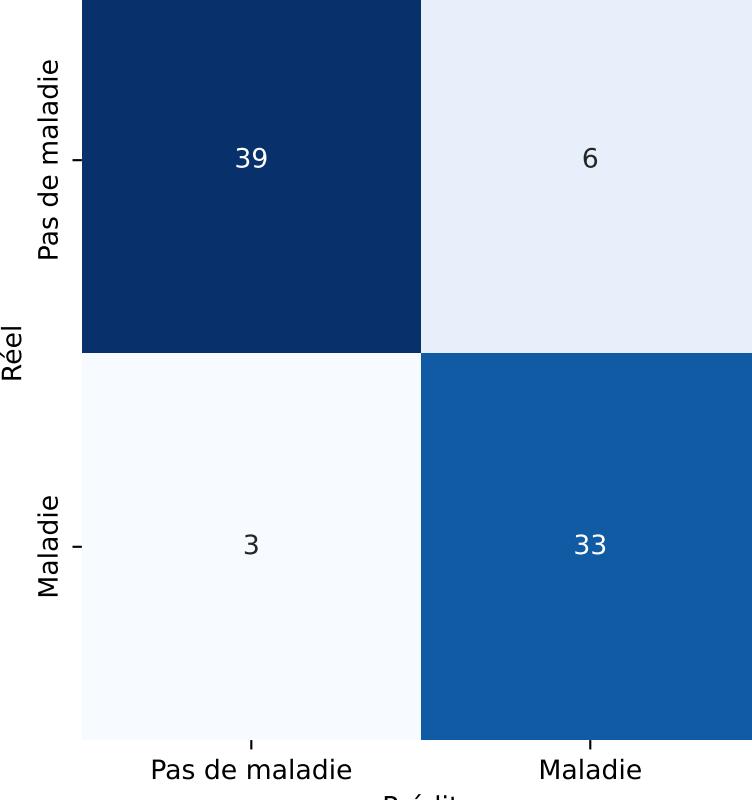
Matrice de confusion - KNN



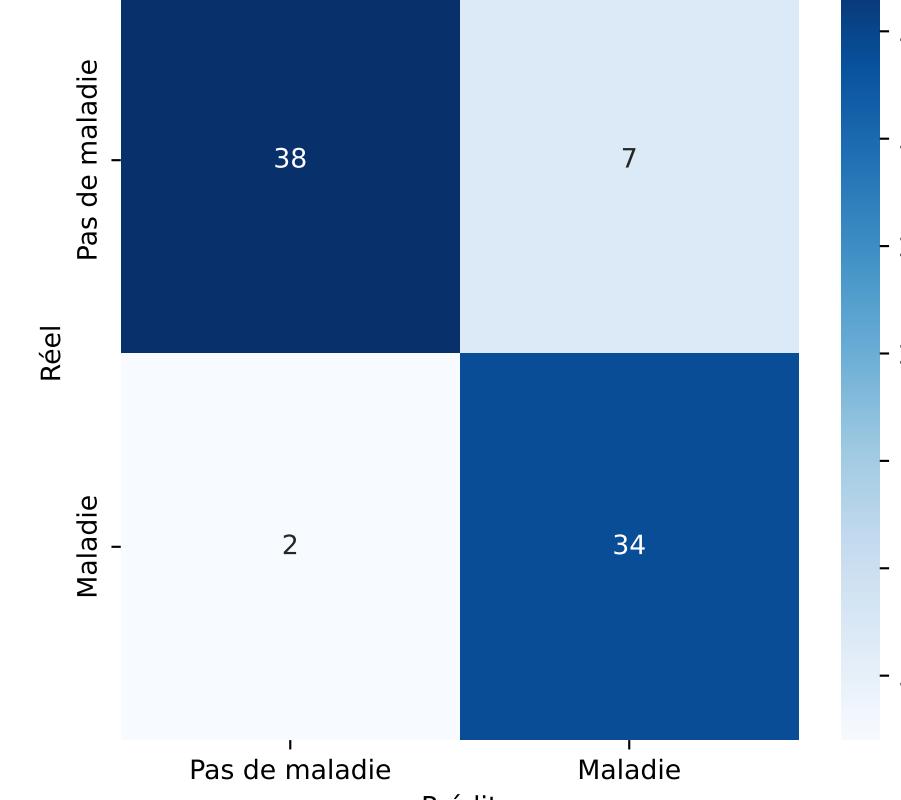
