

Heva



Avec le soutien institutionnel
d'Astellas Pharma France

Le Machine Learning : Exemples de mise en application et perspectives

Intelligence Artificielle et Néphrologie

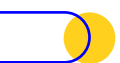
14 et 15 septembre 2023 à Paris
Marie Laurent - Data Scientist

MAT-FR-NON-2023-00297



DOCAPOSTE

l'avenir devient plus simple



Heva : un triple ADN



Analyses de données de vie réelle

**En particulier
sur le SNDS**



Data visualisation

**Pour l'interprétation
et l'ergonomie**



Approches innovantes

**Data science,
intelligence artificielle**

Exemple d'une étude en psychiatrie : Prédiction du risque de réhospitalisation

Data

Un hôpital français de taille moyenne
Base de données composée de 33 000 hospitalisations en
psychiatrie

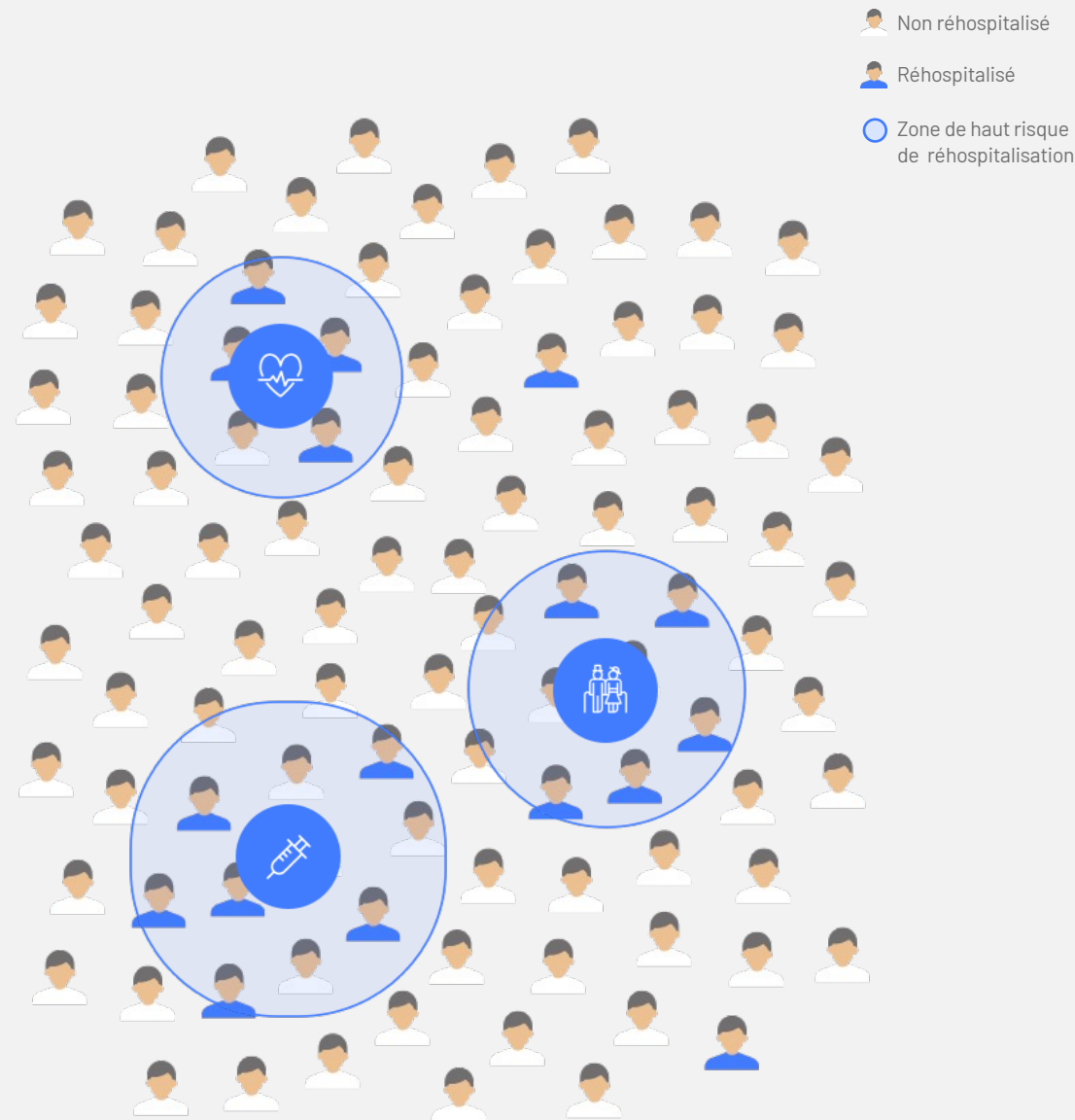
Outcome

Une réhospitalisation dans les 2 ans
pour la moitié des patients



Question d'intérêt pour les médecins