Matrice de confusion - Random Forest



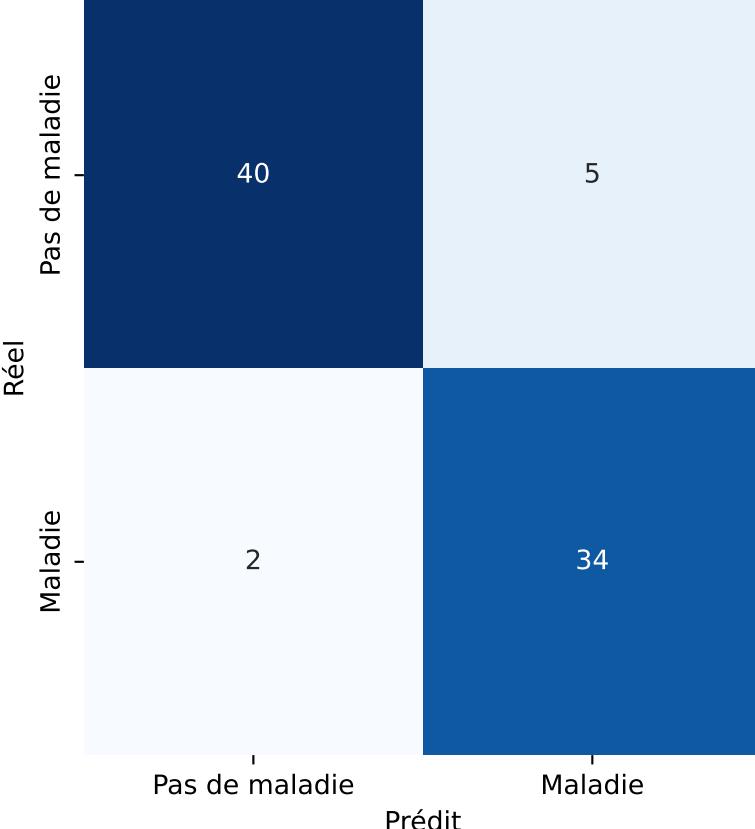
Matrice de confusion - AdaBoost



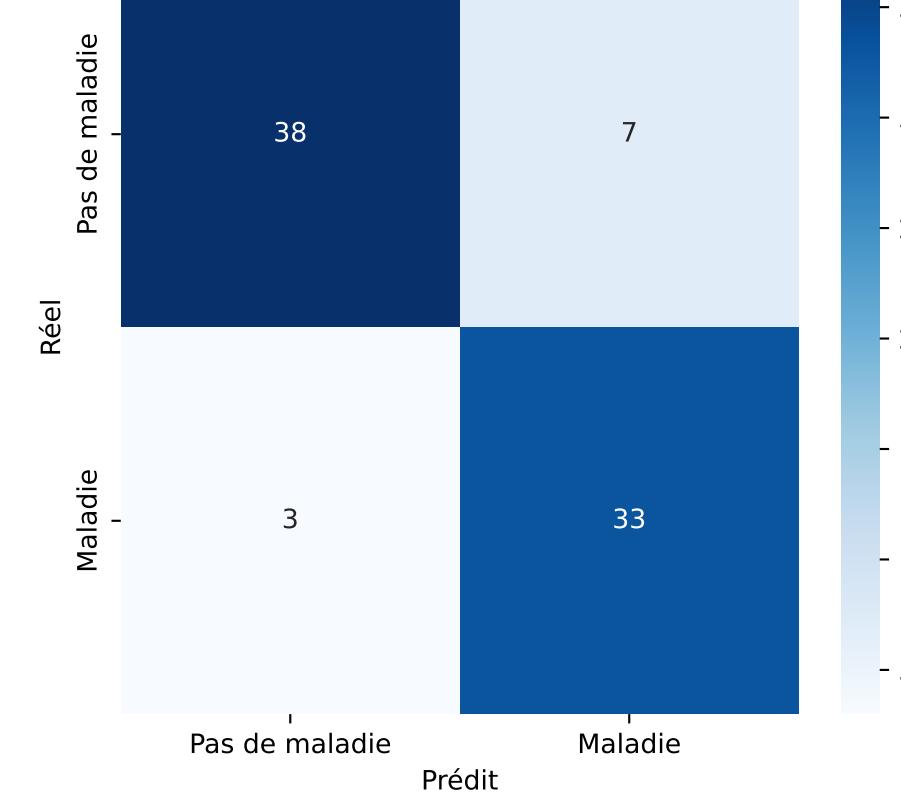
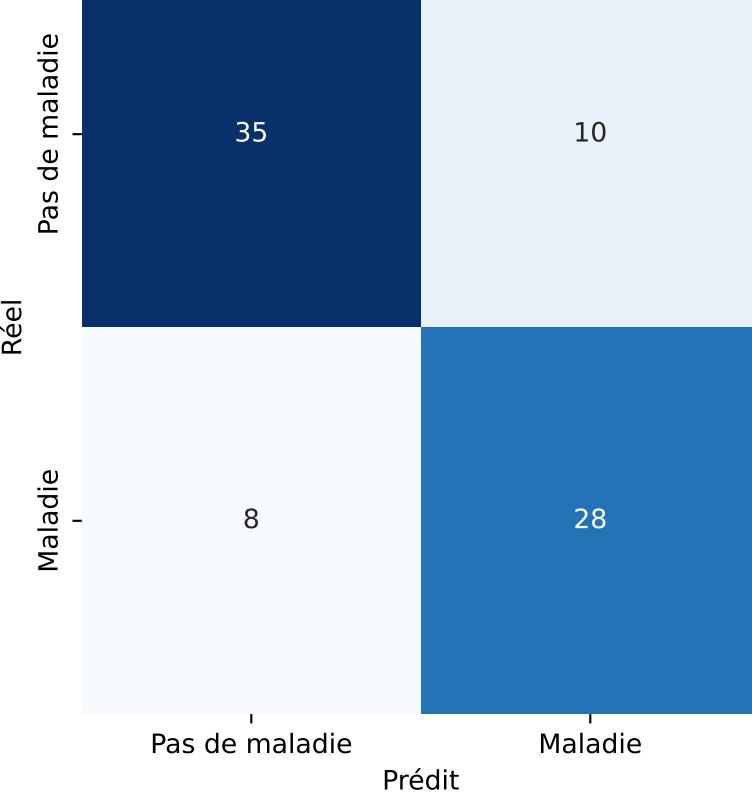
Matrice de confusion - SVM