Qui sont les patients le plus à risque de réhospitalisation ?





Comment répondre à la question de recherche ?

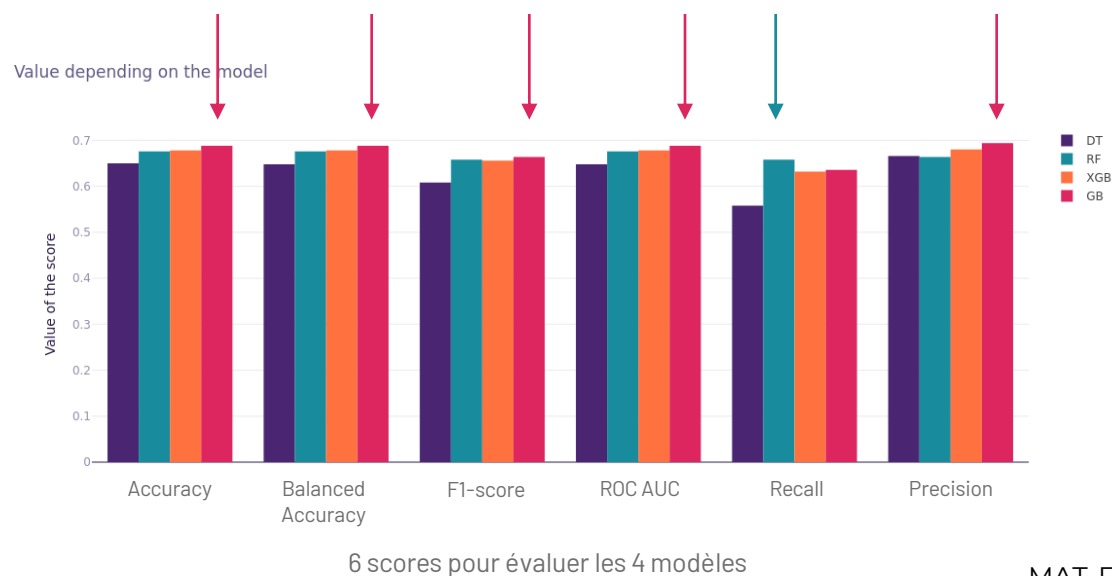
4 algorithmes de machine learning à l'essai

DT Decision Tree

RF Random Forest

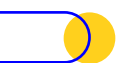
XGB Extreme Gradient Boosting

GB Gradient Boosting



Le grand gagnant est le Gradient Boosting : classé 1^{er} sur 5 des 6 scores évalués.

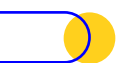
Il peut raisonnablement bien prédire le risque de réhospitalisation des nouveaux patients.



Le résultat d'un modèle de machine learning prédictif...

... c'est une boîte noire, avec des patients en entrée, et un risque en sortie

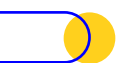




Le résultat d'un modèle de machine learning prédictif...

... c'est une boîte noire, avec des patients en entrée, et un risque en sortie





Le résultat d'un modèle de machine learning prédictif...