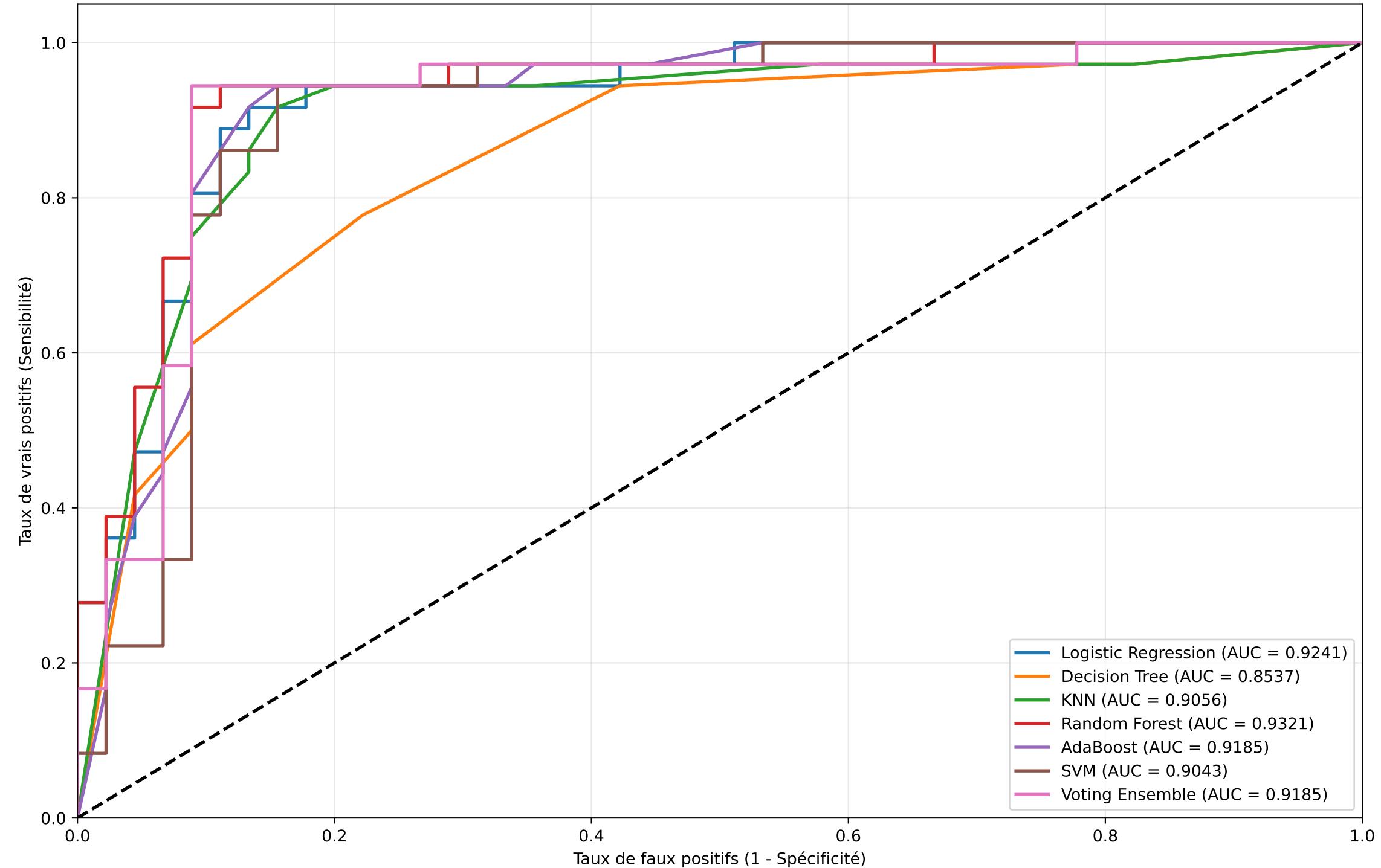
Matrice de confusion - Voting Ensemble



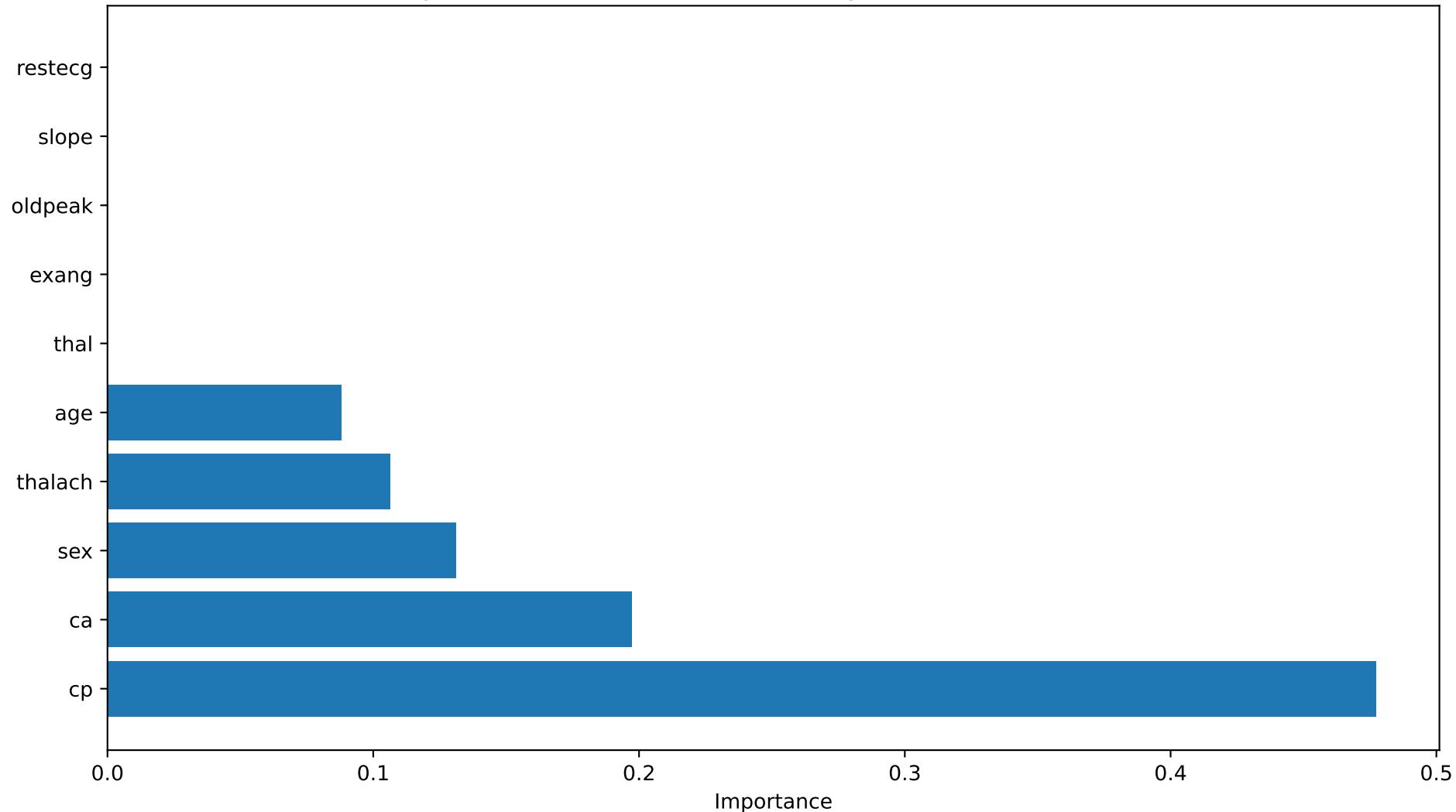
Matrices de confusion des différents modèles