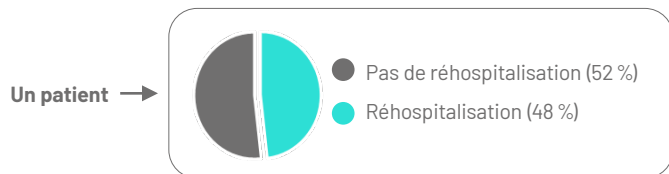
... c'est une boîte noire, avec des patients en entrée, et un risque en sortie



Voulons-nous prédire ou voulons-nous aussi comprendre ?

Prédiction au hasard

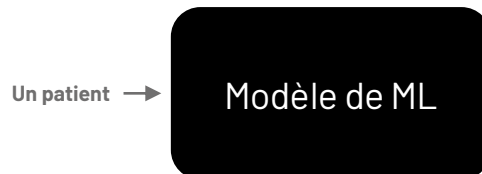
50 % de chances de bien prédire



Risque de réhospitalisation
pour ce patient : **0,48**

Prédiction par un algorithme de ML

70 % de chances de bien prédire



Risque de réhospitalisation
pour ce patient : **0,93**

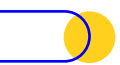
Expliquer la prédiction



**"Au-delà des prédictions de l'outil,
j'ai surtout besoin de savoir ce qui se
passe, quelles sont les variables qui
jouent un rôle."**



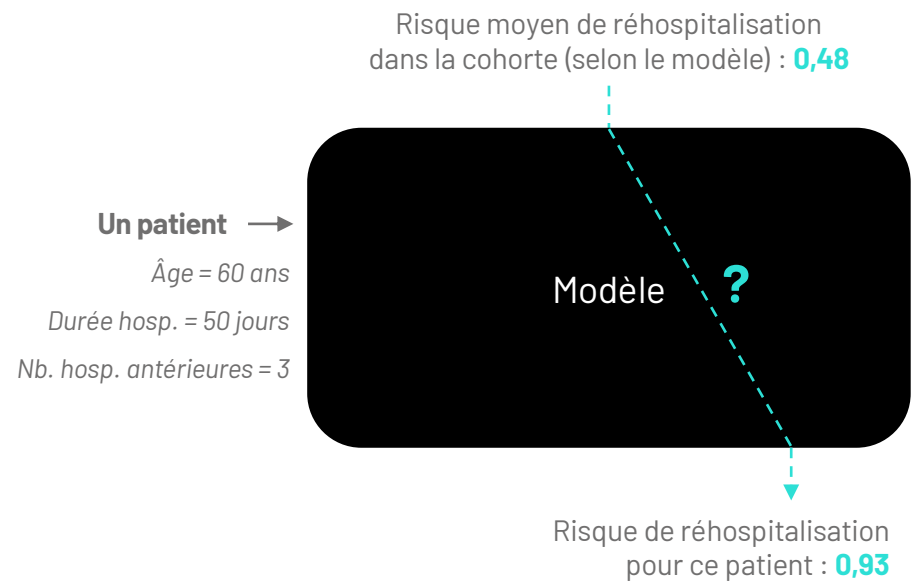
Expert médical



“SHAP values” : Que se cache-t-il derrière la boîte noire ?