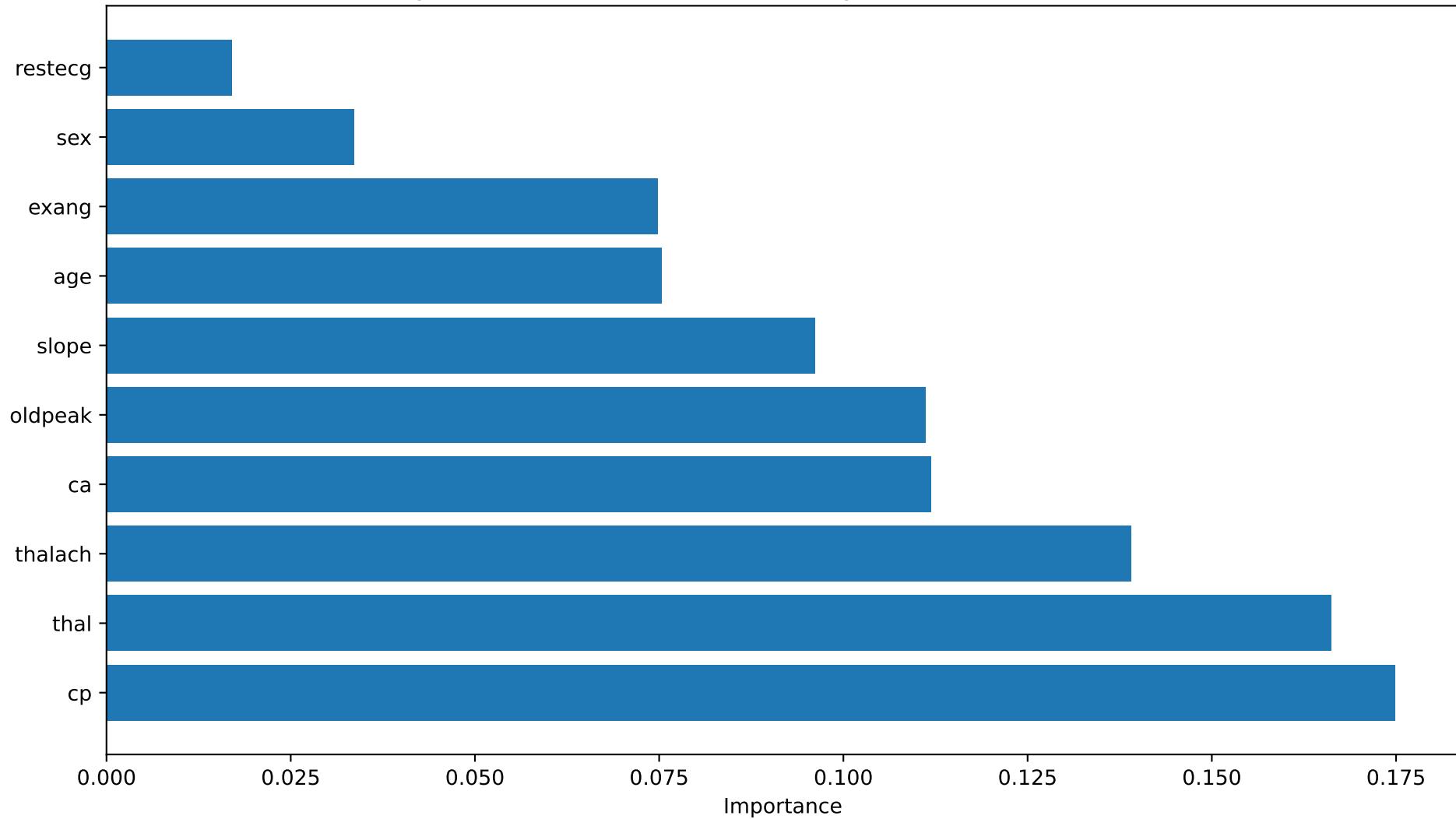
Courbes ROC des différents modèles



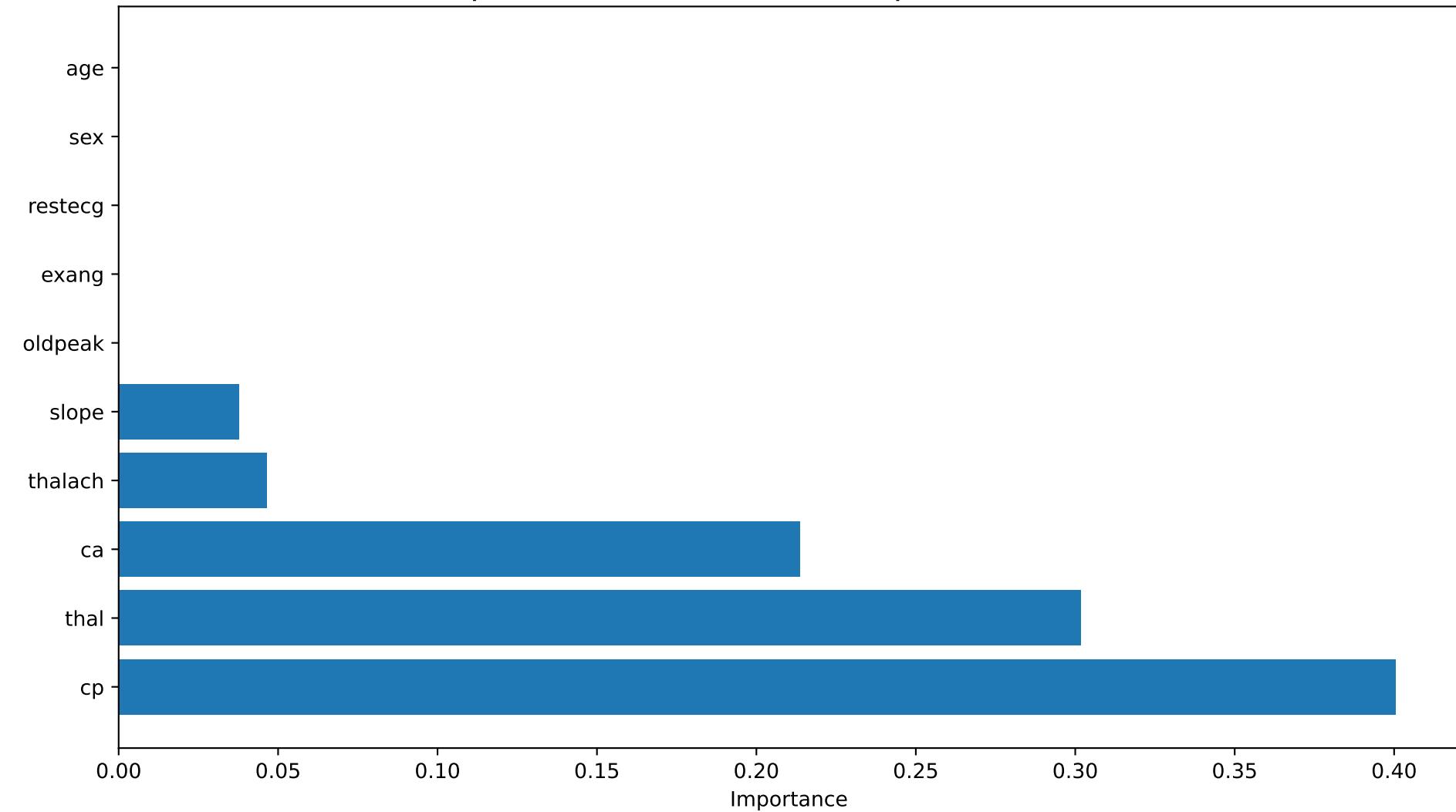
Importance des caractéristiques - Decision Tree