Pour chaque patient, mesurer l'impact de chaque variable sur la prédiction

Explication de la prédiction pour un patient

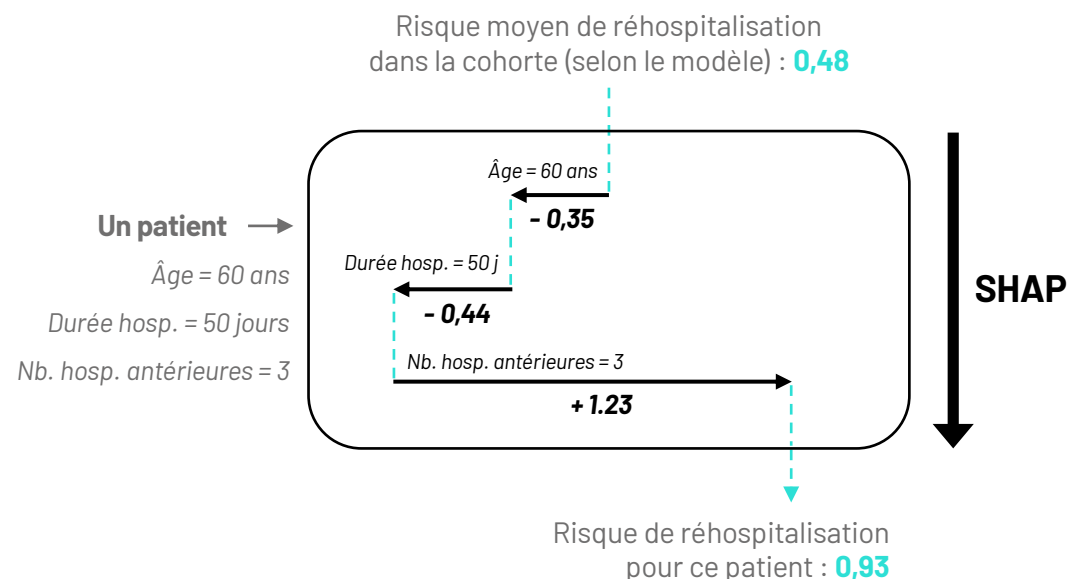




“SHAP values” : Que se cache-t-il derrière la boîte noire ?

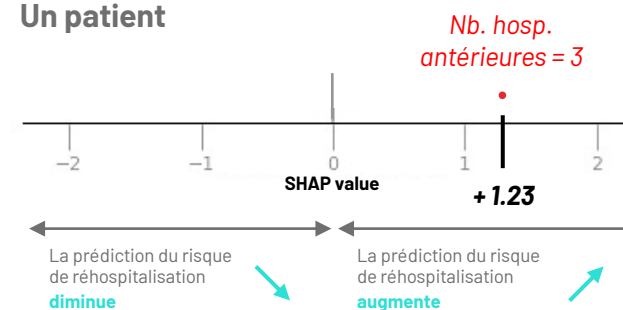
Pour chaque patient, mesurer l'impact de chaque variable sur la prédiction

Explication de la prédiction pour un patient

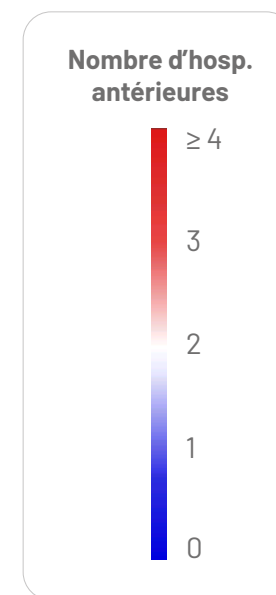
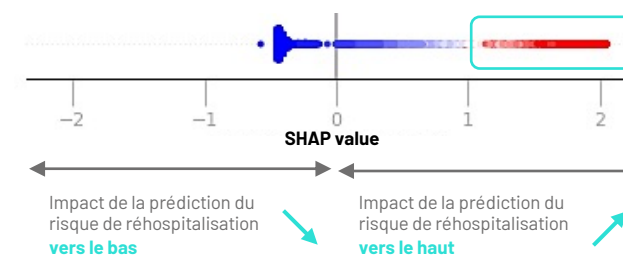


Focus sur l'impact du nb d'hosp. antérieures

Un patient



Tous les patients

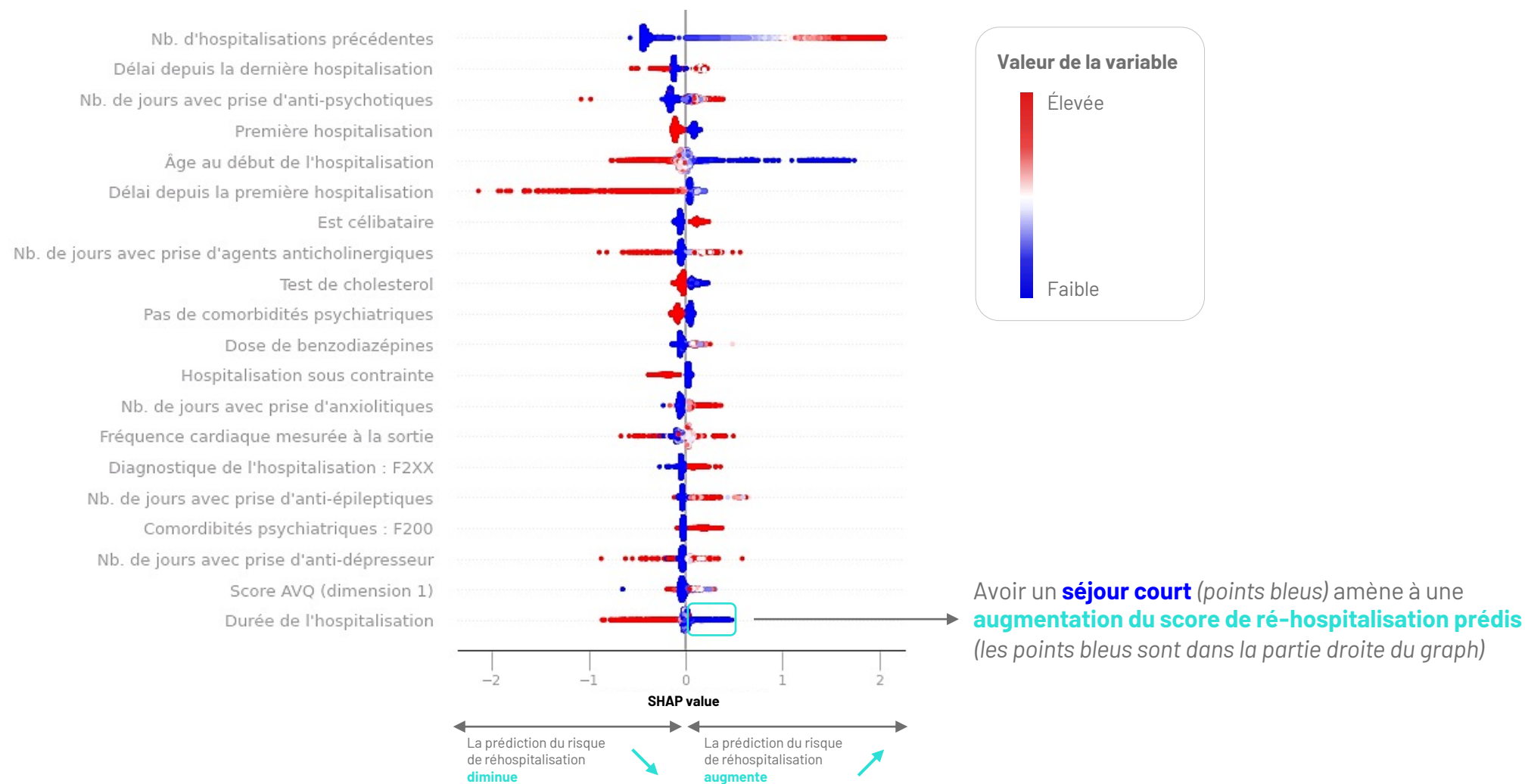


→ Globalement, un nombre **élevé d'hospitalisations antérieures** conduit à **l'augmentation du score de ré-hospitalisation prédis**.



“SHAP values” : Que se cache-t-il derrière la boîte noire ?

Pour toutes les variables



Conclusion

Prédire puis interpréter une prédiction



**Question
de recherche**



**Data et
Algorithmes**



**Prédiction et
Interprétation**

Cette méthodologie peut s'appliquer à un risque de survenue d'évènements indésirables, de décès, de délai vers l'IRCT, etc.