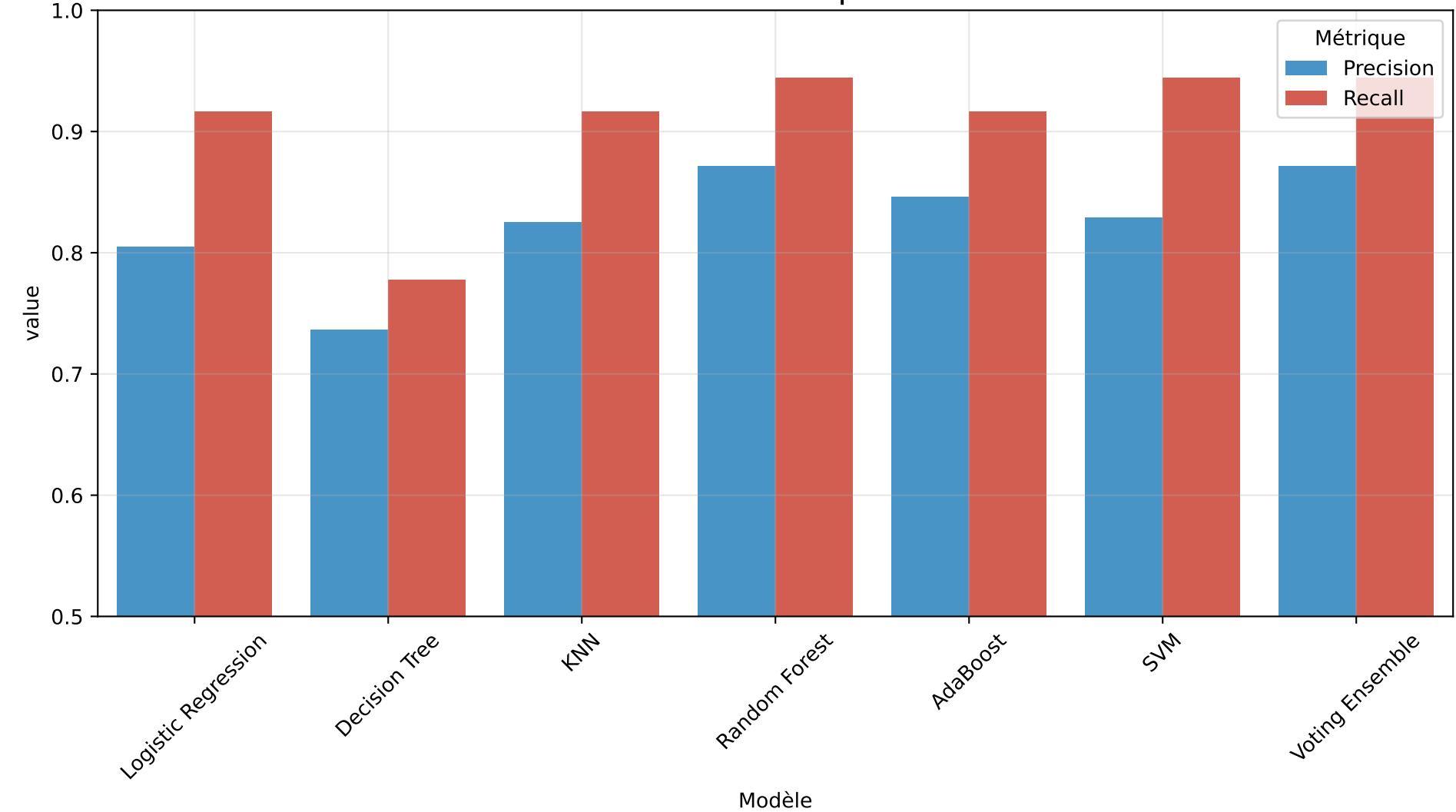
Importance des caractéristiques - Random Forest



Importance des caractéristiques - AdaBoost



Precision vs Recall par modèle



Interprétation des résultats

1. Les caractéristiques les plus discriminantes:

- La fréquence cardiaque maximale (thalach) est généralement plus basse chez les patients malades
- Le type d'angine de poitrine (cp) est fortement corrélé avec la présence de maladie
- La dépression du segment ST (oldpeak) est plus élevée chez les patients malades

2. Différences entre les sexes:

- Les hommes présentent plus souvent des maladies cardiaques que les femmes
- Le profil des symptômes peut varier selon le sexe

3. Performance des modèles:

- Les modèles ensemble (Random Forest, AdaBoost) montrent généralement de meilleures performances
- Le compromis Precision/Recall varie selon les modèles
- L'AUC est un bon indicateur de la capacité discriminante globale

4. Limites de l'analyse:

- Taille relativement limitée des échantillons
- Certaines caractéristiques peuvent être interdépendantes
- Risque de surapprentissage sur certains modèles (KNN notamment)