Analyses prédictives

Etude publiée

Prédiction du risque de réhospitalisation en psychiatrie dans les 2 ans

467047

G. Post¹, N. Assil², A. Schmidt², C. David³, V. Mary³, P. Pospieszyski⁴, P. Lemelle²

¹Psychiatre, consultant médical e-santé, dr.post-geoffrey@gmail.com ; ²Heva, Lyon, France ; ³GIP SIB ; ⁴GIP SYMARIS

DSVR 2023

15^e Colloque Données de Santé en Vie Réelle

20 Juin 2023 Paris (France)

2023



LES ENTREPRISES DE LA
RECHERCHE CLINIQUE

Introduction

- La psychiatrie hospitalière publique est confrontée à une augmentation du volume de patients avec des moyens qui se réduisent.
- En particulier, 50 % des patients hospitalisés seront réhospitalisés dans les 2 ans. Plusieurs mécanismes sont évoqués dans la littérature scientifique pour l'expliquer : situation sociale, diagnostic clinique, comorbidités, antécédents médicaux, comportement durant l'hospitalisation, traitements reçus.
- Dans ce contexte, diminuer le nombre d'hospitalisations apparaît comme un enjeu majeur.

Objectifs

L'objectif de notre étude était de fournir un score de risque de réhospitalisation dans les 2 ans suivant une hospitalisation en psychiatrie.

Méthode

Source de données

- Base hospitalière du CH de Rouffach, issue des dossiers médicaux.
- Contient des données socio-démographiques, cliniques (tests biologiques), diagnostiques, traitements pris pendant le séjour et ceux prescrits à la sortie.

Design de l'étude

- Inclusion des patients entre le 01/06/2008 et le 31/12/2018.
- Suivi des patients jusqu'au 31/12/2018.
- Variable à prédire = risque de réhospitalisation du patient en psychiatrie dans les 2 ans suivant une hospitalisation (par unité patient-hospitalisation).

Résultats Machine Learning

Définition des modèles

- Le GB était le plus performant avec 89 % de justesse et d'aire sous la courbe ROC (vs. 87 % pour le RF). Le modèle GB final était de profondeur 3, avec 500 arbres intermédiaires.

Fixation du seuil

- Un score proche de 0 ou de 1 montre que l'algorithme prédictif est confiant, tandis qu'un score autour de 0,5 montre une incertitude (Figure 1). Il a donc été décidé de fixer une plage de valeurs au sein de laquelle ne sont pas prises en compte les prédictions estimées trop incertaines. La recherche de cette plage se fait sur le jeu de test, l'objectif étant d'obtenir de bonnes métriques de performance tout en gardant une prédiction pour un grand nombre de patients.

Figure 1. Définition des plages de prédiction : exemple sur 4 patients-hospitalisations



→ Le seuil final retenu est 0,25 (Figure 2) :

- haut risque de réhospitalisation : score $\geq 0,75$
- faible risque de réhospitalisation : score $\leq 0,25$

→ Prédiction pour 11 676 patients-hospitalisations (34,7 %) tout en ayant de très bonnes métriques : 88 % de justesse et d'aire sous la courbe ROC.

Figure 2. Choix du seuil

Figure 4. Impact de la durée de séjour sur la prédiction, en fonction de l'âge

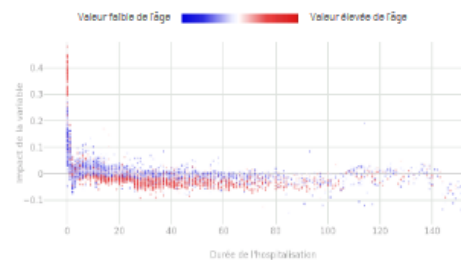
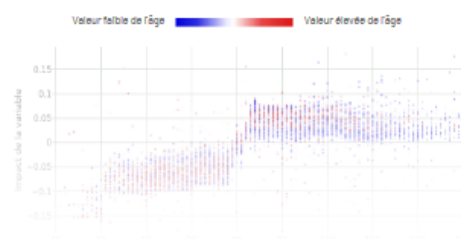


Figure 5. Impact de la fréquence cardiaque sur la prédiction, en fonction de l'âge



Heva



*Avec le soutien institutionnel
d'Astellas Pharma France*

Merci.



DOCAPOSTE

l'avenir devient plus simple

MAT-FR-NON-2023-00297